ESSAY

Inversion Aversion

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

INTRODUCTION

Some objects, like Weebles and lawn darts, resist inversion. The same is true of certain popular legal theories—or so we argue. In this Essay, we explain what it means to "invert" a theory, why one would want to do such a thing, and why it might be difficult to accomplish.

I. WHY AND HOW TO INVERT A THEORY

Suppose a compelling theoretical model embeds a critical assumption that is extremely implausible or patently false.[1] Two responses predominate. One is to ignore the assumption's falsity and embrace the theory anyway. The other is to reject the theory outright. But there is 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 (mostly successful) inversion involves the Coase Theorem, which in its popular formulation

† Max Pam Professor of Law, The University of Chicago Law School. I am grateful for research support from the Harold J. Green Faculty Fund and the SNR Denton Fund.

†† Deputy Dean and Bernard D. Meltzer Professor of Law, The University of Chicago Law School.

For conversations about this Essay, we thank Anupam Chander, Victor Fleischer, Jerry Frug, Calvin Johnson, Michael Knoll, Steven Medema, Richard Schragger, Sloan Speck, Andrew Verstein, and participants in the Harvard Law School conference Celebrating Jerry Frug's Work on Cities. We also thank Reeves Jordan for excellent research assistance. An earlier, longer draft of this Essay circulated under the title *Inverted Theories* and remains available on Chicago Unbound at http://perma.cc/XB7Q-TXYE.

[1] See Dani Rodrik, *Economics Rules: The Rights and Wrongs of the Dismal Science* 27 (Norton 2015) (explaining that "an assumption is critical if its modification in an arguably more realistic direction would produce a substantive difference in the conclusion produced by the model").

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 Ronald 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* (and typically *does*) matter to efficiency.[4] This rearticulation puts the emphasis where Coase 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]

There are other popular theoretical models that are equally good candidates for inversion that have not been successfully inverted to date. Below, we discuss four such theories: Robert Nozick's entitlement theory of distributive justice, the Tiebout Hypothesis, Louis Kaplow and Steven Shavell's principle of tax superiority, and the Prisoners' Dilemma. The persistence of the uninverted, popularized forms of these theories is puzzling, and we explore some possible reasons for it. First, though, we spell out how inversion works, using the Coase Theorem as an example.

A. Conditions for Inversion

There are three basic ingredients that make a theoretical model suitable for inversion. First, it must contain strong and unrealistic critical assumptions. Second, it must have generated a widely shared popular understanding with a simple normative

---

[2] See generally R.H. Coase, *The Problem of Social Cost*, 3 J L & Econ 1 (1960). George Stigler named the Theorem, not Coase himself. George J. Stigler, *The Theory of Price* 113 (Macmillan 3d ed 1966).

[3] Coase, 3 J L & Econ at 15 (cited in note 2).

[4] For example, Mitchell 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, *An Introduction to Law and Economics* 15 (Wolters Kluwer 4th ed 2011). See also Deirdre McCloskey, *The So-Called Coase Theorem*, 24 E Econ J 367, 367 (1998) ("Economists have gotten the 'theorem' wrong; in fact, backwards."); Steven G. Medema, *HES Presidential Address: The Coase Theorem Lessons for the Study of the History of Economic Thought*, 33 J Hist Econ Thought 1, 4–5 (2011) (providing different versions of the Coase Theorem, including an inverted one from McCloskey).

[5] The field of new institutional economics, for example, centers on transaction costs. See Oliver E. Williamson, *Transaction-Cost Economics: The Governance of Contractual Relations*, 22 J L & Econ 233, 233 (1979) ("The new institutional economics is preoccupied with the origins, incidence, and ramifications of transaction costs.").

prescription that hinges (wittingly or not) on the truth of the unrealistic assumptions. Third, confronting the falsity of the assumptions generates an inverted version of the original theory that replaces this normative prescription with a quite different one. For the Coase Theorem, we can state these core ingredients as follows:

TABLE 1: INVERTING THE COASE THEOREM

Popular Understanding	Bargaining Produces Efficiency Regardless of Legal Entitlements
Unrealistic Assumption	Zero Transaction Costs
Inverted Version	Law Matters to Efficiency

We revisit these same three components as we work through our other examples below.

