

2017

Inverted Theories

Lee Anne Fennell

Richard H. McAdams

Follow this and additional works at: [https://chicagounbound.uchicago.edu/
public_law_and_legal_theory](https://chicagounbound.uchicago.edu/public_law_and_legal_theory)

 Part of the [Law Commons](#)

Chicago Unbound includes both works in progress and final versions of articles. Please be aware that a more recent version of this article may be available on Chicago Unbound, SSRN or elsewhere.

Recommended Citation

Lee Anne Fennell & Richard H. McAdams, "Inverted Theories," Working Paper(2017).

This Working Paper is brought to you for free and open access by the Working Papers at Chicago Unbound. It has been accepted for inclusion in Public Law and Legal Theory Working Papers by an authorized administrator of Chicago Unbound. For more information, please contact unbound@law.uchicago.edu.

INVERTED THEORIES

*Lee Anne Fennell** and *Richard H. McAdams†*

Simple theoretical models play a large role in the legal academy. They typically embed strong assumptions, which double as qualifications. When the assumptions are thought to be roughly accurate, or inaccurate in trivial or irrelevant ways, the theory’s primary claim takes center stage and the qualifications may be downplayed or even ignored. But what if a critical assumption is widely known to be patently false or extremely implausible?¹ Here, we see several responses. A surprisingly popular one is to ignore the assumption’s falsity altogether and embrace the theory anyway. Another robust approach is to reject the theory as being built on a foundation of sand, and refuse to have anything more to do with it. We would like to draw attention in this essay to a third alternative: using a theory’s radically unrealistic assumptions to “invert” it. This approach allows scholars to draw lessons from the theory—indeed, sometimes the very ones that the theorist originally had in mind—by turning the spotlight on the implications of the untrue assumptions.

A well-known example of inversion involves the Coase Theorem, which in its popular formulation holds that if transaction costs are zero, an efficient result will always be reached regardless of the initial allocation of entitlements.² The zero transaction cost assumption is, of course, wildly unrealistic—a fact Coase emphasized from the outset.³ A different and better way to articulate the Coase Theorem is to invert it: Because transaction costs are positive, the initial allocation of entitlements *can* matter to efficiency.⁴ This rearticulation puts the emphasis where Coase

* Max Pam Professor of Law and Ronald H. Coase Research Scholar, University of Chicago Law School. I am grateful for research support from the Harold J. Green Faculty Fund and the SNR Denton Fund.

† Bernard D. Meltzer Professor of Law, University of Chicago Law School. For conversations about this piece, we thank Anupam Chander, Jerry Frug, Richard Schragger, and participants in the Conference at Harvard Law School Celebrating Jerry Frug’s Work on Cities. We also thank Reeves Jordan for excellent research assistance.

¹ See DANI RODRIK, *ECONOMICS RULES: THE RIGHTS AND WRONGS OF THE DISMAL SCIENCE* 27 (2015) (explaining that “an assumption is critical if its modification in an arguably more realistic direction would produce a substantive difference in the conclusion presented by the model”).

² R.H. Coase, *The Problem of Social Cost*, 3 J. L. & ECON. 1 (1960). George Stigler named the Theorem, not Coase himself. G.J. STIGLER, *THE THEORY OF PRICE* 113 (3d ed. 1966).

³ Coase, *supra* note 2, at 15 (“The argument has proceeded up to this point on the assumption . . . that there were no costs involved in carrying out market transactions. This is, of course, a very unrealistic assumption.”).

⁴ For example, Mitch Polinsky defines “the more complicated version of the Coase Theorem” as follows: “If there are positive transaction costs, the efficient outcome may not occur under every legal rule.” A. Mitchell Polinsky, *The Coase Theorem*, in *AN INTRODUCTION TO LAW AND ECONOMICS* 15 (4th ed. 2011). See also Deirdre McCloskey, *The So-Called Coase Theorem*, 24 EASTERN ECON. J. 367, 367 (1998) (“Economists have gotten the ‘theorem’ wrong; in fact backwards.”); Stephen G. Medema, *The Myth of Two Coases: What Coase Is Really Saying*, 28 J. ECON. ISSUES 208, 213 (1994) (stressing Coase’s focus on transaction costs and institutional arrangements and observing that “[t]he ‘popular Coase’ is a fiction.” Because the need to “invert” (as we term it)

himself did. Although the popular or “uninverted” form of the Coase Theorem still receives a lot of play, law and economics scholars seem well attuned to the significance of the zero transaction cost qualification.⁵ Of course, it is in their interest to heed the qualification: if law can never matter to efficiency, their life’s work is pointless.

In this essay, we consider four other theoretical models that are good candidates for inversion: Nozick’s entitlement theory of distributive justice, the Tiebout Hypothesis, Kaplow and Shavell’s principle of tax superiority, and the Prisoners’ Dilemma. In the first three cases, the original theorists recognized the dependence of their theories’ main takeaways on strong assumptions that did not align with reality. In the case of the Prisoners’ Dilemma, the specific conditions that produced the dominant equilibrium and associated normative punchline were baked into the game’s structure. Later users of these ideas as inputs into policy prescriptions have, however, tended to downplay the significance of the qualifications. In all of these cases, we submit, the better basis for policy is the inverted version of the theory that puts the qualifications at center stage.

Theories well-suited for inversion should not be taken at face value, we argue, but they should not be discarded either. Rather, they can usefully focus attention on the conditions that render their supporting assumptions false, and thereby push scholarly attention in productive directions. The logical parameters of the original theory offer a structure for the inverted analysis, which differs from a general critique or rejection of the theory.

The discussion below proceeds in three steps. Part I outlines the conditions that make a theoretical construct ripe for inversion, using the Coase Theorem to illustrate. Part II focuses on our four additional examples—theories that we maintain are more useful and powerful when inverted. Part III distinguishes inversion from critiques and warnings, and explores some of the concerns that might surround the inversion exercise.

I. CONDITIONS FOR INVERSION

There are three basic ingredients that make a theoretical model particularly suitable for inversion. First, it must contain strong and unrealistic critical assumptions. Second, it must have generated a widely-

the popular articulation of the Coase Theorem is well-known, we use the example to introduce the concept, but focus below on theories where the needed inversion is less understood.

⁵ But perhaps this development is relatively recent, at least in some quarters. See McCloskey, *supra* note 4, at 368 (“Something like a dozen people in the world understand that the ‘Coase’ theorem is not the Coase theorem. One of this select group is Ronald Coase himself...”); Robert C. Ellickson, *The Case for Coase and Against “Coaseanism,”* 99 YALE L.J. 611 (1989) (observing that in an article by John Donohue, “which in this regard is representative of law-and-economics writing in general, Coase’s name is consistently attached to propositions that he has explicitly repudiated”) (footnote omitted); Medema, *supra* note 4, at 210 (observing that there are “a legion of people who have misinterpreted (or failed to read) Coase’s message”).

shared popular understanding with a simple normative prescription that hinges (wittingly or not) on the truth of the unrealistic assumptions. Third, confronting the falsity of the assumptions must upend this normative prescription, yet identify a productive path for further normative inquiry. Figure 1 summarizes these core ingredients for the Coase Theorem:

Figure 1: *The Coase Theorem as Inverted Theory*

Theory	Popular Understanding	Unrealistic Assumptions	Inverted Implications
THE COASE THEOREM	Bargaining Produces Efficiency Regardless of Legal Entitlements	Zero Transaction Costs	Attend to Entitlements & Transaction Costs

A brief overview of each of these components will help to set the stage for the additional examples examined in Part II.

A. *Unrealistic Assumptions*

For a theory to gain more traction when inverted than when taken at face value, it must have at its core one or more unrealistic assumptions. The clearest type of unrealistic assumption is one that is categorically false, and universally understood to be so. Zero transaction costs is a paradigmatic example. Of course, even categorically false assumptions must be relevantly and significantly false, not just trivially or technically false, in order to support inversion; otherwise, the assumption could stand as a reasonable and convenient approximation of the truth. Again, the zero transaction cost assumption fits the bill. Although there are some circumstances in which costs are closer to zero than others, transaction costs are rarely low enough to be irrelevant to the analysis, even if they are capable of being overcome.

Some unrealistic assumptions are contingently rather than categorically false. *Any* assumption will be false in some circumstances, so contingent falsity is insufficient on its own to make a theory amenable to inversion. Rather, the contingently false assumption must be one that is overwhelmingly false in the setting or settings in which it is invoked. Sometimes a cluster of assumptions is conjunctively necessary to support a theory, none of which is especially unlikely on its own, but the combination of which is so rare as to make the assumption-cluster contingently false in an overwhelming proportion of cases. This, we will argue, is the case with the Prisoners' Dilemma, where the full set of required structural features rarely if ever appear in combination.

As consensus about the falsity of the relevant assumptions starts to unravel, the case for inversion weakens. At the limit, we reach cases where

the truth or falsity of a given assumption is hotly contested, and it is well understood that the prescription will invert depending on which view prevails. Consider in this context the Laffer Curve,⁶ which purports to depict the relationship between the tax rate and the resulting government revenue as an inverted U shape. The normative lesson of the Curve depends on a contested fact—where along the curve one places our current tax regime. If we are on the downward slope, beyond the optimal point, then increasing taxes will decrease revenues, but if we are not yet at the optimal point, increasing taxes will increase revenue.⁷ The Laffer Curve meme is fascinating in its own right,⁸ but because academic debate is focused on the empirical question of where we are on the Curve, it does not benefit from inversion as discussed here. Rather, the theory already embeds inversion in its structure.

B. Popular Understandings

A theory grounded in false assumptions becomes an attractive target for inversion when it has taken on a popular intellectual life of its own that both ignores and depends upon those assumptions. This requires, at a minimum, that the theory's popularized version be well-known and frequently used. The best candidates for inversion will be the scholarly equivalent of household names. They typically attain that status by featuring clear and counterintuitive normative prescriptions of broad significance. This, we believe, is true of all of the examples we discuss below.

Popular understandings are most likely to develop around workhorse concepts that legal scholars use as basic building blocks for other arguments or as shared assumptions in conversations. These are the set pieces that are trotted out in first-year classes and referenced innumerable times thereafter as students progress through their legal education. Faculty mention them regularly and reflexively in workshops and conversations, in hallways and over lunch. They are in the air and under our skin.

