

The Relationship Between Foundations and Principles in IP Law

ROBERT P. MERGES*

TABLE OF CONTENTS

I.	INTRODUCTION	957
II.	ON “WHETHER”: FOUNDATIONS AND STRUCTURE.....	960
III.	MORE ON FOUNDATIONS AND MIDDLELEVEL PRINCIPLES.....	961
	A. <i>Utilitarianism vs. Efficiency</i>	962
	1. <i>Rejecting Utilitarian Foundations</i>	962
	2. <i>The Difference Between Utilitarianism and Efficiency</i>	964
	B. <i>An Example of Midlevel Principles at Work</i>	967
	C. <i>Other Candidates for the Midlevel Principles?</i>	968
IV.	FOUNDATIONS AS TIEBREAKERS.....	971
	A. <i>On Ties, and How To Break Them</i>	972
	B. <i>Reviewing the Three Levels of Analysis</i>	974
	C. <i>Is There a Risk of Unraveling?</i>	976
V.	CONCLUSION	978

I. INTRODUCTION

In my book *Justifying Intellectual Property (JIP)*, I wrote about what I call the “foundations” of the field of intellectual property (IP) law.¹ I tried to distinguish between a foundational level of discourse and another level, the level of basic principles. In the San Diego conference at which my book was discussed—and in several other settings as

* Wilson Sonsini Goodrich & Rosati Professor of Law, and Faculty Director, Berkeley Center for Law & Technology, University of California, Berkeley; Visiting Professor of Law, Stanford Law School, Spring 2013.

1. See ROBERT P. MERGES, *JUSTIFYING INTELLECTUAL PROPERTY* (2011).

well—the most frequent and persistent line of questioning about my book centered on the relationship between these two levels. That is what this brief Article is about.

For me, the foundational question in the IP field is whether IP should exist at all. Should society award state-backed exclusive rights—property—over works of creativity and imagination, and if so, why? In Part I of *JIP*, I say yes. I argue that on the basis of the works of Locke and Kant respecting property rights, IP protection is justified.² I go on to say that, consistent with the work of John Rawls, IP represents a just social institution from the perspective of the fair distribution of resources in society.³ It is not, in other words, fatally elitist.

Having answered the foundational question in the affirmative, I proceed in *JIP* to describe four basic or “midlevel” principles of IP law.⁴ These are the basic conceptual building blocks of the legal field as it is practiced in the world. The four principles themselves are straightforward: (1) efficiency; (2) nonremoval or preservation of the public domain; (3) proportionality; and (4) dignity. Despite my best efforts to be clear, some readers have been confused by the relationship between these principles and the foundations of the field.

Much of the head-scratching stems, I think, from my rejection of utilitarianism as an effective foundation for IP law. This is a function of two related issues. Part I of *JIP* does not say enough, I have come to see, about exactly what I demand of a foundation and what my criteria are for accepting or rejecting one. This I address in Part II of this Article. Another aspect of the first part of *JIP* is the notion of foundational pluralism; this too has led to some interesting questions. How, it was asked, could I both insist on the need for foundations and yet tolerate, or even invite, multiple and inconsistent alternative groundings for the field? Part II of this Article also reviews my thinking in this regard. After Part II, the second major topic I address is how I can reject utilitarianism as a foundation and yet include efficiency as one of the four basic principles of IP law. The answer to this is the heart of Part III below, where I try to show the difference between a foundational justification—which must answer the “whether” question, as in “whether we should have this field”—and a basic conceptual principle that helps a field operate effectively, once it is decided that such a field ought to exist. The “how” of midlevel principles, in other words, is a very different inquiry from the more fundamental question of whether. My argument is that we cannot justify IP rights based on current knowledge

2. *See id.* at 31–101.

3. *See id.* at 102–38.

4. *See id.* at 139–58.

of their costs and benefits. The data are inadequate to the task. But, given the decision to have IP rights, important features of our current IP regime can be explained by recourse to the principle of efficiency. Using the best data and analytic techniques at our disposal, we as a society try to maximize the net benefits of IP rights. Put simply, the argument is that we do not have IP rights because we are sure they are efficient, but given that we have them—because they are justified by Lockean, Kantian, and other precepts—we strive to make our IP system as efficient as possible. So efficiency explains and ties together important aspects of the IP system as practiced, even though we cannot be sure that as a whole, IP systems are a net benefit and thus efficient in an overall or global sense. To state this simply, I would say that efficiency is a good quality or feature for an IP system, but not an acceptable rationale for it.⁵

My insistence on foundational pluralism, together with the idea that foundations serve a very different purpose compared to midlevel principles, leads to a final, obvious question: Is there any sort of working relationship between the two? Even if one accepts the very different conceptual roles of foundations and principles, it still makes sense to ask whether foundations influence principles in any way. This is another issue I have been asked about by several readers of *JIP*.

The answer is in Part IV of this Article. Here I develop some comments from *JIP*, in an attempt to clarify the “tiebreaking” role of the foundations in cases where multiple midlevel principles are in conflict. The basic notion is that in the hardest of cases, one will and should have resort to one’s ultimate normative commitments in taking a personal position. So this type of case represents the rare situation in which the “whether” question influences the “how” of the IP system. This does not change my view that for almost all purposes foundations are functionally independent of the midlevel principles, however.

Part V concludes with some final thoughts on the need for foundations, the role of midlevel principles, and the importance of detailed doctrines and practices in IP law.

5. In some related writing, I respond to two related concerns: (1) the possibility of foundations independent from midlevel principles; and (2) the relationship between midlevel principles and the doctrinal/institutional details of IP systems. See Robert P. Merges, *Foundations and Principles Redux: A Reply to Blankfein-Tabachnick*, 101 CALIF. L. REV. (forthcoming 2013).