B. Inversion and Alternatives

Inversion, as we use the idea, tracks the creation of an inverse in logic.[6] For the Coase Theorem, we start with this categorical if-then statement:

> If transaction costs are zero, an efficient result will always be reached regardless of the initial allocation of entitlements.

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. To reposition emphasis on the untrue assumption, we form this inverse:

> If transaction costs are not zero, an efficient result will not always be reached regardless of the initial allocation of entitlements.

Or more succinctly: when transaction costs are positive, the initial allocation of entitlements can matter to efficiency.

Inverting a theory differs from most other forms of critique in that it grants (at least for purposes of discussion) everything claimed by the theory except the false assumption under scrutiny.[7] To be sure, one can undermine or challenge theories in

---

[6] See Alfred Tarski, *Introduction to Logic and to the Methodology of Deductive Sciences* 39–40 (Oxford 4th ed 1994) (Jan Tarski, ed).

[7] For a similar approach, see Duncan Kennedy, *A Semiotics of Legal Argument*, 42 Syracuse L Rev 75, 87 (1991) (describing "[f]lipping" as "appropriating the central idea of

other ways short of outright attack. For example, a theory's punchline might be watered down by introducing exceptions and qualifications.[8] Or competing narratives might be attached to the theory's central framework in an effort to broaden or challenge its takeaways.[9] Another alternative to inversion effectively applies a warning label to the theory, stressing that it should not be applied to situations in which the specified assumptions do not hold. Yet a proviso that effectively tells readers, "Don't try this at home—or, actually, anywhere else in the real world," invites disregard of either the warning or the theory. Inversion offers a third way, one that turns analytic attention on the implications of the false assumptions.

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.[10] We focus here on four examples: Nozick's entitlement theory of distributive justice, the Tiebout Hypothesis, Kaplow and Shavell's theory of tax superiority, and the Prisoners' Dilemma.

A. Nozick's Theory of Distribution

In his 1974 book, *Anarchy, State, and Utopia*,[11] Nozick worked out a theory of a minimal or "night-watchman" state that would largely limit its interventions to protecting citizens against force and fraud, and that would not involve itself in redistribution.[12] This hands-off approach to distributive issues is rooted in

---

your opponent's argument-bite and claiming that it leads to just the opposite result from the one she proposes").

8 See generally, for example, Jeremy K. Kessler and 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).

9 See, for example, Richard H. McAdams, *Beyond the Prisoners' Dilemma: Coordination, Game Theory, and Law*, 82 S Cal L Rev 209, 218–25 (2009); Wayne Eastman, *Telling Alternative Stories: Heterodox Versions of the Prisoners' Dilemma, the Coase Theorem, and Supply-Demand Equilibrium*, 29 Conn L Rev 727, 740–41 (1997). See also generally Carol Rose, *Game Stories*, 22 Yale J L & Humanities 369 (2010).

10 We have already learned of an additional theory that has been inverted or "reversed" in past work. See generally Michael S. Knoll, *The Modigliani-Miller Theorem at 60: The Long-Overlooked Legal Applications of Finance's Foundational Theorem*, 36 Yale J Reg Bull 1 (2018) (discussing the "reverse" Modigliani-Miller theorem of capital structure irrelevancy).

11 Robert Nozick, *Anarchy, State, and Utopia* (Basic Books 1974).

12 See generally id.

Nozick's "entitlement theory" of distributive justice, which asserts that the legitimacy of a society's distribution does not depend on the patterns of resources that people end up with, but rather on the processes through which people come to hold those resources.[13] A just distribution would result, according to Nozick, if certain unrealistic conditions were met—specifically, if all acquisitions and all subsequent transfers satisfied principles of justice in acquisition and transfer, or if any injustices in acquisition or transfer were fully rectified.[14]

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 targeted at the poor.[15] The claim that a properly understood theory of distributive justice eschewed any particular pattern was viewed as a defense of existing inequalities. Yet Nozick's entitlement theory endorses a pure market distribution only if everyone's holdings came about from combinations of just acquisition and just transfer (or if there has already been proper rectification for any injustice in holdings).