This very popularity produces much of the pressure toward sticking to the theory's canonical, popularized understanding, as a kind of shared language. Incantations like "Coasean bargaining" or "Tieboutian sorting" can instantly bridge a gap in an argument, succinctly frame a critique, or make a student's eyes light up with recognition. In a conversation focused on some specific legal topic, it would seem pedantic and perhaps even uncollegial to point out that the theory's underpinning assumptions are

⁶ See Don Fullerton, *Laffer Curve*, in THE NEW PALGRAVE DICTIONARY OF ECONOMICS 839 (Steven N. Durlauf, and Lawrence E. Blume, eds., 2nd ed. 2008); Mathias Traband and Harald Uhlig, *The Laffer Curve Revisited*, 58 J. MONETARY ECON. 305 (2011).

⁷ The importance of this distinction is understood by academic economists, even if some politicians miss it.

⁸ See, e.g., Robert J. Shiller, *Narrative Economics*, 107 AM. ECON. REV. 967, 979-82 (2017).

utterly false, and the lesson we should take from the theory is a different one entirely from the one being referenced. One goal of this essay is to lower the costs of invoking inverted understandings of key theories in such discussions.

C. Upending Prescriptions

Inversion is only useful if it substantially changes the analysis surrounding a theory, but leaves something of value. Taking seriously the falsity of the theory's supporting assumptions does not require that the normative prescription always or completely reverse. Rather, what is frequently upended is the categorical version of the prescriptive lesson. Attention is then focused on the untrue assumptions themselves, and the analytic leverage that can be gained from thinking carefully about the significance and implications of their falsity.

The analytic move that we have in mind tracks the creation of an inverse in logic from an if-then statement.⁹ For the Coase Theorem, we start with this statement:

If transaction costs are zero, an efficient result will always be reached regardless of the initial allocation of entitlements. [And therefore legal interventions are unnecessary to achieve efficiency].

The popularized version of the Coase Theorem downplayed the “if” clause and instead suggested that the balance of the sentence would hold true across a range of real-world conditions. We can reposition emphasis on the untrue assumption by forming this inverse:

*If transaction costs are **not** zero, an efficient result will **not** always be reached regardless of the initial allocation of entitlements. [And therefore the allocation of entitlements can matter to efficiency].*

It is worth drawing attention to the work that is done by the *always* in the statements above. If we were to omit the word *always* from the original statement, the inverse would be:

If transaction costs are not zero, an efficient result will not

⁹ See ALFRED TARSKI, INTRODUCTION TO LOGIC AND TO THE METHODOLOGY OF THE DEDUCTIVE SCIENCES 40 (Jan Tarski, ed. 1994) (4th ed.) (explaining that the inverse of the statement “if *p*, then *q*” is “if not *p*, then not *q*”).

be reached regardless of the initial allocation of entitlements.

That statement is obviously untrue, at least if we think that efficient results are reached in some cases in the real world, where transaction costs are always nonzero. Adding the “always” is an appropriate way to carry out the inversion, because it is the categorical version of the Coase Theorem that carries force and currency.

A related but distinct logical move can be used to counter a different sort of categorical claim: that a specified result will follow “if and only if” a particular condition is present—or, in other words, that the condition is both necessary and sufficient to produce the result.¹⁰ Consider this claim about property arrangements:

If (and only if) property is held in common (rather than in private ownership), a resource tragedy results.

The “if and only if” specification means that the converse of the original conditional statement is also true—an equivalence.¹¹ The converse looks like this:

If a resource tragedy results, property is held in common (rather than in private ownership).

Frank Michelman (and later Michael Heller) disproved this converse through the notion of the anticommons,¹² and with it any idea that the commons *uniquely* produces resource tragedies, or that private property ownership is sufficient to avert such tragedies. Debunking such claims is interesting and important, but it is not inversion as we use the term here.

Also distinguishable are efforts to leverage the logic of symmetry by identifying opposite but similar possibilities that theorists or jurists have overlooked. There are many examples of this approach, but the most famous is Guido Calabresi and Douglas Melamed’s Rule 4, which observed

¹⁰ See *id.* at 29-30 (Jan Tarski, ed. 1994) (4th ed.) (explaining that “if and only if” equates to a claim that the specified condition is necessary and sufficient).

¹¹ *Id.* at 29.

¹² The anticommons idea originated in Frank Michelman’s conception of a regulatory regime that would be the “converse” of a commons. See Frank I. Michelman, *Ethics, Economics, and the Law of Property*, in NOMOS XXIV: ETHICS, ECONOMICS, AND THE LAW 3, 6, 9 (J. Roland Pennock & John W. Chapman eds., 1982); Frank Michelman, Remarks at Property Panel, Association of American Law Schools: Is the Tragedy of The Common Inevitable? 6-7 (Jan. 1985) (on file with author). Michelman showed that the potential for resource tragedies is not dependent on the form of property holding, but on the difficulty in assembling cooperation (whether in the form of forbearance in a commons, or in the form of permission in an anticommons). Michael Heller later built on Michelman’s idea of the anticommons. See, e.g., MICHAEL HELLER, *THE GRIDLOCK ECONOMY* (2008); Michael A. Heller, *The Tragedy of the Anticommons: Property in the Transition from Marx to Markets*, 111 HARV. L. REV. 621 (1998).

that just as a polluter might pay a victim for the right to keep polluting, so might a victim pay an injurer to *stop* polluting.¹³ That work is also important and interesting, but it's not inversion within our meaning here.¹⁴

II. THEORIES RIPE FOR INVERSION

There are no doubt numerous theories that are good candidates for inversion, and we hope this essay will prompt scholars to identify more of them. We will focus here on four examples we are familiar with from our past work: Nozick's entitlement theory of distributive justice, the Tiebout Hypothesis, Kaplow & Shavell's theory of tax superiority, and the Prisoners' Dilemma. After briefly introducing each of these theories, we will set out its popular understanding, identify the unrealistic assumptions on which that understanding relies, and discuss the inverted version of the theory.

Figure 2 previews what is to come by placing these four theories within the structure established above for the Coase Theorem.

Figure 2: A Summary of Selected Inverted Theories

Theory	Popular Understanding	Unrealistic Assumptions	Inverted Implications
THE COASE THEOREM	Bargaining Produces Efficiency Regardless of Legal Entitlements	Zero Transaction Costs	Attend to Entitlements & Transaction Costs
NOZICK'S THEORY OF DISTRIBUTION	Vast Inequality Is Consistent With Distributive Justice	Past Justice in Acquisition and Transfer (Whole Chain of Title)	Attend to Distributive Patterns Where Case-By-Case Rectification Is Impossible; Tie Distribution to Reasons
THE TIEBOUT HYPOTHESIS	Let People Sort Into their Preferred Communities	Equal Mobility; Choices are Not Blocked or Interdependent	Attend to Spatial Interdependencies and Choice Constraints
KAPLOW & SHAVELL'S THEORY OF TAX SUPERIORITY	Ignore Distributive Consequences of (Non-Tax) Legal Rules	Political Costs Work the Same Way Regardless of How You Redistribute	Attend to Politics; Consider All Redistributive Routes
THE PRISONERS' DILEMMA	Legal Intervention Is Necessary and Improves Payoffs for Everyone	Payoff Structure Makes Mutual Defection Dominant But Mutual Cooperation Yields Higher Payoffs for Both Parties; Game Is One-Shot	Attend to Distributive Implications of Collective Action Problems; Consider Conditions That Enable Cooperation

¹³ Guido Calabresi & A. Douglas Melamed, *Property Rules, Liability Rules, and Inalienability: One View of the Cathedral*, 85 HARV. L. REV. 1089, 1116-23 (1972).

¹⁴ Nor is the fascinating Chapter 6, "Would Flipping It Work? Trying Things the Other Way Around," in IAN AYRES AND BARRY NALEBUFF, *WHY NOT?: HOW TO USE EVERYDAY INGENUITY TO SOLVE PROBLEMS BIG AND SMALL* 115-34 (2006), which is a looser notion of generating ideas by reversing the usual order of things.

A. Nozick's Theory of Distribution

Robert Nozick's 1974 book, *Anarchy, State, and Utopia*, offered a counterpoint to John Rawls's enormously influential *A Theory of Justice* (1972). In it, Nozick worked out a theory of a minimal or "nightwatchman" state that would largely limit its interventions to protecting citizens against force and fraud, so as to facilitate consensual market transactions among private parties. Significantly, such a state would stay out of the business of redistribution.

Nozick's defense of this hands-off approach to distribution is rooted in his "entitlement theory" of distributive justice.¹⁵ This theory does not base the legitimacy of a society's distribution on the patterns of resources that people end up with, but rather on the processes through which people come to hold those resources.¹⁶ Nozick detailed the conditions that would yield a fair distribution under his theory, which he recognized were not met in practice.¹⁷ His philosophical exercise was instead directed at establishing what he saw as the correct benchmark from which to evaluate distribution.

1. Popular Understanding

Both critics and fans of Nozick's work viewed it as an apology for free-market distributive results and as a basis for ending redistributive programs—or at least those programs that redistributed to the poor.¹⁸ Nozick's claim that a properly understood theory of distributive justice eschewed any particular pattern was viewed as a defense of existing inequalities. Although relatively few took Nozick's full-strength "nightwatchman" idea to heart, his libertarian arguments were marshalled in favor of more limited attacks on "the welfare state."

2. Unrealistic Assumptions

Nozick's presentation of his entitlement theory begins with the following "inductive definition":

1. A person who acquires a holding in accordance with the principle of justice in acquisition is

¹⁵ ROBERT NOZICK, ANARCHY, STATE, AND UTOPIA 150-82 (1974).

¹⁶ *Id.* at 153-60. See Hal R. Varian, *Distributive Justice, Welfare Economics, and the Theory of Fairness*, 4 PHIL. & PUB. AFF. 223, 224 (1975) ("Nozick's theory is a procedural theory; the justice of a distribution is entirely dependent on the path used to reach it.").

¹⁷ NOZICK, *supra* note 15, at 150-53.

¹⁸ See Anupam Chander & Madhavi Sunder, *Is Nozick Kicking Rawls's Ass? Intellectual Property and Social Justice*, 40 U.C. DAVIS L. REV. 563 (2007) ("Robert Nozick stands as one of the foremost intellectual antagonists to claims for distributive justice.").

