

HARM, AMBIGUITY, AND THE REGULATION OF ILLEGAL CONTRACTS

*Adam B. Badawi**

INTRODUCTION

Courts have long preferred to remedy illegal contracts by denying state enforcement of the parties' interests in those contracts.¹ The primary justification for this non-enforcement—which leaves the parties in the position that courts find them—is to deter these bargains from forming in the first place.² According to the prevailing logic, a party who has to act first under the terms of an illegal contract would be hesitant to make an upfront payment or render a service on credit if an attempt to enforce the promise would result in a large, legally-sanctioned forfeiture.

But the reasons that often justify the state's imposition of liability for private behavior—the prevention of harm and the maximization of social welfare³—play little or no role in the determination that the non-enforcement remedy should apply.⁴ This structure puts the illegal contract

* Bigelow Fellow and Lecturer in Law, University of Chicago Law School. I thank Douglas Baird, Omri Ben-Shahar, Sam Bray, Mel Eisenberg, Lee Fennell, Mary Anne Franks, Ezra Friedman, Adam Muchmore, Ariel Porat, Arden Rowell, Max Schanzenbach, Lior Strahilevitz, and participants at the Annual Meeting of the American Law & Economics Association and the Northwestern Law & Economics Colloquium for helpful comments and discussion. All errors are my own.

¹ In a famous example from the eighteenth century, an English court refused to enforce a contract between two thieves that sought resolution of a disagreement over the distribution of their spoils. *Everet v. Williams*, (1893) 9 L.Q.R. 197 (Ex.). This non-enforcement rule remains the presumptive remedy for illegal contracts. RESTATEMENT (SECOND) OF CONTRACTS § 178 (1981); *see also* Juliet P. Kostritsky, *Illegal Contracts and Efficient Deterrence: A Study in Modern Contract Theory*, 74 IOWA L. REV. 115, 118 n.8 (1988) (“The Restatement . . . has adopted this no-effect principle by providing generally that neither enforcement nor restitutionary remedy is available to parties to illegal contracts.”).

² *See infra* Part I.A.; *see also* *McMullen v. Hoffman*, 174 U.S. 639, 669-70 (1899) (“To refuse to grant either party to an illegal contract judicial aid for the enforcement of his alleged rights under it tends strongly towards reducing the number of such transactions to a minimum.”). Modern courts continue to apply this reasoning. *See, e.g.*, *Mlynarik v. Bergantzel*, 675 N.W.2d 584, 587 (Iowa 2004) (“We have noted the purpose of [the non-enforcement rule] ‘is to deter future misconduct by denying relief to one whose losses were substantially caused by his own fraud or illegal conduct.’” (quoting *Gen. Car & Truck Leasing Sys., Inc. v. Lane & Waterman*, 557 N.W.2d 274, 279 (Iowa 1996))).

³ Some have suggested that the rationale for prohibiting certain contracts has its basis in the prevention of third-party harm. *See* F.H. Buckley, *Perfectionism*, 13 SUP. CT. ECON. REV. 133, 139 (2005) (explaining that deeming a contract illegal means “leav[ing] the immediate parties to the contract and consider[ing] its effects upon third parties”).

⁴ Courts have long noticed the confusion in this area. *See* *Richardson v. Mellish*, (1824) 130 Eng. Rep. 294, 303 (Ex.) (explaining that the practice of barring contracts based on public policy considera-

framework at odds with the typical model for remedies in private law, where the remedy calculus often takes account of the incentives that liability provides for generating societal wealth.⁵ This decoupling of the remedies for illegal contracts and their consequences creates the potential for undesirable outcomes because it allows for overdeterrence of beneficial contracts and underdeterrence of harmful contracts.

The concerns about overdeterrence can be particularly troublesome because the determination of whether a contract is illegal—and hence whether non-enforcement may apply—is often an uncertain enterprise. Contract restrictions frequently come from statutes, regulations, and common-law doctrines whose vague language and uncertain scope can complicate attempts to determine whether a contemplated contract is legal or not.⁶ Prohibitions on contracts that are against public policy may even be harder to identify because the concept of public policy is such a general one.⁷ This pervasive ambiguity, when combined with the threat of non-enforcement, can amplify the problems of overdeterrence. In some situations, the threat of losing any legal protection can lead parties to refrain from or minimize contracting even when they have strong reasons to believe that a contract is legal.⁸

These policy ills could be minimized through a remedy analysis that takes into account the potential harm a contract can create and the effect of legal uncertainty on ex ante decision-making. But the current framework does neither. The *Restatement (Second) of Contracts* (“the *Restatement*”), for example, specifies non-enforcement as the presumptive remedy for illegal contract terms and provides an open-ended, multi-factor test to govern

tions is “a very unruly horse, and when once you get astride it you never know where it will carry you. It may lead you from the sound law. It is never argued at all but when other points fail.”).