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 reflects 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. Nor has there been anything resembling full rectification for these injustices. Even Nozick regards the validating assumptions of his entitlement theory of distributive justice as so patently false as to make it impossible to use the theory to criticize any actual instances of redistribution.[16]

What happens, then, if we invert the 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 their current holdings. The principle of rectification would 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 to the

---

13 Id at 150–60.

14 Id at 150–82.

15 See, for example, Anupam Chander and Madhavi Sunder, *Is Nozick Kicking Rawls's Ass? Intellectual Property and Social Justice*, 40 UC Davis L Rev 563, 564 (2007) ("Robert Nozick stands as one of the foremost intellectual antagonists to claims for distributive justice.").

16 See Nozick, *Anarchy, State, and Utopia* at 231 (cited in note 11).

extent they were unjust), 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."[17]

Flipping the theory carries both backward-looking and forward-looking implications. 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 by the imperative to rectify—in at least a rough way—the past injustices that the current patterns reflect.[18] Looking forward, Nozick's theory suggests that it matters *how* people come to possess the things they have. Some processes of distribution draw tighter connections between desert and payoffs than others: compare, for example, a decontextualized cash transfer with a living wage. If the history of 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.

---

[17] Or as Hal Varian puts it, the reality of past injustice demands that rectification be treated not as "somehow minor" but rather "central to the issue of justice." Hal R. Varian, *Distributive Justice, Welfare Economics, and the Theory of Fairness*, 4 Phil & Pub Aff 223, 227 (1975).

[18] Nozick himself suggests as much. Nozick, *Anarchy, State, and Utopia* at 230–31 (cited in note 11). Indeed, Nozick observes that even a Rawlsian approach to distributive justice could potentially follow from his entitlement theory, although he notes this "may well be implausible." Id at 231. See John Rawls, *A Theory of Justice* 78–83 (1971) (describing the "difference principle," which holds that social and economic inequality is permissible only to the extent it improves the position of the least advantaged). Entitlement theory might lead to that distributive approach, Nozick explains, if, "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).

Nozick, *Anarchy State, and Utopia* at 231. This caveat, however, is not part of the popular understanding of Nozick's minimal state.

TABLE 2: INVERTING NOZICK'S ENTITLEMENT THEORY

Popular Understanding	Redistribution Is Unnecessary
Unrealistic Assumptions	Past Justice in Acquisition and Transfer (or Full Rectification)
Inverted Version	Past Injustice Makes Redistribution Necessary

B. The Tiebout Hypothesis

Charles Tiebout wrote *A Pure Theory of Local Expenditures*[19] 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.[20] Tiebout recognized that, in a multijurisdictional metropolitan area, "consumer-voters" can select among local governments.[21] If each local government offers a different mix of goods and services, along 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 highly unrealistic assumptions hold, including perfect knowledge, perfect mobility, no constraints associated with employment (all people are assumed to live on dividend income), and no spillovers among jurisdictions.[22]

The theoretical point, which came to be known as the Tiebout Hypothesis (TH), was a powerful one: it showed how entry and exit could substitute for a price mechanism and reveal information about preferences as residents "vote with their feet." In legal academia, 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.[23] The popular version of TH thus holds that each local

19 Charles M. Tiebout, *A Pure Theory of Local Expenditures*, 64 J Polit Econ 416 (1956).

20 See generally id.

21 See id at 418.

22 See id at 419–20.

23 See Richard Briffault, *The Rise of Sublocal Structures in Urban Governance*, 82 Minn L Rev 503, 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 and Sheila R. Foster, *The Mobility Case for Regionalism*, 47 UC Davis L Rev 63,

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.

These prescriptions crucially hinge on the untrue assumptions that Tiebout used in constructing what he described as an "extreme model."[24] Significantly, Tiebout saw sorting as a means to an end: it is through the process of sorting into communities that people register their preferences for different goods and services. But the quality of the information that is revealed through this process is only as good as real-world conditions permit. 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. Treating sorting as the relevant end gets things backwards and ignores the artificiality of Tiebout's central assumptions.

Suppose that we invert the theory to restore the original emphasis on demand revelation. The flipped version of TH suggests that impediments to mobility and extrajurisdictional impacts of local policies scramble the implicit price signals sent by moves and that it is necessary to correct these distortions. An inverted TH thus looks for ways to improve the conditions of mobility for everyone, increase awareness of the extrajurisdictional implications of local policies, and address interdependencies among communities so that the implicit price signals sent by moves are accurate ones.[25] This means examining the ways in which local governmental policies impose costs on other local governments, such as through exclusionary housing policies. More foundationally, it means focusing attention on how *all* households can be given meaningful choices among local jurisdictions.