- entitled to that holding.
2. A person who acquires a holding in accordance with the principle of justice in transfer, from someone else entitled to the holding, is entitled to the holding.
 3. No one is entitled to a holding except by (repeated) applications of 1 and 2.¹⁹

Although Nozick suggests that these three rules would cover the distributive waterfront “[i]f the world were wholly just,”²⁰ he also recognizes that “[n]ot all actual situations are generated in accordance with the two principles of justice in holdings: the principle of justice in acquisition and the principle of justice in transfer.”²¹ This fact requires him to introduce into his entitlement theory the notion of “rectification of injustice in holdings.”²²

Thus, Nozick’s entitlement theory only endorses a pure market distribution conditional on the assumption that everyone’s holdings came about from combinations of just acquisition and just transfer or that there has already been proper rectification for any injustice in holdings. Any deviations from those principles would require an intervention to bring about the results that would have obtained had those principles been followed—assuming it is possible to determine what that distribution would have looked like. Of course, the principles of justice in acquisition and transfer *have been* violated in dramatic and systematic ways. In the United States, for example, today’s pattern of property holdings are affected by a history that includes settlement-by-conquest; chattel slavery; property and contract restrictions on women; the acquisition of family fortunes by fraud, bribery, and other criminality; government corruption and discrimination; and private discrimination. Consequently, it is all but unimaginable that an American’s current holding of, say, land is the result of a chain of entirely just acquisitions and just transfers.

By hypothesis, Nozick’s theory does not support the existing distribution where the principles of justice in acquisition and transfer have not been met, unless rectification has already set things right. Yet Nozick does not argue that *any* existing society satisfies the principles of just acquisition and transfer or that it has, through rectification, cured its injustices. Instead, his principle of rectification, if taken literally, would require uncovering past injustices and rectifying each one individually, as by a transfer payment from the current beneficiary of the injustice to its

¹⁹ *Id.* at 151.

²⁰ *Id.*

²¹ *Id.* at 152.

²² *Id.*

victim.

Recognizing that it may be impossible to apply the principle of rectification to perfectly offset all past injustices in acquisition and transfer, Nozick observes that “[p]erhaps it is best to view some patterned principles of distributive justice as rough rules of thumb meant to approximate the general results of applying the principle of rectification of injustice.”²³ And, interestingly, he suggests that if certain conditions hold, “a *rough* rule of thumb for rectifying injustices might seem to be the following: organize society so as to maximize the position of whatever group ends up least well-off in society.”²⁴ In other words, Nozick suggests it is possible (though perhaps implausible) to derive a Rawlsian theory of justice from his entitlement theory.²⁵ His larger point is that, if the principle of rectification cannot feasibly be carried out through individual case-by-case remediation, it may instead demand some kind of broad social program of redistribution.

Nozick goes on to add that, at least without a well-developed principle of rectification tailored to specific facts of one society, “one *cannot* use the analysis and theory presented here to condemn any particular scheme of transfer payments, unless it is clear that no considerations of rectification of injustice could apply to justify it.”²⁶ Thus, the assumptions that would be necessary in order for his entitlement theory to support a non-redistributive state are, even in Nozick’s own view, so patently false as to make it impossible to use the theory to criticize any actual instances of redistribution. But of course the theory, in its popular incarnation, has been used in just this way.²⁷ This (mis)use may be in part due to Nozick’s expositional choices and the way in which he allocated emphasis. As Hal Varian explains:

The impression one gets from reading Nozick is that the problem of rectification is somehow minor. It seems to me that the reverse is the case: the problem of rectification is

²³ NOZICK, *supra* note 15, at 230-31.

²⁴ *Id.* at 231 (emphasis in original). The conditions Nozick has in mind are that, “lacking much historical information,” we make the following two assumptions: “(1) that victims of injustice generally do worse than they otherwise would and (2) that those from the least well-off group in the society have the highest probabilities of being the (descendants of) victims of the most serious injustice who are owed compensation by those who benefited from the injustices (assumed to be those better off, though sometimes perpetrators will be others in the worst-off group).” *Id.*

²⁵ *Id.* at 231 (noting that “[t]his particular example may well be implausible”). See JOHN RAWLS, A THEORY OF JUSTICE 75-80 (1971) (describing the “difference principle” under which gains to society are only just if they work to the benefit of the least advantaged); *id.* at 152-57 (describing the “maximin rule”).

²⁶ NOZICK, *supra* note 15, at 231.

²⁷ See, e.g., Julian Lamont and Christi Favor, “Distributive Justice”, The Stanford Encyclopedia of Philosophy (Winter 2016 Edition), Edward N. Zalta (ed.), <https://plato.stanford.edu/archives/win2016/entries/justice-distributive/> (observing that “[a]lthough Nozick is fairly candid about” his theory’s inability to speak to current distributive questions, “many of his supporters and critics have ignored it and have carried on a vigorous debate as though, contrary to Nozick’s own statement, his theory can be used to evaluate the justice of current economic distributions.”).

central to the issue of justice. We are interested in the question of justice precisely because we live in an unjust world; injustices have occurred in the past and are occurring now. The question is what we should do about them.²⁸

Some possible answers to that question emerge if we invert Nozick's theory.

3. Inverting Nozick's Entitlement Theory

Suppose we accept Nozick's idea that the justice of a distribution depends on its history, but reject the claim that our particular history gives people morally justifiable entitlements over the specific holdings that they currently have. The principle of rectification would, as we have already seen, flip the advice about the state's role in addressing distribution. Instead of "if all past holdings and transfers were just (or have been fully rectified where not just), then the current distribution is just," the proper lesson is "because not all past holdings and transfers were just (and the injustices have not been fully rectified), the current distribution is not just." Flipping the theory carries both backward-looking and forward-looking implications. Looking back at the history of injustice in acquisition and transfer (unaccompanied by anything like full rectification) establishes that current holdings are not just. Although unwinding each specific injustice is impossible, systematic past injustices that are reflected in current distributive patterns would, on this account, call for distributive interventions that are motivated not by a desire to create a more pleasing distributive pattern, but rather by the imperative to rectify—in at least a rough way—the past injustices that the current patterns reflect.

Flipping the theory, rather than simply rejecting it outright, also carries forward-looking lessons about distribution.²⁹ At the broadest level, Nozick's theory suggests that it matters *how* people came to possess the things they have. This lesson resonates with scholarship suggesting that legal outcomes cannot be evaluated independent of the processes that brought them about.³⁰ Significantly, some processes of distribution draw tighter connections between desert and payoffs than others. If the history of

²⁸ Varian, *supra* note 16, at 227.

²⁹ These lessons turn out to be sharply at odds with the standard law and economics prescription. See *infra* Part II.B. (discussing Kaplow and Shavell's theory of tax superiority).

³⁰ See, e.g., Daphna Lewinsohn-Zamir, *In Defense of Redistribution through Private Law*, 91 MINN. L. REV. 326, 357–60 (2006) (arguing that gains and losses are evaluated based not only on their magnitudes but also on the process by which they came about); Bruno S. Frey, Matthias Benz, & Alois Stutzer, *Introducing Procedural Utility: Not Only What, But Also How Matters*, 160 J. INSTIT. AND THEOR. ECON. 377 (2004) (presenting evidence that people care about the processes that bring about outcomes, and not just the outcomes themselves); see also Daphna Lewinsohn-Zamir, *Taking Outcomes Seriously*, Utah L. Rev. 861 (urging a broader definition of "outcomes" that incorporates elements like the processes that produced them).

holdings bears on the justice of holdings, as Nozick suggests, then distributive policy should be sensitive not only to creating distributive patterns but also to building distributive histories that link payments with rationales.

On this view, a grant that is given for no reason other than to produce a particular income pattern (e.g., to decrease the Gini coefficient)³¹ is normatively different from a grant that is given to make up for a distributive shortfall produced by government action—or inaction. Government makes innumerable decisions, from choosing law enforcement priorities to deciding where and how to build dams, highways, or landfills, that have distributive consequences. Making up for those effects through concurrent distributive efforts (whether in-kind remediations, social insurance, or direct payments to the affected populations) maintains the link between the rationale and the payment in a way that disconnected welfare payments made at some other time would not. To put it a bit differently, certain ways of arranging social policy can clarify or cloud the historical “title” that underpins claims of entitlement on Nozick’s theory. If one subscribes to a theory of justice that depends on accurately tracing title, then one should be willing to invest in making that task easier to accomplish.

B. *The Tiebout Hypothesis*

Charles Tiebout wrote *A Pure Theory of Local Expenditures* as a rejoinder to the then-conventional academic view, articulated by Richard Musgrave and Paul Samuelson, that governmental provision of goods and services is inherently inefficient due to intractable problems of demand revelation.³² Musgrave and Samuelson maintained that people will misrepresent their demand for public goods and services in hopes of free-riding on others, given the lack of any price mechanism to force them to reveal their preferences. Tiebout recognized that in a multi-jurisdictional metropolitan area, “consumer-voters” can select among local governments. If each local government offers a different mix of goods and services, along

³¹ The Gini coefficient is a standard measure of income inequality in a society. Given a curve that depicts the share of income earned by each income percentile (the Lorenz Curve), the Gini coefficient describes the degree by which that curve diverges from a perfectly proportionate income dispersion. *See, e.g.,* LEE S. FRIEDMAN, *THE MICROECONOMICS OF PUBLIC POLICY ANALYSIS* 126–28, fig. 5-1 (2002). Zero represents perfect equality and one represents perfect inequality (where one person receives all national income).

³² Charles M. Tiebout, *A Pure Theory of Local Expenditures*, 64 *J. POLIT. ECON.* 416 (1956); *see also* Wallace E. Oates, *The Many Faces of the Tiebout Model*, in *THE TIEBOUT MODEL AT FIFTY: ESSAYS IN PUBLIC ECONOMICS IN HONOR OF WALLACE OATES* 21, 21-22 (William A. Fischel, ed., 2006) (“Tiebout’s purpose in the paper is to show that, in principle at least, there exists a mechanism or solution to the free-rider problem for a specific class of public goods—namely, local public goods.”). Oates goes on to argue that Tiebout meant to offer more “than just a formulation of a preference-revealing algorithm” insofar as “he does at various points try to justify and explain his model in terms of ‘real’ behavior.” *Id.* at 22.

with associated tax burdens, the person who is choosing where to live can be analogized to a shopper who is selecting among different baskets and price points—if certain strong and unrealistic assumptions hold.