II. ON “WHETHER”: FOUNDATIONS AND STRUCTURE

For me, a proper foundation for IP law must answer one question—whether society should grant IP rights at all. Answering this “whether” question, and providing solid reasons for the answer, is what makes a workable foundation.

One immediate objection to the idea of foundations is this: How can I say whether IP rights make sense unless I know at least the main features of the IP system being proposed? This perspective would reject as artificial the distinction between foundations and basic principles and would insist that “whether” is always dependent on how an IP system is put together at the structural level.

A recent book describes IP law very much along these lines. In *Constructing Intellectual Property*, author Alexandra George argues that “intellectual property exists only after it has been brought into being by law, and . . . is therefore a figment of the collective imagination.”⁶ She does not take an explicitly normative view, opting instead for a more phenomenological approach that takes the existing social practice of IP law as a starting point.⁷ She is not concerned with the “norms contained in [IP] laws,” but with “how and in what form they came to exist.”⁸ She aims to expose the generative story behind IP law as it now stands and to thereby correct the misunderstanding that this body of law is founded on timeless, straightforward principles.⁹

My approach in *JIP* Part I is obviously quite different. I draw on a long tradition of philosophers who seek a pure bedrock starting position. Whether it is Locke with the state of nature, Kant with his purely abstract thought experiments on topics such as property, or Rawls with his original position, the idea is to move away from factual detail toward an idealized account that lays bare the foundational components of the institution or practice in question. This is not to say that I completely reject an inductive or phenomenological approach; this is in fact an aspect of the midlevel principles in Part II of *JIP*. It is just that, for me, Part I’s idealized step back is a necessary first move in thinking about the desirability of IP rights.

Put simply, the “whether” question asks whether there is at least one conceivable IP system, one imaginable set of rules and institutions, that could be justified in awarding IP rights. The only starting point is the requirement that the ideal system in question really be capable of

6. ALEXANDRA GEORGE, *CONSTRUCTING INTELLECTUAL PROPERTY* 6 (2012).

7. *Id.* at 29.

8. *Id.*

9. *See id.* at 31–32.

accurate description as an IP system, that is, that it include individual exclusive rights over discrete intangible objects. Any imaginable instance of such a system will do for purposes of the initial “whether” question. If one or more imaginable instances of an IP system can be justified, the analysis moves on to a consideration of the structural features of an actual, operational IP system. In *JIP*, this second stage is conducted with reference to the existing IP systems of advanced economies, primarily the United States, but with Europe, Japan, and China always in the background too. The midlevel principles described in Part II of *JIP* are meant to be norms that mediate between foundational commitments and the detailed particularities of actual practices; in my account, midlevel analysis comes only after the initial “whether” question has been answered in the affirmative.¹⁰ If an IP system can be envisioned that comports with foundational beliefs, attention then shifts to identifying norms consistent with these beliefs that arise from actual rules, doctrines, and institutions on the ground. These norms both reflect existing practices and help to shape or guide them in the direction of an operational system more in keeping with foundational commitments.

III. MORE ON FOUNDATIONS AND MIDLEVEL PRINCIPLES

As I said earlier, for me a foundation answers the question of “whether.” Principles, on the other hand, are about “how.” For most purposes, there does not seem to be much confusion on this score. The proportionality principle, for example, explains a substantial number of IP rules and practices, yet I have not so far heard an argument that proportionality is adequate as a conceptual foundation for the field of

10. Generations of people with diverse but at least somewhat compatible ultimate commitments have left their imprint on the IP system when deciding difficult cases. This imprint takes the form of philosophical fragments (such as brief references to Locke), intuitions, and archetypes. Judges and policymakers draw on this shared body of ideas when deciding difficult cases; and from these cases the midlevel principles arise. Thus, I see midlevel principles as high-level, repeating metathemes in IP case law. At the conceptual level, I find that Rawls’s account of a shared public space provides a good model for thinking about these principles. Not an empirical account of where they come from, but an idealized way to describe and discuss them. See JOHN RAWLS, *POLITICAL LIBERALISM* 97–98 (expanded ed. 2005) (“[G]iven the fact of reasonable pluralism, citizens cannot agree on any moral authority We adopt, then, a constructivist view to specify the fair terms of social cooperation . . . [based on] fundamental ideas of the public political culture . . .”).

IP.¹¹ It seems readily apparent that given the decision to award IP rights, it makes sense to award them in rough proportion to the effort or contribution embodied in the works they cover.

A. *Utilitarianism vs. Efficiency*

The exception to this is the efficiency principle. A number of *JIP* readers have remarked on the seeming inconsistency of rejecting utilitarianism as a foundation and then adopting efficiency as a basic operating principle. But the answer to this conundrum lies, I believe, in a proper understanding of the differences between foundational concepts and principles. The main point here is this: I do not believe there is adequate data at this time to fully commit to a utilitarian grounding for the IP field. And yet I believe, based on Locke, Kant, and Rawls, that the answer to the “whether” question is “yes.” In a hypothetical sense, my cocitizens have reached a similar conclusion, based on whatever foundational values guide them. Having so decided, society has implemented a robust system of IP protection, which has as one of its most basic and recognizable features the idea that IP rights ought to be awarded and administered as efficiently as possible so as to maximize net benefits as best we can determine.