⁵ Joseph A. Grundfest & Peter H. Huang, *The Unexpected Value of Litigation: A Real Options Perspective*, 58 STAN. L. REV. 1267, 1317-18 (2006) (“A central normative proposition of the economic analysis of law is that litigation induces efficient resource allocation if the substantive rules of liability and damages are properly defined.”). Much of the work in this area has examined existing doctrine to determine whether it comports with this normative proposition. See, e.g., WILLIAM M. LANDES & RICHARD A. POSNER, *THE ECONOMIC STRUCTURE OF TORT LAW* (1987) (arguing that the common law tort doctrine is efficient). But see Eric A. Posner, *Economic Analysis of Contract Law after Three Decades: Success or Failure?*, 112 YALE L.J. 829, 852-55 (2003) (discussing the methodological difficulties with the determination whether doctrine is, or is not, efficient).

⁶ See, e.g., Kostritsky, *supra* note 1, at 116 n.4 (acknowledging that the determination whether a contract is illegal involves “difficult legal questions”).

⁷ *Fairfield Ins. Co. v. Stephens Martin Paving, LP*, 246 S.W.3d 653, 673 (Tex. 2008) (explaining that “the words ‘public policy’ are vague in meaning and dangerous of application, and that, unless we exercise due discrimination, we are likely to fall into error when we come to apply them to the construction of a contract” (citing *Singer Mfg. Co. v. Rios*, 71 S.W. 275, 276 (Tex. 1903))).

⁸ The basic intuition underlying this point is that the risk of losing upfront payments through forfeiture will cause parties to refrain from contracting that would take place in the shadow of a restitutionary measure. See *infra* Part I.

the exceptions that warrant an award of restitution.⁹ None of these factors—which include the parties’ justified expectations and the seriousness of the misconduct involved—direct courts to consider third-party harm, ambiguity, or the likely incentive effects of the remedy. While some courts hint at the concerns that ambiguity and harm implicate, the cases largely ignore these connections.¹⁰

Two cases under the current framework demonstrate how this overdeterrence and underdeterrence can produce undesirable results. The first example comes from a common area of contract restrictions—the regulation of professional licenses and relationships. Many states have laws that restrict the ownership of professional partnerships, like medical practices and law firms, to licensed professionals within those professions.¹¹ These rules mean that only licensed doctors may have ownership interests in medical practices and only licensed attorneys may have ownership interests in law firms.¹² The apparent aim of these laws is to provide some insulation between professional judgment and the profit motive.¹³ But these laws can make it difficult for professionals to exit partnerships because the owners can sell only to a subset of potential buyers.¹⁴

To solve this problem in the context of medical and other health professional practices, lay firms have attempted to design complex transactions where a firm buys a practice’s assets via a large upfront payment in exchange for a percentage of future revenues.¹⁵ These deals allow the practice to economize on procurement and logistics costs while leaving medical decision-making in the hands of doctors.¹⁶ Additionally, these agreements allow both sides to benefit and avoid the harm contemplated by the ownership restrictions. But when doctors sue to have the transactions declared illegal because they violate the restriction on lay ownership of professional

⁹ See RESTATEMENT (SECOND) OF CONTRACTS § 178 (1981). See also *id.* § 197 (instructing courts to avoid restitution as a remedy unless non-enforcement would cause “disproportionate forfeiture”); *id.* § 198(a) (suggesting that restitution is appropriate in cases of excusable ignorance).

¹⁰ Judge Posner has suggested that *ex ante* ambiguity might be a factor in the choice of remedies for illegal contracts. *Nagel v. ADM Investor Servs., Inc.*, 217 F.3d 436, 441 (7th Cir. 2000) (“[I]f the legality of a contract cannot easily be determined in advance, that might be a factor rebutting the presumption noted earlier that illegal contracts are unenforceable.”).