---

73 (2013) ("[S]cholars regularly invoke the Tiebout model to support arguments for devolution and decentralization.").

24 Tiebout, 64 J Polit Econ at 419 (cited in note 19).

25 Tiebout himself suggests as much when he notes the allocative efficiency benefits of "[p]olicies that promote residential mobility and increase the knowledge of the consumer-voter." Id at 423.

TABLE 3: INVERTING THE TIEBOUT HYPOTHESIS

Popular Understanding	Respect Local Autonomy to Facilitate Self-Sorting
Unrealistic Assumptions	Perfect Mobility and Unconstrained Choice; No Spillovers among Communities
Inverted Version	Unconstrained Local Autonomy Distorts Self-Sorting

C. Kaplow and Shavell's Theory of Tax Superiority

Kaplow and Shavell famously modeled the advantages of conducting all redistribution through tax and transfer (hereinafter "tax") rather than through legal rules.[26] Moving money through the tax system is well known to produce distortions in the choice between labor and leisure.[27] However, Kaplow and Shavell argued that redistributive legal rules would embed an equivalent labor-leisure distortion, but would *also* distort choices about the primary behavior that is being regulated.[28] 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 and conduct all redistribution through tax alone.[29]

---

[26] See generally Louis Kaplow and Steven 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 (2000); Louis Kaplow, *The Optimal Supply of Public Goods and the Distortionary Cost of Taxation*, 49 Natl Tax J 513 (1996); Louis Kaplow and Steven Shavell, *Why the Legal System Is Less Efficient than the Income Tax in Redistributing Income*, 23 J Legal Stud 667 (1994); 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). A key antecedent in the economics literature was A.B. Atkinson and J.E. Stiglitz, *The Design of Tax Structure: Direct versus Indirect Taxation*, 6 J Pub Econ 55 (1976). Although Kaplow and Shavell qualify their claim of tax superiority in a number of respects, they note that "these qualifications involve subtle refinements that are tangential to the ordinary view concerning how legal rules might be adjusted to increase redistribution, namely, to favor the poor at the expense of the rich." Kaplow and Shavell, 29 J Legal Stud at 834–35.

[27] See Atkinson and Stiglitz, 6 J Pub Econ at 56–57 (cited in note 26) (explaining that a tax system will inevitably produce distortions when the characteristic upon which the tax is based, such as income, lies under the individual's control).

[28] Kaplow and Shavell, 23 J Legal Stud at 667–68 (cited in note 26).

[29] See id at 677. We term this claim "prescriptive tax superiority" and distinguish it from "formal tax superiority," which focuses only on Kaplow and Shavell's formal result. See Lee Anne Fennell and Richard H. McAdams, *The Distributive Deficit in Law and Economics*, 100 Minn L Rev 1051, 1058–69 (2016).

This prescriptive claim has become dominant in law and economics.[30] But it depends on a core unrealistic assumption—that political impediments to redistribution are insensitive to the method of redistribution. Critics of Kaplow and Shavell's theory have often noted that redistribution through tax may be politically infeasible[31] but have rarely focused on Kaplow and Shavell's rejoinder: that any political impediments to redistribution through the tax system will also block, *to an equal extent*, efforts to redistribute through legal rules.[32] This claim—that the amount of redistribution cannot be altered by the choice of distributive method ("distributive invariance")—is false.[33] For many reasons we have previously explored, it may be more difficult politically to move money through the tax system than through a substantive legal rule.[34] If one can get *different* distributive results by using legal rules instead of or in addition to tax, tax will not always be the exclusively preferred method.

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. 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 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.

---

30 See, for example, Fennell and McAdams, 100 Minn L Rev at 1062 n 32 (cited in note 29) (collecting citations); Kyle Logue and Ronen Avraham, *Redistributing Optimally: Of Tax Rules, Legal Rules, and Insurance*, 56 Tax L Rev 157, 158 (2003).