1. *Rejecting Utilitarian Foundations*

In stating why I personally reject utilitarian foundations for the IP field, I say repeatedly that the data on hand are inadequate.¹² This raises two questions, which readers of *JIP* have been quick to see: (1) Does a

11. It is important to note, however, that midlevel principles do embody and express objectives and values. They are not merely or only operational guidelines. Though they ultimately guide the “how” of IP policy, they can be conceptualized as products of the hypothetical political-deliberative process envisioned by Rawls and best captured in the idea of an overlapping consensus. This consensus is reached in a manner Rawls describes as “public” and “political,” by which he means that values and objectives are characterized in terms distinct from one’s ultimate commitments. Only these ultimate commitments can fully answer the “whether” question; but the deliberative procedure based on them nevertheless yields midlevel principles that reflect an attenuated version of foundational values. Alternatively, one might say that midlevel principles are produced by a public and political procedure; and that this procedure reflects indirectly the ultimate commitments of those who engage in it. So the existence of midlevel principles, on this account, reflects consensus on the “whether” question, despite the fact that their primary function is to guide the “how.”

12. MERGES, *supra* note 1, at 3 (“The sheer practical difficulty of measuring or approximating all the variables involved means that the utilitarian program will always be at best aspirational. Like designing a perfect socialist economy, the computational complexities of this philosophical project cast grave doubt on its fitness as a workable foundation for the field.”).

utilitarian foundation really require such precise data?; and (2) Why am I willing to adopt alternative foundations rooted in fuzzy and amorphous deontological notions of autonomy and fairness while rejecting utilitarianism due to a lack of stringent proof? Both questions in a sense come down to this issue of consistency: am I being consistent when I reject utilitarianism for lack of precise data while championing a set of foundations that themselves hardly seem a model of precision? After all, there is nothing like universal agreement that Locke and company have arrived at the one correct and true basis for awarding private property rights. So why do I say their reasons are superior in the case of IP rights? Is it not better to build on solid data, even if incomplete, rather than vague and sketchy concepts drawn from outmoded and opaque philosophic texts?

On the first question, I feel comfortable saying that I reject utilitarian foundations because data are what this body of thought is all about. It is in the nature of utilitarianism, in other words, to require adequate data in order to answer the question of “whether.” Without data, it is impossible to say whether a single act or an entire practice creates a net social benefit or not. It is meaningless to say that the theory requires data for answers, yet in the face of inadequate data we can use the theory anyway. Data are its lifeblood, its essence. Therefore, lack of data is fatal. The theory has nothing to say when the data are not adequate. If a theory depends on adequate data to answer the “whether” question, and the data are not available, the theory cannot answer the question. Any assertion that it can must resolve to one of two other assertions: the data are in fact adequate, or the theory works even when a rough approximation of adequate data is all we have to work with.

For reasons I explain in *JIP*, I do not see how anyone can reasonably contend that the empirical data on IP protection can make a tight and solid case that IP protection is a net benefit for society. There may come a time when that is true, but I do not believe we are there yet.

As for the second proposition, this strikes me as a highly softened and lightened version of utilitarianism—one that strips it of much of its claim to superiority. The great advance that this theory represents, after all, is that it proposes to resolve centuries of disagreement over first principles. And how? By recourse always and only to the facts, the objective data, the measurable and inarguable record of positive and negative effects. To reject this stringency and adopt a deformed alternative is to give up any claim to the theory’s overall superiority. Because what can one use

to fill in the gaps in the data? On what basis, if not firm data, can one say, “This social practice creates net benefits”? Whatever that basis—reason, intuition, emotion, et cetera—it represents precisely the grounds for social decision that utilitarianism rejected at its birth. “Soft” or “light” utilitarianism represents the worst of all worlds: it is not actually data driven, but it claims the cloak or mantle of objective data. It is a fraud, playing on the natural propensity of many people to respect data and all conclusions that are firmly grounded in it.

Once the skin of objectivity is pierced, what lies underneath in the case of utilitarianism is less convincing than certain alternative normative justifications for property (in general) and IP (in particular). That is the essence of the argument in *JIP* Part I.

2. *The Difference Between Utilitarianism and Efficiency*

With all this said about utilitarianism, I turn now to efficiency. I seem to have done a thorough job casting aspersions on utility. So how can I then champion efficiency as a basic principle of IP law?

The answer, in a nutshell, is this: once we have decided to pursue a social practice—such as the awarding of IP rights—it makes eminent sense to pursue this end in a way that maximizes benefits and minimizes costs *as best we can determine them*. Efficiency, then, describes a way of carrying out a certain end. It is not an independent rationale for that end. Rationales are the domains of foundations, and efficiency is not foundational. It is a quality or characteristic of a social practice, but it cannot provide a reason or ultimate motive for that practice.

Because efficiency has become intertwined with foundational issues, I believe its true meaning is often obscured today. But the *Oxford English Dictionary* definition of *efficient* is quite clear: “Productive of effects; effective; adequately operative. Of persons: Adequately skilled.”¹³ The key here, I think, is the notion of effective or adequate, as opposed to ideal. An efficient way of doing something is a way that works well. Efficiency is a quality or feature of the way a certain end is pursued. But it does not define the thing to be pursued; it does not and cannot provide the end itself. Efficiency then is in some sense always a supporting concept when compared to the end it serves. It is adjectival, so to speak; it describes a way of doing something but in no sense does it guide us in the initial choice of what to do.¹⁴

13. 5 THE OXFORD ENGLISH DICTIONARY 84 (2d ed. 1989).