¹¹ For an overview of restrictions in the medical industry and a critique, see Andrew Fichter, *Owning a Piece of the Doc: State Law Restraints on Lay Ownership of Healthcare Enterprises*, J. HEALTH L. 1 (2006). For a review of the constraints in the legal profession and a critical take on those restrictions, see Larry E. Ribstein, *Ethical Rules, Agency Costs, and Law Firm Structure*, 84 VA. L. REV. 1707 (1998).

¹² Fichter, *supra* note 11, at 6; Ribstein, *supra* note 11, at 1721.

¹³ See Fichter, *supra* note 11, at 4.

¹⁴ See *id.* at 6-7.

¹⁵ *Id.* at 11-12.

¹⁶ *Id.* at 12.

practices, courts have applied the conventional non-enforcement rule.¹⁷ This result means that the lay firms forfeit the large upfront payment and lose any right to the future receipts from the practice. Given the ambiguity of whether the ownership restrictions apply to these types of transactions and the possibility for these types of forfeitures, it is unsurprising that the market for these transactions—including those that may not be illegal—has largely disappeared.¹⁸ The absence of any apparent externality makes this deterrence undesirable, but the lack of any inquiry into the possibility of harm associated with a particular deal and an insensitivity to the consequences of non-enforcement have effectively eliminated innovation in this area.¹⁹

This same approach can produce underdeterrence, as a second example shows. Prohibitions on commercial trade with rogue nations are a common device for achieving foreign policy aims. An express rationale for the restriction on trade with Iran is the fear that contracting with Iranian entities will help to finance a regime that provides resources to terrorist groups.²⁰ Given the harm-based rationale for this rule, the deterrent power of the non-enforcement remedy would appear to be appropriate. Yet when courts apply the multi-factor balancing test provided by the *Restatement*, they take on inquiries that detract from the harm that these contracts can cause.

For example, in *Transfair International, Inc. v. United States*,²¹ a contractor won a government contract to transport relief supplies to Eritrea.²² On three occasions, a subcontractor used an Iranian carrier to make the deliveries in violation of a federal regulation prohibiting the use of carriers licensed and registered in restricted nations.²³ As a consequence, the United States refused to pay the agreed amount and Transfair sued for breach of contract.²⁴ The court denied the United States' attempt to have the non-enforcement rule applied through a motion to dismiss and, instead, favored the fact-based culpability inquiry that the *Restatement* invites—which has little to do with an assessment of the contract's potential to create harm.²⁵ In

¹⁷ See, e.g., *Morelli v. Ehsan*, 756 P.2d 129, 133 (Wash. 1988); *Gerald F. Moore, P.C. v. Orthodontic Ctrs. of Am., Inc.*, No. D035808, 2002 WL 32351, at *7 (Cal. Ct. App. Jan. 11, 2002).

¹⁸ See Fichter, *supra* note 11, at 34 (“Where an agreement between lay persons and professionals is held illegal, it is not unusual for the court to leave the parties as it found them, which is often with the professional holding practice assets acquired with lay capital and the lay person holding an unenforceable debt obligation.”).

¹⁹ See *id.* at 33-34.

²⁰ See Exec. Order No. 12,959, 60 Fed. Reg. 24,757 (May 6, 1995) (describing trade prohibitions with Iran as a means to “deal with the unusual and extraordinary threat to the national security, foreign policy, and economy of the United States”).

²¹ 54 Fed. Cl. 78 (2002).

²² *Id.*

²³ *Id.* at 79.

²⁴ *Id.*

²⁵ *Id.* at 84-85 (applying RESTATEMENT (SECOND) OF CONTRACTS § 178).

doing so, the court missed an opportunity to harness the ability of the non-enforcement rule to deter potential third-party harm. Furthermore, going forward, one could expect similarly situated contractors to be underdeterred from entering into similar, and potentially harmful, contracts.

This Article proposes a more systematic approach to the remedies for illegal contracts that accounts for the risks of deterring desirable bargains and insufficiently discouraging harmful contracts. The proposal suggests a move away from non-enforcement as a presumptive remedy and would instead predicate that remedy on a showing of harm. Doing so would help to align the incentives of parties with the harm-reducing rationale of contract restrictions. But an effective remedy structure also requires an understanding of how the features of potentially illegal contracts and their remedies interact with deterrence. Accordingly, this Article analyzes how non-enforcement, restitution, and other potential remedies can affect the decision-making of parties. This analysis counsels some caution because, as one might expect, in some circumstances the parties will be able to evade the deterrent effect of a remedy by pricing the contract to account for that risk.²⁶ The Article's proposed framework accounts for this issue by highlighting the situations where courts may, and may not, be able to deter harmful contracting and by discussing the role that uncertainty can play in deterrence.