31 See, for example, Fennell and McAdams, 100 Minn L Rev at 1074 n 69 (cited in note 29) (collecting citations).

32 See, for example, Kaplow and Shavell, 23 J Legal Stud at 675 (cited in note 26).

33 See Fennell and McAdams, 100 Minn L Rev at 1079–1109 (cited in note 29) (defining the term "distributive invariance" and demonstrating its implausibility).

34 See generally id. For additional discussion of why distributive changes are not offset, which is one way in which distributive invariance fails, see Zachary Liscow, *Is Efficiency Biased?*, 85 U Chi L Rev 1649, 1664–66 (2018).

TABLE 4: INVERTING KAPLOW AND SHAVELL'S THEORY OF TAX SUPERIORITY

Popular Understanding	Ignore Distributive Implications of (Nontax) Legal Rules
Unrealistic Assumptions	Political Impediments Are Insensitive to the Method of Redistribution
Inverted Version	No Redistributive Routes Should Be Ruled Out

D. The Prisoners' Dilemma

The Prisoners' Dilemma (PD) is an application of game theory that has been widely used by legal scholars. Unlike the above examples, there is no explicit normative theory of the PD, but from many specific uses, we can articulate what has become an implicit theory.[35]

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

[35] A vast literature across many domains of law makes use of the PD. As one of us previously reported, a 2009 Westlaw search in the Journals & Law Reviews database for "prisoner's dilemma" or "prisoners dilemma" (which also retrieves "prisoners' dilemma") resulted in 3,119 documents. McAdams, 82 S Cal L Rev at 214 n 14 (cited in note 9). Westlaw appears to have changed its libraries slightly, but the same search in the Law Reviews & Journals database more recently (on October 14, 2018) retrieved 4,828 documents. Nine articles just since 2012 contain the term "prisoner's dilemma" or "prisoners' dilemma" in the title. Scholars have applied the PD to many legal areas, including contracts, corporate law, banking, international law, the federal judiciary, and civil discovery.

TABLE 5: 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 three years rather than seven) and best to confess if Colin remains silent (getting zero years instead of one). Colin is in exactly the same position: he does best by confessing no matter what Rohn does. Thus, both prisoners confess ("defect") and do three years in prison, though they would both gain if they could both remain silent ("cooperate") and receive one 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 and multiplayer versions of the PD game have been widely used in legal theory, by law professors and political scientists, to build a normative case for legal intervention.[36]

At its core, the PD is widely understood to justify the use of law to solve problems of cooperation.[37] 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. But these twin predictions—that the PD players will defect rather than cooperate and that a resolution exists that will make everyone better off—depend on demanding assumptions that are rarely true in combination.

As a general matter, game theory assumes that the players are perfectly rational and perfectly self-interested. The classic PD game further assumes a one-shot game with a peculiar payoff structure that makes joint defection dominant but that would also deliver mutual gains to the parties if joint cooperation could be achieved. This follows from the way that both of the players are deemed to subjectively rank the options open to them: defecting while the other player cooperates is best of all and cooperating

---

[36] See, for example, McAdams, 82 S Cal L Rev at 214 nn 15–22 (cited in note 9) (collecting citations).

[37] Some have previously dissented from what we describe here as the popular view. See generally id. See also Eastman, 29 Conn L Rev at 740–41 (cited in note 9).

while the other player defects is worst of all, even though mutual cooperation has a higher payoff than mutual defection.

In contrast to the PD's assumptions, most real-world situations have both winners and losers, even when solving a collective action problem. Moreover, when two individuals engage in an indefinite repetition of PD interactions, cooperation can and does emerge.[38] Indeed, even people who are *actually* in a one-shot game are often uncertain that the game will ultimately prove to be one-shot, or they may fear broader reputational sanctions that transfer across one-shot interactions. There are many other ways that reality can diverge from the model's assumptions. For example, the knowledge that another player might become angry at one's selected strategy (even if the other player cannot retaliate) could alter the perceived payoffs in a real-world setting and change the PD into a different game entirely.[39]

Thus, the complete, defining set of PD assumptions is rarely true. Consider the implications. First, if the assumptions that collectively suggest people cannot cooperate are rarely fully met, law may be necessary for cooperation much less frequently than is usually supposed. We might focus instead on how to ensure that the conditions for cooperation exist. This is what Elinor Ostrom's work did with respect to the multiplayer version of the PD, the Tragedy of the Commons: she explored how real-world communities sharing a depletable common-pool resource can maintain it at sustainable levels without resort to law.[40]