14. It might seem that applying efficiency as a global principle counters this argument. The idea would be to list all possible actions or ends, and then weigh out the costs and benefits of each, choosing the highest net-benefit action first, and proceeding through all items of the list that are at least net positive in terms of benefits. However, the data

When put this way, it is perhaps easier to see how I can both reject utilitarian grounds for IP law, yet vigorously support the idea of pursuing IP rights in an efficient manner. The data, I contend, are inadequate to the task of answering the “whether” question. But once an independent (for example, deontological) foundation is established for the field, whatever data can be found are most helpful in determining “how.” And here again the definition offered above comes into play. Efficiency is a matter of something being “adequately operative.”¹⁵ In most areas of IP law and policy, we have at least some data, some useful information to guide our actions. The data will necessarily be incomplete in most if not all cases. But that is not fatal to the project of efficiency. Adequate operational guidance is enough. It is quite helpful to know that it is more efficient to pursue goal *X* by means of action *A* than by action *B*. In this way, an independent basis for deciding that goal *X* is in fact worthwhile shifts attention to the best available information on how to go about achieving goal *X* most effectively.

One way to get a better grasp of what I am talking about here is to see efficiency as essentially an “ordinal,” as opposed to cardinal, concept.¹⁶ That is, when we are talking about efficiency, we often make an implicit comparison between two states, one more efficient than the other. Comparison is the essence of the matter. Action *A* is more efficient—or, as we often say, “a more efficient way of achieving *X*”—only as compared to action *B*, which leads to a lower net social benefit. To determine efficiency in this sense, we only need data on the pluses and minuses of the two alternative actions. This type of analysis does not compare action *A* to all other possible actions, but only to action *B*. The universe of comparison is closed, rather than open. And what closes it, of course, is the independent decision that goal *X* is worth pursuing.

Another way of putting this is that we can pursue efficiency when data are incomplete, so long as we have a given end in mind. We need *some* data—enough to make a comparison between rival courses of action—but we do not need to decide if the ultimate end or purpose is the best among all possible ends. The efficiency calculus takes this end or goal as a given; it requires only some comparative data about the ways that

required for such an approach are obviously not likely to be at hand. Indeed, this global form of efficiency is none other than utilitarianism in disguise, which means that my earlier critique of utilitarianism—at least in the IP setting—applies here.

15. 5 THE OXFORD ENGLISH DICTIONARY, *supra* note 13, at 84.

16. MERGES, *supra* note 1, at 189.

the given end might be pursued. This is, of course, much more tractable. And this feature of efficiency is what makes it a valuable contributor to the tool kit available for IP analysis.

To take a concrete example, consider the term of protection for patents and copyrights. Currently these terms are set at a fixed number of years by statute.¹⁷ Many observers of the IP scene have argued over the years that these fixed terms are inefficient. For patents, a leading proposal is to calibrate patent terms to the requirements of various industries: make software patents shorter than the standard twenty-years-from-filing term, for example,¹⁸ and perhaps retain the current fixed term for pharmaceutical patents. For copyrights, an intriguing proposal was made by William Landes and Judge Richard Posner to permit infinitely long copyright terms that require frequent renewals, on the theory that lack of incentive to renew would shorten the effective copyright term for a huge number of copyrighted works, making the scheme as a whole far superior to the current one-size-fits-all regime.¹⁹

These arguments are typically couched in the phraseology of efficiency, as well they might be. The question of optimal IP terms lends itself to the type of comparative analysis I have described. Will overall social cost be lower with variable patent terms, given the dual realities of higher administrative costs and fewer long-lived patents, that is, with their attendant social costs? And for copyrights, will the heavier burden of frequent renewals be worth the savings in social cost from having fewer long-lived copyrights, many of which may well be forgotten by their owners long before they expire? These questions are inherently about efficiency. They are specific examples of a more general theme, namely, the best way to conduct a patent or copyright system at the lowest net cost. But again, this is true only because the “whether” question has already been asked and answered. At least on these topics, we have decent comparative efficiency data. What we decidedly do not have is the kind of data that would be required to decide whether patent or copyright protection are good ideas overall.

17. There are some nuances, such as requirements for renewals and term adjustments in patent law due to administrative delays, but put these aside.

18. See DAN L. BURK & MARK A. LEMLEY, *THE PATENT CRISIS AND HOW THE COURTS CAN SOLVE IT* 96 (2009) (citing Leo J. Raskind, *The Uncertain Case for Special Legislation Protecting Computer Software*, 47 U. PITT. L. REV. 1131 (1986); Pamela Samuelson, *Modifying Copyrighted Software: Adjusting Copyright Doctrine To Accommodate a Technology*, 28 JURIMETRICS J. 179 (1988); Richard H. Stern, *The Bundle of Rights Suited to New Technology*, 47 U. PITT. L. REV. 1229, 1262–67 (1986); John C. Phillips, Note, *Sui Generis Intellectual Property Protection for Computer Software*, 60 GEO. WASH. L. REV. 997 (1992)).

19. See William M. Landes & Richard A. Posner, *Indefinitely Renewable Copyright*, 70 U. CHI. L. REV. 471, 473–75, 517–18 (2003).

B. An Example of Midlevel Principles at Work

Because of the strong utilitarian origins of the IP field, I have felt it important to explain the efficiency principle and to distinguish it from foundational issues. But efficiency is only one midlevel principle. Equally important are the other midlevel principles. They connect with, and indeed arise from, the detailed doctrines and practices that make up the rich surface texture of IP law. So it is crucial to see how these principles relate to this surface level.

To make this concrete, start with a long-running debate in copyright law. At various times, artists and authors grant licenses for their works to be adapted and exploited in various media. So an author might grant rights for a film adaptation of her story. Controversy arises when new technologies, such as videotape or DVDs, come along after a long-running license agreement is signed. Authors then claim that the contract granting the right to make a “film” does not include the new technologies. But owners of the right to adapt the novel to film of course disagree; they claim that film rights should be interpreted so as to include the new technologies.