The theoretical point, which came to be known as the Tiebout Hypothesis, was a powerful one: it showed how entry and exit could substitute for a price mechanism and reveal information about preferences as consumer-voters “voted with their feet.”³³ Over the last six decades, the idea has become a mainstay of local government analysis. However, at least in legal academia, its primary thrust has morphed from an analytic point about methods of demand revelation to a normative point about local autonomy and consumer choice.

1. Popular Understanding

The Tiebout Hypothesis (TH) is broadly associated with the positive claim that people sort into the communities that suit them best, and that the communities shape themselves to compete for consumer-voters. This corresponds to a normative claim: that local governments should have autonomy, so as to induce optimal sorting.³⁴ Although typically addressed to local government autonomy at the municipal level, it has also been cited in connection with smaller subunits such as private communities.³⁵

The normative use of TH takes on particular significance in the context of local land use controls, where it is often invoked.³⁶ The claim might simply be that people will wish to sort themselves among communities based on their preferences for particular land use policies (which can affect density, architecture, city layout, and so on), just like they might wish to

³³ Wallace Oates greatly raised the visibility of Tiebout’s work among economists in 1969 when he published an empirical study testing the degree of capitalization of tax-service packages into home prices. Wallace E. Oates, *The Effects of Property Taxes and Public Spending on Property Values: An Empirical Study of Tax Capitalization and the Tiebout Hypothesis*, 77 J. POLIT. ECON. 957 (1969); see William A. Fischel, *Footloose at Fifty: An Introduction to the Tiebout Anniversary Essays*, in THE TIEBOUT MODEL AT FIFTY, *supra* note 32, at 1, 5 (describing the effects of Oates’s 1969 publication).

³⁴ See, e.g., Richard Briffault, *The Rise of Sublocal Structures in Urban Governance*, 82 Minn. L. Rev. 503 (1997) (“The dominant law and economics model of local government, based on the work of Charles M. Tiebout, assumes that decentralization of power to local governments promotes the efficient delivery of public goods and services.”); Nestor M. Davidson & Sheila R. Foster, *The Mobility Case for Regionalism*, 47 U.C. DAVIS L. REV. 63, 73 (2013) (“[S]cholars regularly invoke the Tiebout model to support arguments for devolution and decentralization.”). These normative claims attached to the Tiebout Hypothesis can be distinguished from the writings of Tiebout himself. See Davidson & Foster, *supra*, at 73 n. 31 (“The proliferation of structural prescriptions associated with Tieboutian localism is akin to the outgrowth of a kind of normative Coaseanism that is arguably removed from Ronald Coase’s actual work—what Robert Ellickson contrasted as the cardboard Coase and the real Coase.”) (citing Ellickson, *supra* note 5, at 612-13).

³⁵ See, e.g., Robert C. Ellickson, *Cities and Homeowners Associations*, 130 U. PA. L. REV. 1519, 1547-48 (1982). Tieboutian logic has also been extended to the analysis of states within a federal system. See, e.g., Roderick M. Hills, Jr., *Compared to What? Tiebout and the Comparative Merits of Congress and the States in Constitutional Federalism*, in THE TIEBOUT MODEL AT FIFTY, *supra* note 32, at 239.

³⁶ See, e.g., William A. Fischel, *Footloose at Fifty: An Introduction to the Tiebout Anniversary Essays*, in THE TIEBOUT MODEL AT FIFTY, *supra* note 32, at 1, 18 (“Scholarly discussions of land use regulation do not go far before invoking the Tiebout model.”).

sort themselves based on any other local policy. But zoning plays a unique role in supporting the TH's shopping metaphor. Without zoning, in-movers could obtain the same local services at a lower tax price by adding cheaper homes;³⁷ zoning can enforce an equal property tax burden by requiring everyone to consume a minimum amount of housing.³⁸

Yet zoning does not just enforce equal tax burdens within a given community consume, it also heavily influences who can move into the community. For example, if the only types of housing stock that the municipality's zoning ordinance permits are single-family homes on sizable lots and luxury studio apartments, low-income people with children are locked out. The popular TH supports zoning that undermines mobility and free choice of some citizens, even though those same characteristics of mobility and choice lie at the heart of TH's popular appeal.

2. Unrealistic Assumptions

Tiebout developed his novel intuition that consumer-voters could shop for a local government "by postulating an extreme model" that made several explicit assumptions.³⁹ In addition to assuming perfect mobility and full knowledge on the part of consumer-voters, Tiebout's model assumed that employment played no role in anyone's locational choice; rather, he "assumed that all persons are living on dividend income."⁴⁰ And, perhaps most crucial and unrealistic of all, Tiebout's model assumed that "[t]he public services supplied exhibit no external economies or diseconomies between communities."⁴¹

As Tiebout recognized, the unrealistic nature of these assumptions placed limits on his theory's implications. For example, externalities produced by law enforcement in particular jurisdictions might call for at least some functions to be handled at a more centralized level. The mobility assumption is also sharply limited by, among other things, the real world demands of employment and the need for proximity to particular people or firms for other reasons.⁴² In addition, zoning has the effect of altering opportunities for mobility.

³⁷ See Bruce W. Hamilton, *Zoning and Property Taxation in a System of Local Governments*, 12 URB. STUD. 205 (1975) (explaining that Tiebout's idea of shopping for local services required some way of stabilizing the prices that residents and would-be residents confront through taxation—which zoning could provide).

³⁸ See *id.* Hamilton later refined his argument to explain that the level of housing consumption did not have to be uniform across the jurisdiction for a Tieboutian pricing system to work—it was only necessary that housing supply be fixed so that the tax bargain could be properly capitalized into the price of homes. Bruce W. Hamilton, *Capitalization of Intra-jurisdictional Differences in Local Tax Prices*, 66 AM. ECON. REV. 743 (1976); see also WILLIAM A. FISCHEL, *THE HOMEVOTER HYPOTHESIS* 68-69 (2001).

³⁹ Tiebout, *supra* note 32, at 419.

⁴⁰ *Id.*

⁴¹ *Id.*

⁴² See David Schleicher, *The City as a Law and Economic Subject*, 2010 U. ILL. L. REV. 1507 (noting tensions between sorting and agglomeration).

These limits did not, however, change Tiebout's theoretical point that if it were possible to move closer to the world of the assumptions, it would also be possible to improve the efficiency of government. Yet Tiebout's important observations about demand revelation lie far from the normative propositions generally associated with TH: that local governments should be given free rein in order to induce optimal sorting. Focusing on Tiebout's original point leads us instead to a different, inverted interpretation of TH.

3. Inverting the Tiebout Hypothesis

To begin, note that under Tiebout's model, sorting is a means to an end, not an end in itself: it is through the process of sorting into communities that people register their preferences for different goods and services. The quality of the information that is revealed through this process is only as good as the sorting itself can make it. Every barrier to mobility, every obfuscation of information about services and costs, and every externality that attenuates the connection between what is paid and what is received, makes location choices that much less revealing.

Restoring the emphasis on demand revelation flips TH from its usual orientation. The popular version of TH holds that each local government must remain free to set its own policies (including exclusionary land use policies) without restraint, so that people—people with perfect mobility, that is—can sort into their preferred communities. This gets things backwards. The flipped version of TH suggests that public policy should work to improve the conditions of mobility for everyone, increase awareness of the implications of various policies, and address interdependencies among communities so that the implicit price signals sent by moves are accurate ones. As Tiebout himself put it, "Policies that promote residential mobility and increase the knowledge of the consumer-voter will improve the allocation of government expenditures in the same sense that mobility among jobs and knowledge relevant to the location of industry and labor improve the allocation of private resources."⁴³

An inverted TH looks at ways to improve the shadow market of local government so that its performance with respect to public goods can approach the private market's performance with respect to private goods. What would this inverted perspective entail? For one thing, examining the ways in which local governmental policies impose costs on other local governments, as through exclusionary housing policies that increase education and social service burdens elsewhere, and environmental policies that place costs (such as those associated with landfills) elsewhere. But more foundationally, it would involve examining how all households can be

⁴³ Tiebout, *supra* note 32, at 423.

given meaningful choices among local jurisdictions, whether through enhanced voucher programs that open up more housing options in more places, policies to alter the amount and variety of housing stock across jurisdictions, or through other mechanisms that give low-income people realistic opportunities to self-sort based on their preferences.⁴⁴ Sorting based on what one likes and values *is* attractive, but the popular TH primarily promotes economic stratification among communities—something that cuts against rather than in favor of meaningful choice and mobility.

C. Kaplow & Shavell's Theory of Tax Superiority

Kaplow and Shavell, building on earlier economic treatments of commodity taxation, developed a formal model showing the advantages of conducting all redistribution through tax-and-transfer (hereinafter “tax”) rather than through legal rules.⁴⁵ The intuition behind their formal model relies on minimizing distortions to behavior. Moving money through the tax system is well known to produce distortions in the choice between labor and leisure, which had caused some legal scholars to suggest that using legal rules could be a preferable alternative.⁴⁶ Kaplow and Shavell argued that redistributive legal rules would embed an equivalent labor-leisure distortion, because rational actors would understand that, just as in the realm of taxation, having a higher or lower income would harm or benefit them.⁴⁷ However, legal rules would *also* distort choices about the primary behavior that is being regulated. For example, a tort rule that granted higher recoveries to low-income people and assessed higher damages against high-income people would distort choices about taking care.⁴⁸

Thus, under the Kaplow and Shavell model, moving a quantum of money through tax would cost less in terms of distortions than moving that

⁴⁴ Significantly, it is not actually necessary to have homogeneous housing stock (with similar property tax burdens) throughout a jurisdiction to determine how much people value local public goods and services. What is important is that the tax price be capitalized into the housing cost, which can occur even with heterogeneous housing stock, although it does require a constraint on supply to enable capitalization to occur. *See supra* note 38.

⁴⁵ *See* Steven Shavell, *A Note on Efficiency vs. Distributional Equity in Legal Rulemaking: Should Distributional Equity Matter Given Optimal Income Taxation?*, 71 AM. ECON. REV. 414 (1981); Louis Kaplow & Steven Shavell, *Why the Legal System Is Less Efficient Than the Income Tax in Redistributing Income*, 23 J. LEGAL STUD. 667 (1994) [hereinafter Kaplow & Shavell, *Less Efficient*]; Kaplow & Shavell, *Should Legal Rules Favor the Poor? Clarifying the Role of Legal Rules and the Income Tax in Redistributing Income*, 29 J. LEGAL STUD. 821, 825-34 (2000) [hereinafter Kaplow & Shavell, *Legal Rules*]; Louis Kaplow, *The Optimal Supply of Public Goods and the Distortionary Cost of Taxation*, 49 NAT'L TAX J. 513 (1996). A key antecedent in the economics literature was Anthony B. Atkinson and Joseph E. Stiglitz, *The Design of Tax Structure: Direct Versus Indirect Taxation*, 6 J. PUB. ECON. 55 (1976).