Second, the popular theory suggests that legal intervention is uncontroversially justified because solving the PD creates only winners. In fact, we must almost always decide how to trade off gains against losses, winners against losers. One reason is that the defect/defect outcome may not be worse for everyone than cooperate/cooperate. A consumer of a common-pool resource who has an attractive outside option may prefer to jointly deplete the resource rather than to cooperate to maintain it; this person is made worse off by limits on consumption. Yet even if everyone potentially gains from cooperation, there is almost always more than one way to cooperate, and the choice of *how* to cooperate may controversially involve differing distributional consequences.[41] To

---

[38] See generally Robert Axelrod, *The Evolution of Cooperation* (Basic Books revised ed 2006); Charles Holt, Cathleen Johnson, and David Schmidtz, *Prisoner's Dilemma Experiments*, in Martin Peterson, ed, *The Prisoner's Dilemma* 243 (Cambridge 2015).

[39] See McAdams, 82 S Cal L Rev at 226–28 (cited in note 9).

[40] See generally Elinor Ostrom, *Governing the Commons: The Evolution of Institutions for Collective Action* (Cambridge 1990).

[41] See McAdams, 82 S Cal L Rev at 228–30 (cited in note 9).

prevent overfishing a common lake, for example, one might cooperate by limiting either the size of fishing boats or the days of fishing. If the fishers who already have big boats prefer to limit the days of fishing, while those with the smallest boats (who need to fish every day to support themselves) prefer to limit boat size, a normative conflict remains.

Inverting the implicit theory of the PD game pushes us to recognize that cooperation failures are not inevitable, but distributive disagreement is (nearly) inevitable.

TABLE 6: INVERTING THE PRISONER'S DILEMMA

Popular Understanding	Legal Intervention Is Necessary and Improves Payoffs for Everyone
Unrealistic Assumptions	Mutual Defection Is Everyone's Dominant Strategy but Mutual Cooperation Yields Higher Payoffs for All Parties
Inverted Version	Legal Interventions Are Rarely Necessary for Cooperation and Rarely Improve Payoffs for Everyone

III. WHY ARE UNINVERTED THEORIES SO STICKY?

As we have shown, inverting a theory is different from rejecting it. The theory inverter sees something valuable in the theory's internal logic and the connections it makes, despite the lack of alignment between the theory's supporting assumptions and reality. She then works to extract and apply that lesson in a manner consonant with real-world conditions.[42] Put this way, the approach sounds useful, even compelling. Why then do popular uninverted theories have such staying power? The Sections below take on this question from two angles: first by considering how inversion can fail or misfire, and then by examining the appeal of uninverted theories.

[42] 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 diverges from the rational actor model as to the content of those regularities. See Christine Jolls, *Behavioral Economics Analysis of Redistributive Legal Rules*, 51 Vand L Rev 1653, 1654 (1998).

A. Inversion Trouble

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 admiration and criticism. The inverter grants enough of the initial theory to alienate the theory's usual critics, yet the inversion itself irritates the theory's supporters. Indeed, the original theorist may assert that the inverter is attacking a straw man, if the theory initially made explicit the assumptions on which it depended. Why harp on the fact that a theory depends on its own stated assumptions or proclaim (as if it were news) that the theory doesn't work if the assumptions are not true? Popularizers, for their part, have long since 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. As a result, inversion efforts may fail to change the scholarly discourse—or may not be undertaken at all.

This Essay hopes to promote inversion as a valid and useful mode of analysis, one that adds value through both critical and constructive moves. The critical move involves delinking the formal results of a theory that depends on false assumptions from the normative prescriptions associated with that theory. Put simply, assumptions that are false in policy-relevant ways cannot be the basis for prescribing anything.[43] The constructive move is to use the theory's core insights to develop a normative analysis that takes seriously the falsity of the theory's assumptions. Consider how these two elements worked together in what is perhaps the most successful inversion to date, that of the Coase Theorem. Even though Coase himself emphasized the existence of positive transaction costs, a popular view of "Coaseanism" that made law irrelevant had to be upended in order for transaction costs to take center stage. This cleared the way for what became a rich and influential vein of scholarly innovation.