Getting the rules for illegal contracts right is particularly important because of the immutable character of the remedy regime. When legislators and courts supply an undesirable default rule, the social cost comes from the resources that parties must expend to contract around the rule.²⁷ But, unlike most other default rules and remedies, parties cannot opt out of the rules that govern the remedies for illegal contracts. In contrast to other doctrines that allow a party to walk away from a contract—such as defects with offer and acceptance, incapacity, or statute of frauds issues—the concern with illegal contracts is not the lack of cognizable bargain. Rather, the doctrines that prohibit illegal contracts can pose an inflexible barrier to an otherwise satisfactory deal. While courts often prohibit these bargains for good reasons, the inability to evade illegal contract rules places added importance on their proper selection.

In selecting the proper rules to govern illegal contracts, it is appropriate to focus on the cases that are particularly hard because they are most likely to end up in court. Agreements that are obviously illegal and impose

²⁶ See *infra* Part II.

²⁷ Robert E. Scott, *Rethinking the Default Rule Project*, 6 VA. J. 84, 94 n.4 (2003) (“Then choosing a default rule on the basis of some normative conception of fairness would be wrong, in the sense that it would not increase the amount of fair contracts in the world, but it would increase the amount of contracting costs.”). Some scholars have questioned the view of whether default rules should seek to minimize transaction costs. See Ian Ayres & Robert Gertner, *Filling Gaps in Incomplete Contracts: An Economic Theory of Default Rules*, 99 YALE L.J. 87, 91 (1989) (arguing that default rules should act as penalties in order to provide transactors with incentives to reveal their preferences).

significant harm are unlikely to end up in court because they often carry collateral consequences like criminal and civil liability. Taking these cases to court can expose the existence of the bargain to prosecutors, potential plaintiffs, and others who can impose the collateral consequences. Contracts that are only arguably illegal and do not obviously impose harm are much more likely to be litigated than a drug deal gone wrong or a dispute over a murder hit.²⁸ Compare these obviously illegal and harmful contracts to a close question in antitrust law where there is doubt whether the arrangement creates any harm—the latter is far more likely to result in litigation. Consequently, the rules governing these contracts should be directed at providing the right incentives in the close cases rather than structured around deals where the illegal contract remedy is unlikely to make a difference in the behavior of parties.²⁹

Despite the frequency with which courts analyze illegal contracts,³⁰ commentary on this issue has been sparse. To the degree that scholarship engages the problems associated with illegal contracts, it tends to focus on the effective use of non-enforcement to deter illegal bargains without analyzing the transactors' decision-making with respect to specific remedies and the effect of ambiguity on those choices. For example, one commentator argues that deterrence would be more effective if courts placed any loss on the party who is able to learn of the illegality at least cost, but she does not focus on the ability of parties to evade the deterrent effect of a chosen remedy and the "difficult legal questions" presented by the threshold question of whether a contract is illegal.³¹ Another piece suggests that there is a connection between illegal contracts and regulating externalities, but does

²⁸ Insofar as the common law will entertain enforcing an obligation that arises from an illegal contract, it is nearly always in the context of civil disputes. Contracts barred by the criminal law, however, are almost never enforced in any measure. See Neal Kumar Katyal, *Conspiracy Theory*, 112 YALE L.J. 1307, 1341 (2003) (characterizing criminal law as "a restriction on the ability to contract"). Judge Posner pointed to this inability to enforce criminal contracts as a reason that organized crime sometimes resorts to violence as a means of enforcement and why organized crime must rely on trust more than parties to legal contracts. See RICHARD A. POSNER, *ECONOMIC ANALYSIS OF LAW* 241-42 (7th ed. 2007); see also Andrew R. Dick, *When Does Organized Crime Pay? A Transaction Cost Analysis*, 15 INT'L REV. L. & ECON. 25, 28 (1995) (predicting the effects for organized crime where third-party enforcement of contracts is not available).

²⁹ The ability to keep proceedings confidential can limit the collateral consequences of the illegality by preserving the secrecy of the bargain. Arbitrating disputes will tend to enhance the ability to limit the knowledge of the existence of the contract. See *Buckeye Check Cashing v. Cardegna*, 546 U.S. 440, 449 (2006) (holding that an arbitrator should determine a claim of illegality of an allegedly usurious contract, per the terms of an arbitration clause).