⁴⁶ *See, e.g.*, A. Mitchell Polinsky, *Resolving Nuisance Disputes: The Simple Economics of Injunctive and Damage Remedies*, 32 STAN. L. REV. 1075, 1084-85 (1980) (observing that “[d]ue to the substantial distortions in work effort, redistribution through the tax system would be quite costly in terms of efficiency”).

⁴⁷ *See, e.g.*, Kaplow & Shavell, *Less Efficient*, *supra* note 45, at 667-68.

⁴⁸ *See id.* *But see* Christine Jolls, *Behavioral Economic Analysis of Redistributive Legal Rules*, 51 Vand. L. REV. 1653 (1998) (disputing this equivalence based on cognitive biases).

same amount of money through the design of a legal rule. Thus, it is possible for everyone to be made better off by shifting all redistribution into the domain of tax.⁴⁹ Kaplow and Shavell accompanied this formal vision of tax superiority with policy advice that has become associated with the popular understanding of their work: that welfarists should ignore the distributive consequences of legal rules, except within the realm of tax.⁵⁰

1. Popular Understanding

Kaplow and Shavell's theory can be understood in (at least) two very different ways.⁵¹ One way focuses on the formal result—that any given increment of redistribution can be accomplished at lower cost (in terms of behavioral distortions) through tax than through legal rules. But this claim of “formal tax superiority” has been thoroughly overshadowed by what we call “prescriptive tax superiority.”⁵² This understanding of Kaplow and Shavell maintains that legal rules should never take distribution into account, but should instead be shaped entirely by efficiency considerations. Such an approach leaves all redistribution to the tax system, where (by hypothesis) it can be accomplished at lower cost.

Prescriptive tax superiority has become dominant in law and economics.⁵³ The idea that distribution is best left to tax has become so mainstream that it often arises in classrooms, workshops, and written scholarship without specific attribution. This categorical prescription might at first appear to be a logical and indeed inevitable extension of formal tax superiority. But on closer examination, the move between these two flavors of tax superiority depends on a crucial assumption—that it is *politically* no easier to achieve redistribution through one mechanism rather than another.

2. Unrealistic Assumption

Critics of Kaplow and Shavell's theory have noted that redistribution through tax may be politically infeasible, which might seem to argue for considering legal rules as an alternative redistributive mechanism. Few of these critics have focused on Kaplow and Shavell's response to this point, however. That response, which can be found in scattered observations in

⁴⁹ See, e.g., Kaplow & Shavell, *Less Efficient*, *supra* note 45

⁵⁰ *Id.* at 677.

⁵¹ These alternative interpretations are emphasized in Lee Anne Fennell & Richard H. McAdams, *The Distributive Deficit in Law and Economics*, 100 MINN. L. REV. 1051, 1058-68 (2016).

⁵² See *id.* at 1062-65.

⁵³ See, e.g., *id.* at 1062 n. 32 (collecting cites); Kyle Logue & Ronen Avraham, *Redistributing Optimally: Of Tax Rules, Legal Rules, and Insurance*, 56 TAX L. REV. 157, 158 (2003) (“[W]e believe it is a safe bet that a majority of legal economists hold the following view: Whatever amount of redistribution is deemed appropriate or desirable, the exclusive policy tool for redistributing to reduce income or wealth inequality should always be the tax-and-transfer system.”).

Kaplow and Shavell's solo and joint work, is that any political impediments to redistribution through the tax system will also block, to an equal extent, efforts to redistribute through legal rules. If a court or legislature does manage to produce a legal rule that alters distribution to a greater degree than could be achieved through tax, Kaplow and Shavell suggest that the victory for the winners will be short-lived: Congress will simply offset what has been done to regain its preferred distributive pattern.⁵⁴

In other words, Kaplow and Shavell suggest that no legal actor will be able to “get away with” trying to redistribute to a greater degree than Congress's own preferences allow. Therefore, one should redistribute through tax; if the distributive pattern is essentially inevitable, all that one can do is raise the cost of achieving it by trying to make use of a non-tax distributive method. We refer to this claim as one of “distributive invariance.”⁵⁵ Although Kaplow and Shavell do not set out distributive invariance as a formal assumption (which distinguishes this example from the prior two), it is nonetheless critical to their claim of prescriptive tax superiority.⁵⁶

Yet, as we once argued at length, the assumption is false.⁵⁷ It would be true only if the costs associated with the political resistance to redistribution—“political action costs”—were equal across all modes of redistribution. And, for many reasons, it may be more difficult politically to move money through the tax system than through a substantive legal rule. For example, legislative inertia and entrenchment may prevent Congress from changing the distributive pattern.⁵⁸

Significantly, the false assumption of distributive invariance underpins the shift from formal tax superiority to prescriptive tax superiority.⁵⁹ If one can get *different* distributive results by using legal rules instead of or in addition to tax, it would no longer be clear that tax is the exclusively preferred method. This is because in a welfarist analysis, both efficiency and distribution matter.⁶⁰ Whatever efficiency losses legal rules might produce through greater behavioral distortions might be more than made up for by their capacity to enhance distribution. It is therefore untrue that it is always best, regardless of one's distributive goals, to pursue those goals

⁵⁴ See, e.g., Kaplow & Shavell, *Less Efficient*, *supra* note 45 at 675.

⁵⁵ See generally Fennell & McAdams, *supra* note 51

⁵⁶ To some degree, they acknowledge as much. See Shavell, *supra* note 45, at 417; see also, Kaplow & Shavell, *Less Efficient*, *supra* note 45, at 675.

⁵⁷ See Fennell & McAdams, *supra* note 38. Circumstantial evidence for the falsity of invariance can be found in the tax system's pervasive non-responsiveness to changes in material inequality. See KENNETH SCHEVE AND DAVID STASAVAGE, *TAXING THE RICH: A HISTORY OF FISCAL FAIRNESS IN THE UNITED STATES AND EUROPE* (2016) (examining the history of twenty nations over two centuries and finding that rising inequality was virtually never sufficient to cause the government to raise taxes on the rich).

⁵⁸ Fennell & McAdams, *supra* note 51, at 1078-1108.

⁵⁹ *Id.* at 1069-71.

⁶⁰ Distribution matters both because material inequality adversely affects welfare in various ways and because welfare economics is open to the use of social welfare functions that value equality of welfare. See *id.*

through the tax system rather than through legal rules. In some cases, the opposite advice would be better.

3. Inverting the Theory of Tax Superiority

Suppose we approached the theory of tax superiority with an appreciation for the fact that political costs can vary among modes of distribution—that is with an understanding that the assumption of distributive invariance is false. Kaplow and Shavell’s theory of formal tax superiority would remain intact, showing that—in theory—there are always gains to be had by shifting more distributive work to the tax arena. But prescriptive tax superiority no longer holds. Instead of: “If the pattern of distribution is invariant to the means of redistribution, then it is always optimal to address distribution only through tax (and not through other legal rules),” we get: “Because the pattern of distribution varies with the means of redistribution, it is not always optimal to address distribution only through tax (rather than through other legal rules).”

Inversion invites several generative lines of inquiry. For starters, it highlights the importance of studying the political barriers to redistributing through different modes, similar to the focus on transaction costs that followed Coase’s theoretical breakthrough. As Nobel Laureate James Heckman has recognized, law and economics is “analytically incomplete” because it lacks “a satisfactory framework within which to analyze redistribution.”⁶¹ He elaborates: “A fully satisfactory analysis would require a careful accounting of the politics of redistribution and the gap between ideal policies and those that are actually used by governments as they emerge from political compromises.”⁶² The assertion of invariance obviates the need to examine the politics of distribution and therefore blocks the necessary analysis. The inversion of the Kaplow & Shavell thesis recognizes the centrality of political action costs, opening the way to the analysis Heckman describes.

From there, work might proceed in a number of possible directions. One approach would involve studying the efficacy of particular legal rules at achieving welfare-enhancing distributive improvements and considering whether these gains are *uniquely* achievable through that method (*i.e.*, are otherwise politically infeasible). If so, the question becomes whether the welfare improvements outweigh the losses caused by the behavioral distortions of the inefficient rule.⁶³ Another line of research might consider

⁶¹ See James J. Heckman, *The Intellectual Roots of the Law and Economics Movement*, 15 LAW & HIST. REV. 327, 332 (1997).

⁶² *Id.*

⁶³ For some examples of legal rules that may have both efficiency and distributive consequences, see Fennell & McAdams, *supra* note 51, at 1065-68 (discussing examples of arbitration clauses, teacher tenure, military

how psychological biases, framing, or perceptions of fairness interact with different modes of distribution and consider whether there are design features that might be tweaked to make redistribution through tax work better. Explicitly linking legal rules that have distributive consequences with particular transfers that can be carried out through the tax system might be one way to harness the fairness and desert intuitions that might accompany redistribution through legal rules.

Recognizing the potential for variable distributive results through different mechanisms also opens up significant institutional questions. Should particular institutions be given priority in determining distributive policy? If different governmental actors subscribe to different social welfare functions (some of which weight equality more heavily than others), whose vision should be allowed to prevail? For welfarists, answering these questions requires considering not only the impacts of behavioral distortions and distributive patterns on welfare, but also whether there are any welfare implications associated with privileging a particular institutional actor in the distributive arena.

D. The Prisoners' Dilemma

Unlike the prior examples, the Prisoners' Dilemma is not an independent theory; it is a particular application of game theory. Nonetheless, by generalizing lessons from legal scholars' many specific uses of the Prisoners' Dilemma (PD) in a vast literature across many domains of law, we can articulate what has become an implicit theory.⁶⁴

In the vivid story by which this abstract game gets its name, a prosecutor creates a PD by making each of two prisoners the same offer (where each prisoner knows the offer is made to both of them): "If you both confess, I will give you both a lenient sentence of 3 years in prison; if you both remain silent, I only have evidence to convict you of a minor crime and you will each serve 1 year in prison; if one of you confesses and the other remains silent, the confessor will walk free, while the non-confessor will get no leniency and serve 7 years." Here is the decision matrix that the prisoners confront, as it is typically presented:

recruitment, criminal enforcement, and environmental regulation).