But inversion carries risks. Indeed, as one of us has argued in other work, the Coasean inversion did not get things quite right.[44] Transaction costs have been turned into objects of scorn,

---

[43] We refer here to assumptions that are essential to the prescriptions, not ones that are peripheral. See Rodrik, *Economics Rules* at 27–29 (cited in note 1) (distinguishing critical assumptions from those that are not essential to the conclusions). See also id at 213 ("Unrealistic assumptions are OK; unrealistic *critical* assumptions are not OK.").

[44] Lee Anne Fennell, *The Problem of Resource Access*, 126 Harv L Rev 1471 (2013).

things to be attacked and minimized.[45] In fact, they represent just one way in which access to resources might be blocked once we move away from a zero transaction cost world. Focusing on only one way of improving that access is a mistake. Among other things, it might cause us to pour too many of our resources into reducing transaction costs.[46] Likewise, efforts to attain the ideals of absolute justice in acquisition and transfer, perfect mobility, invariant political action costs, and Pareto-efficient solutions to intractable collective action problems may be impossible or prohibitively costly. Inversion does not dictate focusing solely on those objectives. We can—and should—think broadly and creatively about alternative ways to confront the false assumptions at the heart of the respective theories.

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 examples[47] involve theories that, in popular form, are associated with 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 PD shows.[48] We hope that this Essay will prompt others to find their own favorite candidates for inversion. Nonetheless, certain kinds of uninverted theories may be especially likely to take root, notwithstanding their false assumptions. The next Section considers why that might be.

B. What's the Attraction?

Our examples show considerable success of theories in what we call their uninverted form. Why do theories become—and remain—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,

---

45 See, for example, Carl J. Dahlman, *The Problem of Externality*, 22 J L & Econ 141, 161 (1979) ("[I]n the theory of externalities, transaction costs are the root of all evil.").

46 See Fennell, 126 Harv L Rev at 1501–02 (cited in note 44).

47 See Parts II.A–C.

48 See Part II.D. See also Eastman, 29 Conn L Rev at 732 (cited in note 9) (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"). Albert Hirschman makes an analogous observation about the theories that he uses to illustrate the rhetorical patterns of "futility, perversity, and jeopardy" and discusses how liberal theories might employ similar patterns. Albert O. Hirschman, *The Rhetoric of Reaction: Perversity, Futility, Jeopardy* 149–63 (Belknap 1991).

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.

One possibility is that certain kinds of assumptions, like those paired 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 (for example, helping the poor) and then suggesting that the proposed approach (for example, rent control) will actually work at cross-purposes with that goal because a crucial assumption is false (for example, that supply will remain unchanged). Such a "grant the goal" approach may be a less common response to more stereotypically conservative positions. But more data points are necessary to determine if there is a systematic ideological skew in this regard.

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 nonexpert. Because the theorist concedes somewhere that the assumptions limit the exciting implications, the expert can assure the nonexpert that the theory is valid, even while the popular nonexpert audience embraces and deploys the theory in uninverted form. At least in our experience, the casual reader often takes confidence in the fact that 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 appear to resolve or avoid otherwise intractable distributive disagreements, enabling scholars to set distribution aside in conducting their own analyses. Most legal academics are

interested in questions about what the law should be, yet normative analysis is difficult to get off the ground when the ground itself embeds deep distributive controversies.

Nozick's theory demonstrates 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 enough, and there are no spillovers among jurisdictions). Kaplow and Shavell enable legal academics to ignore 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 PD 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.[49] 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. If our supposition is correct, inversion will be most difficult to achieve—but perhaps most important to achieve—when it requires giving up the extraordinarily useful illusion of a distribution-free foundation for legal analysis.

CONCLUSION

Categorical, counterintuitive theories are attractive. But they can also mislead. When theories rely on patently false assumptions, ignoring the assumption's falsity means missing the point. Yet to reject such a theory, root and branch, is to discard a valuable route to insight. The theory inverter seeks instead to identify and preserve what is useful in the theory's internal logic

[49] See, for example, R.H. Coase, *The Firm the Market and the Law* 13, 174 (Chicago 1988).

and in the connections it makes. In this way, inversion can turn false assumptions into fulcrums for useful scholarship. Understanding why there might be aversion to such inversions can help clear the way for this promising approach.