The cases on this issue invoke a number of doctrines, drawn from copyright as well as contract law.²⁰ In the three-part analysis described in *JIP*, this means that at the level of detailed rules and practices, there is no clear or consistent answer to the problem of old copyright grants and new technologies. So it is natural to conduct policy discourse using the midlevel principles. These are generalized from multiple sources; they represent the distillation of broad concepts that span a multitude of

20. See, e.g., *Boosey & Hawkes Music Publishers Ltd. v. Walt Disney Co.*, 145 F.3d 481, 487 (2d Cir. 1998) (involving the licensing of film rights, where Boosey, owner of a movie copyright, argued it should be able to sign a new contract with another company for a videocassette version of the movie, and holding for Disney, to whom Boosey originally licensed film rights, stating, “The words of Disney’s license are more reasonably read to include than to exclude a motion picture distributed in video format”); see also *Paramount Publix Corp. v. Am. TriErgon Corp.*, 294 U.S. 464, 465–66 (1935) (interpreting rights to “sound recordings”); *Cohen v. Paramount Pictures Corp.*, 845 F.2d 851, 852 (9th Cir. 1988) (interpreting “television” rights). For a summary of the approaches courts have taken to these cases, see 3 MELVILLE B. NIMMER & DAVID NIMMER, *NIMMER ON COPYRIGHT* § 10.10[B], at 10-89 to 10-91 (2012), which contrasts the narrow, pro-licensor-creator view in the cases that licenses cover only media within the unambiguous core of the license grant, with the broad, pro-licensee view that licenses cover any use reasonably related to the use contemplated by the original grant. To the extent that Nimmer expresses a preference for the latter, broad view—which favors licensees—I disagree, for the reasons set forth in the next few pages of text.

discrete doctrines. They therefore help to bring clarity when doctrines and rules conflict.

In particular, in our film rights example, it could be said that because smaller units of production are more likely to promote the dignity interests of individual creators, a policy favoring these units would be an expression of the importance of the dignity principle in IP law. Yet there is more to it than dignity. As is well known by now, awarding ever more IP rights to ever larger numbers of discrete producers raises the specter of rather serious transaction costs. In terms of midlevel principles, efficiency concerns are a major obstacle to the promotion of creativity in smaller units of production.

In the end it is not important for present purposes to arrive at an answer to the question of copyright licenses and new technologies. Depending on one's estimate of the dignity value of creating alone or in a small team, as opposed to within a large company, and also on one's estimate of the increased transaction costs that would stem from a small-unit-favoring policy, the argument could come out one way or the other. The point here is to show how midlevel principles work in practice. What they do for this debate is to provide a common language, a standard conceptual toolkit, that permits sophisticated analysis and argumentation beyond the level of individual cases, norms, and practices. This allows discourse at a higher level of generality. It also spurs the field on to a deeper and richer plane—one that connects the pressing issues of today, with their inevitable technical and commercial details, with the ruptures caused by controversies of the past.

C. Other Candidates for the Midlevel Principles?

My plea for a middle level of discourse in the IP field seems to have caught the attention of at least a few readers.²¹ There seems to be support for the idea of such principles, but it is also safe to say that others have strong and divergent ideas about the precise contours they should follow. To some extent, I anticipated a good deal of push and pull over this part of *JIP*; as I said, “There are also other plausible candidate principles in addition to these four. . . . [T]here is room for a diversity of organizing and explanatory principles. . . . One chapter cannot do it all.”²² Yet even so, the critiques of the four principles I described are such that I feel the need to say a few more words in defense of my approach.

21. See Eric R. Claeys, *On Cowbells in Rock Anthems (and Property in IP): A Review of Justifying Intellectual Property*, 49 SAN DIEGO L. REV. 1033 (2012); David H. Blankfein-Tabachnick, *Intellectual Property Doctrine and Midlevel Principles*, 101 CALIF. L. REV. (forthcoming 2013).

22. MERGES, *supra* note 1, at 158.

A good example of the critique I am describing is found in the companion article for this volume by Eric Claeys. Professor Claeys believes that the principles I identified are deficient in that they do not draw enough on the structural and theoretical elements of property. In other words, there is not enough property theory in the middle layer of my three-part structure for his taste.²³

At the outset let me say that in writing *JIP* I tried for what might be described as “meta-midlevel” principles, ones that encompassed as broad a sweep of the IP field as I could take in. So I think if I had more time, I could integrate Professor Claeys’s property-oriented principles into the structure of *JIP*, perhaps with some major tweaks. There is much to be said for the idea of taking fine-grained aspects of property theory—roughly congruent with each stick in the property entitlement bundle, so to speak—and integrating them into more fine-grained midlevel principles. I think control rights, usufruct, trespass, and the like are all useful overarching concepts for IP law, and might effectively find expression as useful midlevel principles.

Nevertheless, I still feel that property concepts are often not going to be quite adequate to explain high-level themes in IP law. The parable of the bridge in my book,²⁴ for example, has a good deal of overlap with the concept of accession in the law of property. But I meant this parable to do more work than accession generally does. I wanted it to include an element of distributive justice and desert theory that I do not necessarily see in most accounts of accession.²⁵ Consider Professor Claeys’s own description of how accession would apply to the bridge parable. He describes the basic problem as whether A1, the owner of the land along the river bank, or the bridge builders, who construct a bridge touching on

23. See Claeys, *supra* note 21, at 1037, 1054–58.

24. MERGES, *supra* note 1, at 162–65.

25. At least one well-reasoned account of the accession principle explicitly recognizes that this principle is in conflict with certain normative impulses at work in the legal system:

[A]ccession creates a multiplier effect enhancing the wealth of owners of discrete assets that does not apply to those who have only a claim on a general share of societal resources. Finally, accession contributes to an expanding sphere of property and a concomitant shrinking of the public domain. If one believes that there is inherent value in preserving the public domain—a proposition widely endorsed today by many intellectual property scholars—then accession has problematic consequences for this reason, too.