⁶⁴ As one of us previously reported, a 2009 Westlaw search in the Journals & Law Reviews (JLR) database for "prisoner's dilemma" or "prisoners dilemma" (which also retrieves "prisoners' dilemma") resulted in 3119 documents. See Richard H. McAdams, *Beyond the Prisoners' Dilemma: Coordination, Game Theory, and Law*, 82 S. CAL. LAW REV. 209, 214 n.14 (2009). Westlaw appears to have changed its libraries slightly, but the same search more recently (July 7, 2017) in the Law Reviews & Journals database retrieved 4531 documents. Ten articles in just the last 5 years (since 2011) contain the term "prisoner's dilemma" or "prisoners' dilemma" in the title.

Table 1: Prisoners' Dilemma (Payoffs for Rohn, Colin)

	Colin Remains Silent	Colin Confesses
Rohn Remains Silent	(-1, -1)	(-7, 0)
Rohn Confesses	(0, -7)	(-3, -3)

Rohn does best to confess if Colin confesses (getting 3 years rather than 7), and best to remain silent if Colin remains silent (getting 0 years instead of 1). Colin is in exactly the same position: he does best confessing no matter what Rohn does. Thus, both prisoners confess (“defect”) and do 3 years in prison, though they would both gain if they could both remain silent (“cooperate”) and receive 1 year.

This “both defect” prediction is driven by the particular assumptions built into the structure of the game. Far from being treated as an odd puzzle that might arise under rarified circumstances, however, the PD game has been widely used in legal theory, by law professors and political scientists, to build a normative case for legal intervention.

1. Popular Understanding

At its core, the PD is understood to justify the use of law to solve problems of cooperation.⁶⁵ First, the PD shows that cooperation is impossible because the dominant strategy is to defect. Second, the PD shows that legal sanctions can “solve” the problem of cooperation in a way that makes everyone better off.⁶⁶ (We set aside here situations where third parties are harmed, such as potential victims who would prefer that criminals be punished for their crimes). Scholars’ applications of the PD to law are legion even though, as we will see, all of the strong assumptions built into the game are rarely true at the same time.⁶⁷ Applications include

⁶⁵ Some have previously dissented from what we describe here as the popular view. See McAdams, *supra* note 64; Wayne Eastman, *Telling Alternative Stories: Heterodox Versions of the Prisoners’ Dilemma, the Coase Theorem, and Supply-Demand Equilibrium*, 29 CONN. L. REV. 727, 740 (1997) (“The message of the standard Prisoners’ Dilemma is not simply that self-interested parties will fail to achieve optimal outcomes, but that self-interested parties given the ability to make agreements and a functioning contract law system will achieve optimum outcomes, and that self-interested parties engaging in repeated interactions may well do so.”). These understandings, which line up in some ways with the inversion of the PD we discuss below, seem to be overshadowed by the more gloomy interpretations we emphasize.

⁶⁶ There are various ways that the law can change the payoffs to alter the dominant strategy from one of universal defection to one of universal cooperation, whether by sanctioning defection, subsidizing cooperation, or otherwise altering the relative attractiveness of the alternatives.

⁶⁷ The same might be said of the multi-player version of the PD, the Tragedy of the Commons. See, e.g., ELINOR OSTROM, GOVERNING THE COMMONS: THE EVOLUTION OF INSTITUTIONS FOR COLLECTIVE ACTION 3-5(1990) (noting structural equivalence between the PD and the Tragedy of the Commons); Lee Anne Fennell, *Commons, Anticommons, Semicommons*, in RESEARCH HANDBOOK ON THE ECONOMICS OF PROPERTY LAW 35, 36 (Kenneth Ayotte & Henry E. Smith, eds., 2011) (noting that the standard presentation of the tragedy features conditions that make cooperation more likely). Like the PD, the Tragedy of the Commons is a staple of legal scholarship. A recent Westlaw search (July 7, 2017) for “tragedy of the commons” in Law Reviews & Journals

contracts,⁶⁸ corporate law,⁶⁹ banking,⁷⁰ international law,⁷¹ the federal judiciary,⁷² and civil discovery.⁷³

2. Unrealistic Assumptions

The prediction that the PD players will fail to reach the mutually desirable outcome – that they will defect rather than cooperate – is based on several strong assumptions. As a general matter, game theory assumes that the players are perfectly rational and perfectly self-interested. The classic PD game further assumes a peculiar payoff structure that both makes joint defection dominant but that would also deliver mutual gains to the parties if joint cooperation could be achieved. It is also assumed that the players interact against each other in the game only once.

These assumptions produce two positive claims that are crucial to the popular version of the PD: First, that people cannot cooperate, and second, that forcing them to choose the cooperative solution (through law) will improve everyone's payoffs. Together, these positive claims support the normative prescription to intervene through law whenever a PD is afoot. These are highly demanding assumptions. A PD is only present if the conditions mentioned above hold true in combination, and they will do so only rarely. To see why the PD a relatively rare game structure to encounter in the real world, it's important to consider some of the ways that the assumptions can fail to hold, especially in combination.

To begin, consider the singular payoff ordering necessary for a PD. The PD arises when and only when each player ranks the outcomes as follows:

yielded 3537 documents.

⁶⁸ See, e.g., Robert L. Birmingham, *The Duty to Disclose and the Prisoner's Dilemma: Laidlaw v. Oregon*, 29 WM. & MARY L. REV. 249, 275-82 (1988); Robert E. Scott, *Conflict and Cooperation in Long-Term Contracts*, 75 CAL. L. REV. 2005 (1987).

⁶⁹ See, e.g., Steven H. Kropp, *Corporate Governance, Executive Compensation, Corporate Performance, and Worker Rights in Bankruptcy: Some Lessons from Game Theory*, 57 DEPAUL L. REV. 1 (2007).

⁷⁰ Several scholars describe bank runs as PDs. See e.g., Mark E. Van Der Weide & Satish M. Kini, *Subordinated Debt: A Capital Markets Approach to Bank Regulation*, 41 B.C. L. REV. 195, 201 (2000); Daniel R. Fischel, Andrew M. Rosenfield & Robert S. Stillman, *The Regulation of Banks and Bank Holding Companies*, 73 VA. L. REV. 301, 307-10 (1987); Jonathan R. Macey, *The Business of Banking: Before and After Gramm-Leach-Bliley*, 25 J. CORP. L. 691, 696-97 (2000).

⁷¹ See, e.g., Maya Steinitz and Paul Gowder, *Transnational Litigation as a Prisoner's Dilemma*, 94 N.C. LAW REV. 751 (2016); George Norman & Joel P. Trachtman, *The Customary International Law Game*, 99 AM. J. INT'L L. 541, 548-53 (2005).

⁷² See, e.g., David S. Law, *Appointing Federal Judges: The President, the Senate, and the Prisoner's Dilemma*, 26 CARDOZO L. REV. 479, 502-04 (2005); Thomas W. Merrill, *Pluralism, the Prisoner's Dilemma, and the Behavior of the Independent Judiciary*, 88 NW. U. L. REV. 396, 401-09 (1993).

⁷³ See, e.g., Paul W. Grimm and Heather Leigh Williams, *"The [Judicial] Beatings Will Continue until Morale Improves": the Prisoner's Dilemma of Cooperative Discovery and Proposals for Improved Morale*, 43 U. BALT. LAW FORUM 107 (2013); Mitchell London, *Resolving the Civil Litigant's Discovery Dilemma*, 26 GEO. J. LEGAL ETHICS 837 (2013).

1. Best to defect when the other cooperates;
2. Second best when both cooperate;
3. Third best when both defect; and
4. Fourth best (worst) to cooperate when the other defects.

Note that what matters here is how the players subjectively rank the outcomes, not the number of years (or dollars, or any other objective metric) associated with each outcome. For example, if a person values “beating the prosecutor” as an end, or has reason to care how much time the other prisoner spends incarcerated, then the above prosecutor’s deal may not create a PD. A change in the rank order of any two outcomes converts the game into something other than a PD, such as one of the other well-known two-by-two games with entirely different properties.⁷⁴ In these other games, rational and selfish players may cooperate without legal sanctions.

The PD structure creates the prospect of mutual gains from cooperation (or forced cooperation through law). Yet, most real-world situations have both winners and losers, even when solving a collective action problem. Take the simple example of overfishing, an instance of the tragedy of the commons. Everyone appears to lose if the fish in the common lake are depleted; everyone appears to gain if there are sufficient limits to preserve a stable population of fish. Yet some people might face an endgame problem: if they are happy to leave the area when the pool is depleted, they might be worse off by the limitation on fishing. And then there is the problem where there is more than one way to define the limits. One might just limit the number of fish per person. But if monitoring each fish caught is costly, an alternative might be to limit fishing technology – the size of a boat or boat net – factors that are highly observable. Assume that the tradeoffs make the two regulations equally efficient. Nonetheless, those whose business model depends on the scale economies of a larger-than-permitted boat or net may find that they cannot continue fishing if the equipment limit is chosen—making them worse off than if they had been free to exploit the fish supply without constraint until it collapsed.

A final assumption is that the game is one-shot. The PD might be one-shot because the situation is entirely one-of-a-kind for that individual. Yet, if the person has never encountered a context like the one he is in, he might not recognize its strategic essence is the PD. Though players would likely learn through repeated interactions, repetition makes cooperation more likely. When the PD game is iterated – when two individuals engage in an indefinite repetition of PD interactions – cooperation can and does

⁷⁴ See RICHARD H. MCADAMS, *THE EXPRESSIVE POWERS OF LAW: THEORIES AND LIMITS* 29-39 (2015). As one example, if the rank of the first and second alternatives flip for both players the situation becomes an Assurance game in which a selfish player might rationally choose to cooperate. *Id.* at 39.

emerge.⁷⁵

Indeed, even people who are *actually* in a one-shot game are often uncertain that the game will actually prove to be one-shot, or they may fear reputational sanctions that transfer across one-shot interactions. For example, if the players are an Uber driver and passenger, they may decide to cooperate because the ratings each generates of the other affects their future interactions, even if the two people will never see each other again. It is also assumed that the players are perfectly self-interested, driven neither by altruism or animosity toward their opponents. There are many ways this assumption might fail. For example, at least outside of perfectly anonymous interactions, the knowledge that another player might become angry at one's selected strategy (even if the other player cannot retaliate) might be enough to alter the perceived payoffs in a real world setting and change the PD into a different game entirely.⁷⁶

Because the full set of conditions for a PD are rarely met simultaneously, the assumptions that follow from that game structure are very often false. This makes the theory ripe for inversion.