³⁰ A Westlaw search in the ALLCASES database for (illegal /2 contract*) within the last ten years yields a result of 1,912 cases (search conducted on Sept. 14, 2009).

³¹ Kostritsky, *supra* note 1, at 116 n.4; see also *infra* Part IV.A for a discussion of the differences between this Article's proposal and Professor Kostritsky's approach.

not engage the incentive effects posed by specific remedies and uncertainty in a systematic way.³²

To analyze the problem of optimally deterring illegal contracts, Part I of this Article examines the incentive effects of different potential remedies and their consequences for efficiency. This discussion begins with an examination of the non-enforcement remedy's ability to deter illegal contracts and whether that ability aligns with the harm that such contracts can cause. The part also explores the potential of tying remedies to the amount of social harm they cause and why such remedies are not likely to be effective. Ambiguity and variance in the cost of performance also play an important role in the deterrent effect of different remedies and the ability of parties to evade attempts at deterrence. Part I concludes with an analysis of the effects of these variables on decision-making.

Part II analyzes the use of remedies and incentives to develop a structured framework for the analysis of arguably illegal contracts that breaks from the ad hoc nature of the current approach. This Article's proposed framework gives harm and ambiguity a prominent role in the approach to illegal contracts that is missing from the present analysis. Emphasis on these features should minimize the overdeterrence of innovative contracting and the underdeterrence of harmful contracting that present doctrine invites.

Part III centers on applied examples, beginning with more complete discussions of the examples above of overdeterrence in the corporate practice of medicine and underdeterrence in trade with rogue nations. This Part also analyzes the insurable interest requirement as an area where overdeterrence has been a persistent problem. I apply the framework developed in Part II to each of these areas and show how courts might improve their approach to these cases. Part IV engages some potential objections and concludes.

I. INCENTIVES AND ILLEGAL CONTRACTS

This Part examines the arguments that have been used to justify different remedies for illegal contracts and the relationship of those arguments to the prevention of harm. The analysis begins with a look at the non-enforcement remedy's ability to deter the formation of illegal contracts and contrasts this effect with the relative inability of restitution to inhibit contracting. To examine what these effects mean for efficiency, the model then incorporates third-party harm as a variable to examine whether any remedy—including penalties or bounties that attempt to measure the social harm—can deter illegal contracting at the optimal level. Finally, this Part analyzes how the pervasive ambiguity that accompanies many contract re-

³² See Note, *A Law and Economics Look at Contracts Against Public Policy*, 119 HARV. L. REV. 1445, 1458-59 (2006) (briefly speculating about the role that uncertainty may play in illegal contracts).

strictions can affect decision-making under the threat of non-enforcement. While this ambiguity can exert a strong deterrent effect on some contracts that the parties believe are likely to be legal, the introduction of variance can allow the parties to price the risk of non-enforcement into their bargains. This ability to evade the deterrent effect can facilitate the formation of desirable contracts because the variance dilutes the deterrent effect, but this ability can also inhibit the ability of the non-enforcement remedy to deter harmful bargains.

A. *Non-Enforcement, Restitution, and Deterrence*

Courts have, for several generations, justified the non-enforcement rule with the claim that it deters the formation of illegal contracts. The Supreme Court explained this logic over one hundred years ago in *McMullen v. Hoffman*.³³

To refuse to grant either party to an illegal contract judicial aid for the enforcement of his alleged rights under it tends strongly towards reducing the number of such transactions to a minimum. The more plainly parties understand that when they enter into contracts of this nature they place themselves outside the protection of the law, so far as that protection consists in aiding them to enforce such contracts, the less inclined will they be to enter into them. In that way the public secures the benefit of a rigid adherence to the law.³⁴

Underlying this reasoning is the standard contract model—one party acts first by providing a payment or conferring a benefit, followed by a future decision by the other party deciding to perform the promised act or breach and pay damages.³⁵ As long as there is a credible threat of enforcement,³⁶ the second party should carry out the promise when the cost of doing so does not exceed the cost of expectation damages—at least according to the conventional economic account.³⁷

³³ 174 U.S. 639 (1899).