Thomas W. Merrill, *Accession and Original Ownership*, 1 J. LEGAL ANALYSIS 459, 462 (2009).

Al's land, should be the sole owner of the land-plus-bridge combination.²⁶ As he says,

Although different doctrines treat accession differently depending on context, at the level of policy, accession focuses on: (1) whether the objects at issue are more likely to be used productively as separate entities or as a single package entity; and (2) whether observers are likely to perceive the objects as one "thing" or multiple "things."²⁷

This description emphasizes first, efficiency or productivity, and second, the perceptions of third parties as the key elements in working through the bridge scenario. What is missing, it seems to me, is a direct discussion of the respective contributions of the two parties. The accession story focuses on efficient and commonsense accounts of appropriate asset bundles, whereas the bridge story in Chapter 6 of *JIP* is meant to capture something broader: the relative contributions of the two parties—the degree and significance of their effort, the value unlocked by the part they play in creating the land-plus-bridge combination. Although IP cases and doctrines can be found that illustrate the importance of proper asset bundling, proportionality is meant to describe a wider variety of doctrines whose essence is the weighing of relative contributions. That explains why I pitched the proportionality principle at such a high degree of generality. Put another way, the accession story can be explained strictly in efficiency terms. Yet I find that in IP law, the handling of accession-type problems often includes a comparison of the relative contributions of the two parties to the dispute. This additional dimension implicates a principle other than straightforward efficiency of optimal asset bundling. I call this additional dimension proportionality.

I realize that all four of the midlevel principles are pitched at a high level of abstraction. This is no accident; it flows from the nature of the IP field and the major conceptual crosscurrents that cut through it. The notion of a midlevel principle comes from Jules Coleman, who first employed the idea to explain important features of tort law.²⁸ For Coleman, the principle of corrective justice is central to understanding tort law.²⁹ This served as a template for me—an example of a general overarching principle that connects and explains disparate aspects of a broad field. But IP law is very different from tort law. To be sure, there are ways in which IP law overlaps with torts. Wendy Gordon's work on

26. Claeys, *supra* note 21, at 1040–42, 1054.

27. *Id.* at 1054.

28. See JULES L. COLEMAN, *THE PRACTICE OF PRINCIPLE: IN DEFENSE OF A PRAGMATIST APPROACH TO LEGAL THEORY* 5 (2001).

29. See *id.*

IP law as unjust enrichment explores these overlaps in detail.³⁰ But there are vast swaths of IP law that are not adequately explained by concepts drawn from torts. For example, even though a defendant blatantly copies a patented invention, patent law will let the defendant off the hook if information about the invention in question was theoretically available in an extremely obscure locale that was unknown to the defendant. This appears to defy principles of corrective justice, but makes sense when the IP midlevel principle of nonremoval from the public domain is taken into account. Likewise, exceptions to IP infringement liability, such as fair use for copyright and experimental use for patents, are applied despite the fact that infringement is thought to harm right holders. Countervailing considerations, such as efficiency and proportionality, outweigh the harm in these cases.

My point is this: neither Coleman's corrective justice nor Claey's property-based principles are broad enough to embrace the diverse rules and doctrines at work in IP law. If this means that the midlevel principles of *JIP* seem highly general—if they seem like metaprinciples—then the cause must be said to be the nature of IP law itself.

IV. FOUNDATIONS AS TIEBREAKERS

So it is that the midlevel principles arise from the details of everyday IP practice, and then come to cabin and shape controversies in the future. But what about the foundational commitments I discussed earlier, those that take up Part I of *JIP*? Where do they fit in? How do they connect with and relate to the other two levels of analysis?

Basically, foundational commitments play the same role as the midlevel principles, only at a higher level of analysis. Foundational values underlie all the midlevel principles. Yet the foundations of the IP field serve a different purpose from the other levels of analysis, and this makes a difference in how they can be applied in controversial cases. Foundations answer the “whether” question: whether it makes sense for society to make a social commitment to IP. Because they serve a very different function, they cannot be expected to operate interchangeably with midlevel principles and detailed doctrines. Sometimes, in rare cases, foundational commitments can help resolve deep-seated conflicts between principles,

30. See Wendy J. Gordon, *Of Harms and Benefits: Torts, Restitution, and Intellectual Property*, 21 J. LEGAL STUD. 449 (1992).

but in the general run of cases, it is the principles themselves that must be used to resolve conflicts of this nature.

This is so because of the inherent pluralism built into the foundational level. As I put it in Part I of *JIP*, there is “room at the bottom” when it comes to foundational justifications for IP law.³¹ Different and nonoverlapping justifications are possible. So one who finds the ultimate rationale for IP protection in the pages of the Torah and one who finds it in a detailed cost-benefit analysis based on the best available data can both support and contribute to the basic institutions of an IP system. Yet for those institutions to work well, it is important that these competing justifications not come into conflict too often. It should not be necessary in most cases to resort to these ultimate commitments to resolve a pressing problem; that is what the midlevel principles are for. They provide a shared language and a common set of conceptual tools with which to discuss difficult policy issues. They are the equivalent to Rawls’s “overlapping consensus” in the world of IP.³² Their entire purpose is to make it unnecessary in most cases to resort to foundational commitments to resolve the clash of competing policies.