3. Inverting Prisoners' Dilemma Theory

One might invert the PD theory in two ways. First, if the PD were the primary justification for law, one might use the theory to demonstrate how infrequently law is necessary. If the assumptions that collectively suggest people cannot cooperate are rarely fully met, they will rarely need law to induce them to cooperate.⁷⁷ The same inversion might focus our attention on how to ensure that the conditions for cooperation – the absence of the PD – exist. This is what Elinor Ostrom's work did with respect to the multi-player version of the Prisoners' Dilemma, the Tragedy of the Commons.⁷⁸ By examining design principles that enable long-term

⁷⁵ See, e.g., ROBERT AXELROD, *THE EVOLUTION OF COOPERATION* (1984); Charles Holt, Cathleen Johnson, and David Schmitz, "Prisoner's Dilemma Experiments," in *THE PRISONER'S DILEMMA* 243-264 (Martin Peterson, ed., 2015). Later work uses evolutionary methodology to explain the emergence of cooperation. See, e.g., ROBERT AXELROD, *THE COMPLEXITY OF COOPERATION* (1997). In theory, iteration can unravel if people know exactly how many rounds will be played, due to what game theory calls "backward induction." Players reason that they will defect in the last round n , which means there is no reason to cooperate in round $n-1$. Expecting mutual defection in round $n-1$, there is no reason to cooperate in round $n-2$, and so on, to the first round. Nonetheless, if people are a little less than perfectly rational, or doubt that others are fully rational, they might not work through this logic; they might cooperate in at least in the early rounds of a lengthy set of interactions. See Philip Pettit and Robert Sugden, *The Backward Induction Paradox*, 86 *J. PHIL.* 169 (1989); David Kreps, Paul Milgrom, John Roberts and Robert Wilson, *Rational Cooperation in the Finitely Repeated Prisoner's Dilemma*, 27 *J. ECON. THEORY*, 245 (1982).

⁷⁶ Even among perfect strangers who know they will never interact again, the emotion of shame or the desire for esteem may motivate cooperation in a PD. See, e.g., Richard H. McAdams, *The Origin, Development, and Regulation of Norms*, 96 *MICH L. REV.* 338 (1997).

⁷⁷ Perhaps people need law for other purposes, such as redistribution, but this justification is not supported by the PD.

⁷⁸ See generally OSTROM, *supra* note 67.

cooperative solutions, the focus was taken off the need for government coercion or strong private property rights as the inevitable solutions.⁷⁹

There is a second way to invert PD theory. The constructed theory suggests that legal intervention is uncontroversially normatively justified because solving the PD creates only winners. An intervention to solve a classic PD easily satisfies the principle of Pareto efficiency because it makes at least one person (in fact, both people) better off, and makes no one worse off. But if the PD is the leading exemplar for when law can make Pareto-superior improvements, the inverted theory emphasizes how infrequently such moves are possible in the real world. Instead, we must almost always decide how to trade off gains against losses, winners against losers—something that cannot be done through simple allegiance to Pareto efficiency. Just as there are no free lunches, there are no normative cakewalks.

In short, the uninverted PD theory makes the normative issue seem simpler than it is. The inverted theory says that it will be rarely be the case that solving a cooperation problem is Pareto-efficient. What the inverted theory pushes us to recognize about the situation is the inevitability of difficult distributive judgments. Nor will it do to suggest that the answers to these distributive questions are unimportant or can be safely ignored, on the grounds that tax and transfer can address any distributive shortfalls. As we saw in the inversion of the K&S theory above, that response relies on another set of false assumptions.

III. THE VALUE AND LIMITS OF INVERSION

Inverting a theory is different from critiquing it on its own terms or rejecting it outright. The theory inverter sees something valuable in the connections the theory makes between variables, despite the lack of alignment between the theory's supporting assumptions and reality, and works to extract and apply that lesson in manner consonant with real-world conditions. But inversion can raise some concerns of its own, as we will explain.

A. *Inversion and Other Critiques*

In thinking about the significance of the inversion exercise, it is helpful to consider how inverting a theory is different from engaging in other kinds of critiques, including rejecting the premises of the theory outright, watering it down, or attaching warning labels to it. Inversion differs from other forms of critique in that it grants (at least for purposes of discussion)

⁷⁹ See *id.*

everything claimed by the theory except the false assumption under scrutiny. In the examples above, it was possible to grant the internal logic of the theoretical claim while turning attention to the falsity of the assumption that linked that logic to a normative takeaway.

It is of course possible to attack such theories in other ways. For example, Nozick's entitlement theory might be rejected on the grounds that some patterns of holdings are unacceptable regardless of how they came about.⁸⁰ The Tiebout Hypothesis might be rejected on a view of human nature that distinguishes the preferences of a citizen from her acts of consumption.⁸¹ And Kaplow and Shavell's theory of tax superiority might be rejected on the grounds that they are just wrong about the extra distortion argument.⁸² Alternatively, a theory's punchline might be watered down in various ways by introducing exceptions and qualifications.⁸³ Or competing narratives might be attached to its central framework in an effort to broaden or challenge its takeaways.⁸⁴

Another alternative to inversion might be termed the "warning label" approach. Here, theorists or later commentators might apply a "use only as directed" warning to ideas, stressing that the theory cannot be applied in situations where the specified assumptions do not hold true to even a rough or approximate degree. As this description suggests, the warning label is best suited to theories that might be applied across a wide range of contexts, at least some of which feature conditions that are relatively close to those stated in the assumptions. The Coase Theorem is sometimes qualified in this way, where low transaction cost settings are thought to be close enough to a zero transaction cost world to validate the conclusion that efficiency will be reached. But as that example shows, it is often unclear what is going to count as "close enough": zero transaction costs turn out on closer inspection to have a truly other-worldly character that cannot be said to very closely resemble *any* real situations.⁸⁵

When the validating assumptions represent dramatic and pervasive departures from reality, a warning is unlikely to be of much use. A proviso that effectively tells readers "don't try this at home—or, actually, anywhere else in the real world" invites one of two responses. The first response is to

⁸⁰ Or that a variety of more subtle problems exist in Nozick's arguments. See, e.g., the essays in *THE CAMBRIDGE COMPANION TO NOZICK'S ANARCHY, STATE, AND UTOPIA* (Ralf M. Bader and John Meadowcroft, eds., 2011).

⁸¹ See, e.g., Gerald E. Frug, *City Services*, 73 N.Y.U. L. REV. 23, 32 (1998) ("It is commonly said, for example, that human beings see themselves not simply as consumers but also as citizens--and that they think differently in these two different roles.") (citations omitted).

⁸² See, e.g., Jolls, *supra* note 48.

⁸³ See Jeremy Kessler & David E. Pozen, *Working Themselves Impure: A Life Cycle Theory of Legal Theories*, 83 U. CHI. L. REV. 1819 (2016) (arguing that prescriptive legal theories tend to become increasingly complicated and compromised as they mature, ultimately dying out or persisting in a much-adulterated state).

⁸⁴ See Eastman, *supra* note 65; Carol Rose, *Game Stories*, 22 YALE J. LAW & HUMAN. 369 (2010).

⁸⁵ See, e.g., R.H. COASE, *THE FIRM, THE MARKET, AND THE LAW* 15 (1988) ("It would not seem worthwhile to spend much time investigating the properties of such a world.").

ignore the warning. The second response might be to set aside the theory as useless. That may well be the right move in many cases, but a combination of this response with the one above can stifle dialogue: those attracted to the theory apply it sans warning, while everyone else ignores the theory itself. Inversion offers a third way, one that neither rejects the theory outright nor accepts it at face value. Instead, the logic of the theory is harnessed by focusing analytic attention on the implications of the false assumptions.⁸⁶

B. Inversion Concerns

Despite the advantages noted above, inversion invites some objections and comes with some limitations. We can start by observing that inversion is not an especially popular mode of scholarly engagement. It is easy to understand why. Inversion requires approaching an established theory with an awkward combination of embrace and pushback. The method grants enough of the initial theory to alienate critics, yet the inversion is unlikely to garner much appreciation from the popularized theory's supporters.

Indeed, the original theorist is likely to assert that the inversion exercise is redundant or turns their theory into a straw man, if the original theory already made explicit the assumptions on which it depended. The whole point of assumptions is that they assume something that may not be strictly true. Models simplify reality to make a point,⁸⁷ and it may seem pointless and uncharitable to harp on the fact that a theory depends on the assumptions that it already says it assumes, or to assert (as if it were news) that the theory doesn't work if the assumptions are not true. Popularizers, for their part, have long ago stopped thinking about the unreality of the assumptions on which the theory's normative takeaways hinge and are unlikely to relish reminders on this score.

One thing that can go wrong, then, is that the effort at inversion doesn't work to change scholarly views. One aspiration of this essay is to show that inversion can be a valid and useful mode of discourse. Inversion has both a critical and a constructive function. The critical function is to carefully separate out the formal results of a theory from the normative prescriptions associated with it. Quite simply, one cannot prescribe anything based on assumptions that are false in policy-relevant ways.⁸⁸ The constructive

⁸⁶ Although we focus here on specific theories rather than broader methodological approaches, we note a parallel between inversion and the rise of behavioral law and economics in response to the traditional rational actor model. Significantly, behavioral law and economics accepts the premise that regularities in human behavior can form the basis for predictions and hence for policy, but holds different assumptions from the rational actor model about the content of those regularities. See Christine Jolls, *Behavioral Economic Analysis of Redistributive Legal Rules*, 51 VAND. L. REV. 1653, 1654 (1998).

⁸⁷ See RODRIK, *supra* note 1, at 44 ("Economic models are relevant and teach us about the world *because* they are simple.").

⁸⁸ Again, we would distinguish simplifications that remain valid for predictions and policy, and those that are indispensable to the prescriptions. See *id.* at 27-29 (distinguishing critical assumptions from those that are not

function of inversion is to recognize the contribution that the formal result makes and to show what its normative implications might be given the falsity of the assumptions. In this way, a compelling theory can be transformed from one that is irrelevant to policy to one that can be used as a basis for policy.