³⁴ *Id.* at 669-70. Courts continue to apply this logic to illegal contracts. *See, e.g.*, *Mlynarik v. Bergantzel*, 675 N.W.2d 584, 587 (Iowa 2004) (“We have noted the purpose of [the non-enforcement rule] ‘is to deter future misconduct by denying relief to one whose losses were substantially caused by his own fraud or illegal conduct.’” (quoting *Gen. Car & Truck Leasing Sys., Inc. v. Lane & Waterman*, 557 N.W.2d 274, 279 (Iowa 1996))).

³⁵ *See* Mark Pettit, Jr., *Freedom, Freedom of Contract, and the “Rise and Fall”*, 79 B.U. L. REV. 263, 284-87 (1999) (discussing the pattern of initial performance followed by an obligation to perform in the future as a defining element of the institution of contract).

³⁶ The fear of court enforcement may not, of course, be the only reason that parties perform on a contract, but it is a primary rationale for the institution of contract. *See* Alan Schwartz & Robert E. Scott, *Contract Theory and the Limits of Contract Law*, 113 YALE L.J. 541, 557-59 (2003) (explaining when and why state enforcement of contracts is likely to be favored over reputational enforcement).

³⁷ *See, e.g.*, ROBERT COOTER & THOMAS ULEN, *LAW AND ECONOMICS* 201-05 (4th ed. 2004).

By taking away the ability to enforce the contract through public courts, the non-enforcement rule changes this calculus. If parties depend on the threat of judicial enforcement to support their exchange, the non-enforcement rule can act as a deterrent because the second-to-act party can breach and keep the benefit conferred without fear of having to pay any damages. As long as the first party knows of this threat, that knowledge can deter formation of the contract when the second party stands to gain more through breach than through performance.

A simple example demonstrates the dynamics of this decision.³⁸ Imagine a buyer values a service at 400 and a seller can provide it for 200. The contract calls for the buyer to make an upfront payment of the agreed price between 200 and 400, followed by the seller choice's whether to make the investment in performance. If, however, the contract is clearly illegal, and thus subject to the non-enforcement rule, the contract will not form. The seller would not invest in performance because the seller has already received payment, and, as a consequence of the non-enforcement, may keep the entire payment. If the buyer is aware of this possibility, the contract will not form because the buyer will fear a complete loss of the upfront payment without legal recourse.

This exercise shows the deterrent power of non-enforcement, but it also demonstrates some of its limitations. Many exchanges take place in the absence of the ability to enforce them through courts. Often, these agreements rely on reputation or an expectation of reciprocity as a means to ensure performance.³⁹ Even without these mechanisms, it may be possible to structure individual transactions in a manner that minimizes the deterrent effect of the non-enforcement rule. Parties may stagger payments over the course of a deal in order to ensure future performance⁴⁰ or they may exchange hostages to make commitments more credible.⁴¹ If this approach is able to sustain an agreement, the only social cost would be the additional

³⁸ For convenience and clarity I omit transaction costs from the example.

³⁹ The ability of parties to contract in the absence of formal state enforcement has been the subject of extensive study. See, e.g., Avner Greif, *Contract Enforceability and Economic Institutions in Early Trade: The Maghribi Traders' Coalition*, 83 AM. ECON. REV. 525 (1993); Avner Greif, Paul Milgrom & Barry R. Weingast, *Coordination, Commitment, and Enforcement: The Case of the Merchant Guild*, 102 J. POL. ECON. 745 (1994). For empirical work on the role of reputation in supporting exchange in communities that have opted out of state enforcement, see Lisa Bernstein, *Opting Out of the Legal System: Extralegal Contractual Relations in the Diamond Industry*, 21 J. LEGAL STUD. 115, 148 (1992); Lisa Bernstein, *Merchant Law in a Merchant Court: Rethinking the Code's Search for Immanent Business Norms*, 144 U. PA. L. REV. 1765, 1788-92 (1996).

⁴⁰ This solution would resemble the use of staged transactions to diminish the losses associated with the holdup problem. See Darwin V. Neher, *Staged Financing: An Agency Perspective*, 66 REV. ECON. STUD. 255 (1999). But see Vladimir Smirnov & Andrew Wait, *Hold-up and Sequential Specific Investments*, 35 RAND J. ECON. 386 (2004) (showing that the potential for sequential investment can disadvantage some parties, thus diminishing their incentive to invest).