A. *On Ties, and How To Break Them*

Yet even so, there are occasions when foundational commitments will color one’s view as to which competing policy wins out. In most cases, it will be enough to choose one midlevel principle over another: efficiency over proportionality, or dignity over nonremoval, for example. But in a few cases the competing principles may be in equipoise. A compelling efficiency story may run headlong into a persuasive threat to the dignity interest. In such a case, a foundational commitment may come into play as a “tiebreaker.”

Consider the example of copyright licenses and later-developed media, which we discussed earlier in Part III.B. When an author licenses, say, “film rights” to a licensee, there are good reasons to permit the licensee to exploit the licensed work in videotape or DVD format. One important argument is that reading the grant broadly in this way prevents the author from slowing down the diffusion of works in the new media. For to read the grant narrowly is to permit the author to renegotiate the original license—a process which permits the author to revisit the terms of the original deal and to propose new terms more in the author’s favor. The licensee’s settled expectations may be upset. And in any event, the transaction costs of all these renegotiations increase the net cost of the

31. MERGES, *supra* note 1, at 10.

32. *Id.* at 9–10 (quoting RAWLS, *supra* note 10, at 134).

new media format, thus delaying its adoption in the market. The midlevel efficiency principle is, as I argued earlier, surely implicated here.

The counterarguments in this situation largely revolve around issues of fairness—that is, proportionality—and dignity. Why should the licensee enjoy the windfall that flows from the development of a new media format? Was it not the author who created the work that now may reach potential new markets via the new medium? And if renegotiation occurs, is this not a good thing in that the author may be able to obtain a bigger slice of the profits than he or she was able to obtain in the original licensing deal? Surely this implicates the author's ability to make a living as an author, an important aspect of the dignity principle as I describe it in *JIP*.

A conflict of principles at this level may thus lead one back to foundational commitments. If one believes, as I do, that autonomy lies at the very heart of Kant's defense of property rights, one might conclude that because smaller production units increase individual autonomy, the policy favoring small-unit production should be followed as often as it is reasonably feasible. This is true despite the fact that transaction costs will increase under such a policy.³³ In the specific case under discussion, the conclusion would be to read the initial grant of film rights narrowly, so that the original author does indeed have a right to negotiate a new license when new media are later developed.

Kant's emphasis on autonomy supports an appeal to the value of small, autonomous units of production even in the face of higher than optimal transaction costs. This could take many forms in IP law, but the basic idea would be to provide some special encouragement for the creation of IP-protected works by individuals or small companies. Such a proposal would stress the benefits not only of creative works in general, but more specifically of a certain locus of production for such works—creation by lone individuals or within small companies or teams.

In the context of the film adaptation example given earlier, resort to Kantian foundations would support a policy of favoring the original creator in a new technology controversy. If we take seriously Kant's concern for autonomy, we would favor the original creator who granted film rights at an earlier time. Deciding that this grant does not encompass

33. See generally Robert P. Merges, *Autonomy and Independence: The Normative Face of Transaction Costs*, 53 ARIZ. L. REV. 145 (2011).

videotape or DVD formats would of course favor the creator-grantor; it would mean that the licensee in the original contract does not automatically have rights to the videotape and DVD version of the film they created under the original license. The author-creator who made the earlier grant could then choose to reach a separate license contract with another company. Interpreting the original agreement in a way that permits this new contract gives the creator—the author of the novel in our example—greater freedom to strike a new deal. This he or she might choose to do if the original film version was not satisfactory, not true to the spirit of the novel, or simply not as profitable as hoped.³⁴ The author could even strike a new deal with the holder of the original film-rights license, in which case the only effect of the contract interpretation would be to enhance the income of the creator of the original work. The overall point remains the same: an emphasis on Kantian autonomy leads to an interpretation of the original contract that favors the creator or author, because doing so increases the viability and independence of the small-unit-of-production entity—the author.

B. Reviewing the Three Levels of Analysis

The new technologies example provides a good opportunity not only to see how foundational commitments can act as tiebreakers, but also to review all three levels of analysis described in *JIP*. By way of summarizing the preceding thought, we can think of the three levels in the following way as applied to the problem at hand:

34. There is an obvious corollary here in those provisions of the Copyright Act permitting authors to terminate earlier transfers of rights after the passage of time or upon the grant of an additional copyright term by the legislature. See 17 U.S.C. §§ 203, 304 (2006); Peter S. Menell & David Nimmer, *Pooh-poohing Copyright Law's "Inalienable" Termination Rights*, 57 J. COPYRIGHT SOC'Y U.S.A. 799 (2009).

Level	Analysis
3. Detailed Rules and Practices	Cases interpreting copyright licenses in light of new technologies are mixed and somewhat inconsistent. Leading treatises provide nuanced guidance, but close cases are nevertheless difficult to decide on the basis of existing doctrine.
2. Midlevel Principles	Efficiency might dictate lesser fragmentation of rights, reduction in holdup opportunities; this might run in favor of the original licensee, meaning that a holder of film rights should also have the right to sell the movie in videotape and DVD format. But the dignity interest of the original creator might push in the direction of permitting an author to enter a new exploitation contract if dissatisfied with the original film version.
1. Foundations	Kantian autonomy for original authors means greater control of and remuneration from exploitation of works in different media, using new technologies developed after the initial grant of rights.