Consider what might be the most successfully inverted theory to date, the Coase Theorem. Of course, one can argue that it never really needed inversion in the first place, since Coase himself subscribed to a view that emphasized the existence of positive transaction costs. Nonetheless, a popular view of “Coaseanism” that equated it with the capacity of bargaining to solve all ills had to be upended in order for transaction costs to take center stage. And this move founded an influential vein of scholarly innovation.

Yet, as one of us has argued in other work, the Coasean inversion did not get things quite right either.⁸⁹ Transaction costs were turned into objects of scorn, things to be attacked and minimized.⁹⁰ In fact, they represent just one way in which access to resources might be blocked once we move away from a zero-transactions-cost world. Resource access can be blocked not only by thwarted transactions but also by property structures that fail to keep resources in place when their current possessor is the high valuer. And, very significantly, another way that access to resources might go awry would be by pouring too many of our resources into *reducing* transaction costs.⁹¹

This example contains an important lesson about inversion. One cannot simply take the false assumption—“if transaction costs were zero”—and build policy around trying to make it as true as possible, no matter the cost of doing so. The other examples above can be approached similarly. Absolute justice in acquisition and transfer, perfect mobility, invariant political action costs, and Pareto-efficient solutions to intractable collective action problems are just as unattainable as zero transaction costs. While pursuing each holds value, efforts to move closer to these ideals should be pursued only insofar as the costs of doing so are worth incurring. Once that is no longer the case, the remaining gap between reality and theory must be accounted for in other ways.

A final concern about inversion is whether scholars might press it into service to serve an agenda of some kind. Readers may note that three of our examples⁹² involve theories that, in popular form, are associated with

essential to the conclusions); 213 (“Unrealistic assumptions are OK; unrealistic *critical* assumptions are not OK.”).

⁸⁹ Lee Anne Fennell, *The Problem of Resource Access*, 126 HARV. L. REV. 1471 (2013).

⁹⁰ See, e.g., Carl J. Dahlman, *The Problem of Externality*, 22 J.L. & ECON. 141, 142 (1979) (“[I]n the theory of externalities, transaction costs are the root of all evil.”).

⁹¹ See Fennell, *supra* note 89, at 1501-02.

⁹² See *supra* Parts II A., B., and C.

politically conservative messages, and that in inverted form support more politically liberal messages. But the inversion template could be applied to all sorts of theories, including ones that are associated with politically liberal takeaways, as the example of the Prisoners' Dilemma shows.⁹³ We hope that this essay will prompt others to find their own favorite candidates for inversion. Nonetheless, there are interesting questions about what kinds of uninverted theories tend to take hold, notwithstanding their false assumptions—a question we take up next.

C. *What's the Attraction?*

Our examples show considerable success of theories in what we call their uninverted form. One might ask why. Why do theories become popular in this implausible form instead of the more plausible inversion? It is not hard to understand why both the consumers and producers of theories would be attracted to strong, counterintuitive normative takeaways. In academic theorizing, fortune favors the bold. People are more likely to notice a theory and to find ways of using it, when it appears to demonstrate decisively some startling result. In all the cases we discuss, the uninverted theories are bolder than the inverted ones precisely because they minimize the apparent significance of questionable assumptions. This makes them fun to discuss and easy to propagate.

What is perhaps more perplexing is how false assumptions gain the staying power necessary to keep theories uninverted even when they make no sense that way. Here, it may be helpful to consider theories that are effectively “inverted on arrival” so that the falsity of the assumption always takes center stage. Consider this much-quoted line from Federalist 51: “If angels were to govern men, neither external nor internal controls on government would be necessary.”⁹⁴ The predicate—governance by angels—is obviously false, but perhaps no more obviously false than zero transaction costs, perfect mobility, or complete justice in acquisition or transfer. Yet to our knowledge no one has ever used an uninverted version of Federalist 51 to argue that there is no need to restrain government. Why are other theories grounded on false assumptions allowed to take hold in uninverted form?

One possibility is that certain kinds of assumptions, like those paired

⁹³ Cf. ALBERT O. HIRSCHMAN, *THE RHETORIC OF REACTION: PERVERSITY, FUTILITY, JEOPARDY* 149-63 (1991) (making a similar observation about the theories that he uses to illustrate rhetorical patterns emphasizing “futility, perversity, and jeopardy” and introducing some ways that liberal theories employ similar patterns); Eastman, *supra* note 65, at 732 (presenting alternative narrative versions of familiar models, including the Coase Theorem and the PD, “that have different moral and political implications from the canonical accounts but that accord with and illustrate the models' logical twists equally well”).

⁹⁴ *The Federalist*, 51 (James Madison), in *THE FEDERALIST PAPERS* 256, 257 (Lawrence Goldman, ed., 2008).

with mathematical models, require some degree of engagement and deciphering before debunking is possible, something that many critics may not be in a good position to do. It's also possible that attacking assumptions is simply a more common strategy in some lines of discourse than others. For example, our anecdotal sense is that it is common to respond to a stereotypically liberal argument by first granting the goal (e.g., helping the poor) and then suggesting that the proposed approach (e.g., rent control) will actually work at cross-purposes with that goal because a crucial assumption is false (e.g., that supply will remain unchanged). More stereotypically conservative positions, such as those rejecting redistribution to lower-income people, may elicit a different standard response that involves outright rejection of the theory, rather than separating out the assumptions for detailed scrutiny. But more data points are necessary to determine if there is any ideological skew in the picture.

Second, we believe there is a certain resilience to theories that exist in inverted form for the true expert and mostly in uninverted form for the academic non-expert. Because the theorist concedes somewhere that the assumptions limit the exciting implications, the expert can assure the non-expert that the theory is valid, even while the popular non-expert audience embraces and deploys the theory in uninverted form. At least in our experience, the casual reader often takes confidence in the fact that some experts they trust endorse the theory without absorbing that the endorsement is as carefully framed by the same unrealistic assumptions as the original theory. This effect may give the uninverted theory remarkable resilience.

Third, and potentially most importantly (albeit most speculatively), we detect a theme common to all four uninverted theories we review, a possible common reason for their popularity. The theories promise sturdy scaffolding upon which normative analysis can be constructed, bypassing intractable distributive disagreements at the foundation level. Most legal academics are interested in questions about what the law should be,⁹⁵ yet normative analysis is difficult to get off the ground when the *ground itself* embeds deep distributive controversies.⁹⁶

Nozick's theory offers to show that the existing distributional pattern in a society need not matter to justice (if each step leading to each current property holding was just). Tiebout shows that local autonomy produces useful competition that satisfies consumer demand (if people are mobile

⁹⁵ We would contend that many legal academics who are not interested in normative questions (but who pursue purely positive theories, as do many social sciences) have arrived at their position from despair at the possibility of making normative progress.

⁹⁶ Cf. Carol M. Rose, *Canons of Property Talk, or Blackstone's Anxiety*, 108 YALE L.J. 601, 605 (1998) ("The hidden skeleton in property's closet is what I shall call the Ownership Anxiety--that is, anxiety over the foundations for existing distribution.").

enough, and there are no spillovers among jurisdictions). Kaplow and Shavell show that legal academics should ignore the intractable distributional issues whenever they are discussing legal topics other than tax (if the resulting distribution will be the same regardless of the method of redistribution). The Prisoners' Dilemma reveals a situation in which all affected parties benefit from government intervention to coerce cooperation (if the situation is exactly a one-shot PD with no distributional issues in how to cooperate). In each case, the theory offers a novel and counterintuitive means of breaking a potentially paralyzing normative impasse.

What about the Coase Theorem, which has enjoyed parallel lives in popular and inverted form? Perhaps its successful inversion can be explained by the fact that Coase himself worked very hard to debunk the popularized version of the theory and press forward his original point.⁹⁷ Perhaps the legal academy was also more open to receiving the inverted message because, unlike the other examples we have provided, inversion did not reopen any sealed-off distributive questions. Neither the popular nor the inverted version of the Coase Theorem has anything to say about distribution—both look only at efficiency.⁹⁸ If our supposition is correct, inversion will be most difficult to achieve—but perhaps most important to achieve—where it requires giving up the extraordinarily useful illusion of a distribution-free foundation for legal analysis.

CONCLUSION

Much like the numbered jokes (apocryphally) traded among traveling salesmen,⁹⁹ certain theories represent stock moves in scholarly conversations. They serve as placeholders for conclusions that are rarely given close scrutiny, despite their dependence on false assumptions. Inversion shifts the focus to the theory's false assumption to see what use

⁹⁷ See, e.g., COASE, *supra* note 85, at 13 (explaining that his objective in discussing “what would happen in a world in which transaction costs were assumed to be zero” was “to provide a simple setting in which to develop the analysis and, what was even more important, to make clear the fundamental role which transaction costs do, and should, play in the fashioning of the institutions that make up the economic system”); *id.* at 174 (“The world of zero transaction costs has often been described as a Coasian world. Nothing could be further from the truth.”); see generally Ellickson, *supra* note 5 (discussing Coase's efforts to repudiate the mainstream understanding of his ideas).

⁹⁸ Significantly, the inverted theory provided a much more fertile foundation for normative legal analysis than the popular version, which proclaimed the law's irrelevance to efficiency. However, Coase has suggested that his focus on positive transaction costs did open up questions that took economists out of *their* comfort zone. See COASE, *supra* note 85, at (“The world of zero transaction costs, to which the Coase Theorem applies, is the world of modern economic analysis, and economists therefore feel quite comfortable handling the intellectual problems it poses, remote from the real world though they may be.”). He went on to explain that this discussion should be regarded as “but a preliminary to the development of an analytical system capable of tackling the problems posed by the real world of positive transaction costs.” *Id.* Law and economics has attempted to supply that system, but has not yet built out the additional analytic constructs necessary to accommodate distributive questions. See text accompanying notes 61-62, *supra*.

⁹⁹ See G.J. Stigler, *The Conference Handbook*, 85 J. POLIT. ECON. 441 (1977) (recounting an “ancient joke” along these lines and providing a humorous catalog of stock comments made at conferences).

can yet be made of the theory. The method does not mandate any particular way of dealing with the assumption's falsity—except in its insistence that we continue to see what we can learn from the underlying theory.

Unqualified theories, perhaps especially those supporting status quo market and power structures, are more likely to be rejected root and branch by opponents rather than inverted in the way we suggest here. This is a mistake, at least when there remains something to learn from the theory. Engaging with the implications of the theories under more realistic assumptions offers a way of starting a dialogue rather than ending one.