⁴¹ See, e.g., Oliver E. Williamson, *Credible Commitments: Using Hostages to Support Exchange*, 73 AM. ECON. REV. 519, 519-20 (1983).

transaction costs necessary to structure the deal in this manner.⁴² But this will not be possible with all deals; some transactions require large amounts of capital prior to performance or expensive performance prior to payment.⁴³ Indeed, a prevailing rationale for state enforcement of contracts is this institution's ability to prevent opportunistic behavior.⁴⁴ To the degree that the state enforcement of contracts accomplishes this goal, the use of the non-enforcement remedy will undermine it.⁴⁵

Restitution is used as an alternative to non-enforcement as a remedy for illegal contracts.⁴⁶ This remedy returns any upfront payment made by a party prior to breach, which tends to encourage performance and thus helps to support the formation of contracts. Using the example above can illustrate how this effect operates. The buyer and seller, with reservation values of 400 and 200, respectively, would presumably agree to a price between those two values. The buyer would make the upfront payment of the agreed price and the seller would then choose whether or not to perform or breach. The restitution remedy would require the seller to return the upfront payment, so the net gain for the seller would be zero. But by performing, the seller would gain the difference between the cost of performance (200) and the price, so there will be an incentive to perform.

While restitution does much more to encourage performance, it will not provide the optimal means to do so. This remedy does not compensate for the reliance or expectation interests. So, under the conventional account, restitution creates a risk of inefficient breach.⁴⁷ But in some contexts this remedy may amount to full enforcement of the contract, or something close to it. Where a seller renders a service and the buyer refuses payment on the

⁴² See Anthony T. Kronman, *Contract Law and the State of Nature*, 1 J.L. ECON. & ORG. 5, 10-11 (1985) (arguing that, in the state of nature, the lack of state enforcement can create a need for simultaneous exchange, which will increase costs associated with contracting).

⁴³ See COOTER & ULEN, *supra* note 37, at 196-200 (explaining that a primary benefit of contract law is that it converts non-cooperative games into cooperative games).

⁴⁴ See Russell Hardin, *Trust*, in 3 THE NEW PALGRAVE DICTIONARY OF ECONOMICS AND THE LAW 623, 624 (Peter Newman ed., 1998) ("We have the law to back contracts for major exchanges, to protect us in our ordinary dealings, and to reduce the likelihood of at least some massive end-game losses.").

⁴⁵ Peter Leeson attempted to quantify the effect of enforcement on trade by examining how membership in the New York Convention ("NYC") affects the amount of trade that a country engages in. See Peter T. Leeson, *How Important is State Enforcement for Trade?*, 10 AM. L. & ECON. REV. 61, 72-73 (2008) (finding a benchmark result of a 38 percent increase in trade when both countries are members of the NYC and a 15 percent increase when one country is a member, a result that Leeson says is "modest compared to intuition about the importance of state courts, the long-term growth of trade, and the effect of many of the other variables impacting trade").

⁴⁶ RESTATEMENT (SECOND) OF CONTRACTS § 178 (1981) (providing for non-enforcement as a default remedy and providing a balancing test that, if satisfied, would permit restitution as a remedy).

⁴⁷ Andrew Kull, *Restitution as a Remedy for Breach of Contract*, 67 S. CAL. L. REV. 1465, 1497 n.83 (1994) ("Where the producer's cost exceeds the contract price but is less than the value of the goods to the buyer, a remedy limited to return of the contract price will invite inefficient breach.").

basis of illegality, a court cannot return the performed service in the same manner that a court can return an upfront payment. Instead, a court that wishes to encourage performance would have to order the buyer to pay the seller an agreed price, or perhaps a lower amount that excises some or all of the seller's profit from the service.⁴⁸ Restitution will thus tend to encourage performance, but it may not do so at the efficient level, and it may be difficult or impossible to administer in some situations.

B. *Efficiency, Harm, and the Choice of Remedies*

The analysis above shows that the choice of remedies can affect the breach decision and, in some cases, can affect whether or not parties will enter contracts in the first place. For policy makers and judges, a central question posed by the problem of illegal contracts is how to align these incentives in a way that encourages the wealth-producing benefits of some contracts without inviting the social costs produced by other bargains. The following subsections identify how non-enforcement and restitution do not always provide efficient incentives, and then analyze whether the use of penalties or bounties to internalize any potential externality can correct these deficiencies.