I do not mean to make this all seem too algorithmic. Yet at the same time, the relationship between the three levels of analysis in *JIP* can, I think, usefully be seen as a progression of sorts. Starting at the highest level, that of detailed rules and practices, a specific problem is first placed into its proper doctrinal frame. Standard practices are consulted. If an easy answer emerges, that may well be the end of things. Of course, conventional understandings may come to seem somehow at odds with more basic principles, so the analysis does not necessarily end here. But in many cases this is as far as one needs to go. Where conventional doctrine and practices represent a stable embodiment of midlevel principles, and where those principles rest on a convincing conceptual foundation, the conventional answer may well be adequate and indeed satisfying.

Closer cases require a drop down to the midlevel principles. Doctrines may be in tension; practices may be in flux, or under pressure, from changed conditions. When this is so, a move to deeper conceptual principles is in order. Often, doctrines will come into tension because underlying principles dictate opposing results in a given case. Resorting to the principles themselves then makes sense. And once a doctrinal question is posed in terms of an explicit choice between competing

principles, it may make a decision easier. It will at any rate deepen the conversation—and recall that a key benefit of the midlevel principles is that they provide a shared conceptual space, or common vocabulary, in which to conduct such conversations. This is because the midlevel principles are derived from a host of different doctrines and practices. Conversation at this level thus necessarily abstracts away from strictly doctrinal details. It allows for a more general, more policy-oriented type of analysis than the specifically doctrinal level permits.

Often the choice of principle will be only slightly contentious. Where this is so, the analysis need not go beyond the midlevel. But, as described above, there are times when the midlevel analysis leaves one quite at sea as to the proper direction to take. Major principles such as proportionality and efficiency or dignity and nonremoval may be in equipoise. On these occasions, resorting to the foundational level as a tiebreaker becomes necessary. Only reference to one's fundamental commitments is sufficient to resolve a knotty issue that ends up stalemated in the realm of midlevel principles.

C. Is There a Risk of Unraveling?

The tiebreaker metaphor works well enough, but it raises a potential difficulty. What prevents every issue, or at least very many issues, from being declared a deadlock, requiring resort to the next lower level of analysis? What if, in effect, most controversies are or can be characterized as ties? Then resort to the tiebreaker is necessary in most if not all cases, which collapses this fine-tuned, three-stage analysis into what is effectively only one stage. And what is worse, from the point of view of predictability, the tiebreaker level is one where pluralism is quite intentionally tolerated. Which of course means that many issues will result not only in contentious arguments, but also in battles over deeply held, fundamental convictions. What then? Does the entire structure necessarily collapse under the heavy weight of too many tiebreakers? Does the whole elaborate system unravel?

I don't think so for one simple reason: there aren't that many ties to break. There is no need to resort to lower levels of analysis in many cases. Not every detailed doctrinal case is hard, so it is not always necessary to invoke midlevel principles. And even where doctrine or operational rules leave room for doubt, requiring resort to midlevel principles, this does not inevitably mean a head-on clash of principles that in turn necessitates resort to foundations for resolution. There are ample resources *within* each of the first two levels with which to resolve many disputed issues, and this eliminates the need in many cases to go beyond each level for answers.

It would require a great deal of exposition to make a full-fledged defense of this position. Given my space limitations here, let me limit my comments to one (necessarily inadequate) example. The first concerns issues that are of workaday interest in copyright and patent law: doctrines of authorship and patent validity. Often these may be applied in a straightforward way to obtain a predictable answer. The logic of the legal requirements is embodied in a number of technical doctrines and rules, but the basic pattern is easy enough to describe. Once the facts surrounding the origin of the work in question are ascertained, the relevant doctrines will become apparent. For authorship, it may be necessary to know what two or more alleged coauthors did and what they contributed to a given work. For patent validity, the nature of the invention and the most important pieces of prior art must be ascertained. But in any event, once the facts are known, a court or lawyer looking at them can often reach a fully defensible judgment about the proper legal outcome.

This does not mean that midlevel principles are not at all involved. Quite the contrary. These principles are derived from, and supply the guiding force behind, the IP doctrines that apply in these cases. The principles are *embedded into* the doctrines. So, for example, questions of coauthorship bring together two midlevel principles. First, efficiency is central, in that the high transaction costs that follow from a fragmented ownership structure for a single work covered by copyright militates in favor of a significant minimum threshold requirement for the status of an author. Second, dignity comes into play, in that fairness seems to dictate that one who makes a sizeable contribution to a copyrighted work ought to share in the ownership rights.

The point here is that these midlevel principles will seldom have to be explicitly invoked in a run-of-the-mill question of coauthorship status. The specific lines of cases and subdoctrines that make up the details of this area of copyright already embody the relevant principles. Although there may be controversy about the facts of a case, once it is identified as a coauthorship dispute the midlevel principles will seldom need to be explicitly invoked. They are latent within the cases and doctrines that will be deployed to resolve the issue. Once in a while, a new type of case might arise, or an important court might decide to revisit or revise an old established doctrine. In such a case, one or more of the midlevel principles may be discussed more explicitly than usual. But in the main this is an unusual event, and overall the threat of persistent unraveling from level three to level two does not appear very great.

V. CONCLUSION

The analysis in *JIP* was meant to show the benefits of three distinct levels of analysis in IP law. I took pains in that book to explain how each level differed from the other two, and to emphasize the distinct functions of the separate levels. In this brief Article, I have tried to address the relationship between the three levels—the points of contact and occasions when two adjoining levels interact with each other. I still believe in the benefits of this tripartite division. But I hope I have also shown that there is, and should be, something less than complete acoustic separation between them. That at any rate has been my goal.