1. Efficiency and the Conventional Remedies

The imposition of externalities on third parties is a chief source of negative social costs from illegal contracts. These costs can vary widely; a contract to inflict bodily harm can create enormous costs for the victim while individual transactions that violate a licensing statute may pose little or no harm. The current non-enforcement and restitution remedies are not, however, sensitive to the social costs and benefits created by contracts because the application of these remedies does not depend on third-party harm.⁴⁹ This feature distinguishes these remedies from other areas of law where deterrence is a central concern. So, for example, the tort remedy for

⁴⁸ Compare Ariel Porat, *Expanding Restitution: Liability for Unrequested Benefits* 1 (Univ. of Chi. Law & Econ. Working Paper No. 388, 2008) (arguing in favor of expanding an award of restitution of this sort), available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1088796, with Oren Bar-Gill & Lucian Arye Bebchuck, *Consent and Exchange* 3 (Harvard Univ. Law & Econ. Working Paper No. 590, 2007) (arguing that restitution of this sort would permit inefficient entry by sellers and inefficient exit by buyers), available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=998486.

⁴⁹ See *infra* Part III for a discussion of the current approach to doctrine with a focus on the relevant section of the *Restatement*.

external harm typically depends on the harm inflicted by the liable conduct.⁵⁰

An example shows how this insensitivity to harm operates. Assume that: (1) courts will, on average, correctly identify those contracts that are illegal; and (2) the social harm created by a contract is invariant to its legality (i.e., some legal contracts produce third-party costs and some illegal contracts produce no externalities). Returning to the valuations from the example in the last section, assume that a seller values performance at 200, a buyer values performance at 400, and that performance of the contract would produce an externality of -300 . This latter assumption means that it would be inefficient for the contract to form because the private surplus of 200 is less than the social cost of -300 .

When there is no doubt about the illegality of the contract, and the consequent use of the non-enforcement remedy, the contract should not form due to the risk that the seller will breach.⁵¹ In this example, the outcome is efficient insofar as it deters the contract, but, of course, the non-enforcement remedy can deter any contract that it applies to, regardless of whether the contract causes any external harm. As a consequence, even where a contract creates no external harm, the contract would be deterred.

Restitution has familiar effects—the prospect of returning any benefit conferred will tend to encourage performance of the contract.⁵² Like the non-enforcement rule, the seller's decisions under a restitution rule will be invariant to the external social costs imposed by the contract. The seller will only care about the expected damages from breach, so the seller will not breach unless the costs of performance rise above the initial estimate, and even then the costs must rise above the level of expected damages before the calculus favors breach. This incentive to perform will support the formation of contracts, which means that restitution may be particularly undesirable where the expected social costs associated with contract are high because it will encourage performance.

It would improve the use of these remedies to condition the choice of non-enforcement or restitution on the basis of a contract's expected social effects. So, in the example above, non-enforcement would be preferable because the surplus created by the bargain is smaller than the negative social costs associated with the contract. But when the gain from the contract exceeds the perceived social costs, restitution would be appropriate because it tends to encourage performance. To be sure, this approach may be too blunt to work as an optimal deterrent. As noted earlier, in some instances

⁵⁰ Likewise, criminal sanctions usually vary based on the expected magnitude of the culpable conduct.

⁵¹ Again, the lack of a formal contract does not preclude exchange of some sort occurring, albeit with higher transaction costs and perhaps lower stakes. *See supra* Part I.A. If the trade does occur, there is, of course, the chance that it will produce the externality associated with the contract.

⁵² *See id.*

the non-enforcement remedy would result in the restructuring of contracts rather than their elimination. This type of rule may also suffer from insensitivity to marginal analysis in the sense that it bases the remedy on a contract's overall effects rather than on marginal investment decisions in ensuring performance or minimizing external harm. Furthermore, it may prove difficult to measure external harm because, unlike most tort actions, the harm has not occurred (as a consequence of breach of the contract).⁵³ Nevertheless, a more express focus on third-party harm in deciding whether non-enforcement will apply should improve outcomes relative to a rule—like the status quo—that is insensitive to such harm.

2. Penalties and Bounties

As an alternative to non-enforcement and restitution, it is worth exploring whether courts could improve deterrence through remedies that apply the conventional economic wisdom of obtaining more efficient results by internalizing the externality created by an illegal contract.⁵⁴ There are several potential ways to structure this type of remedy. One alternative would be for the breaching party to pay—in addition to any damages—a penalty to the court calibrated to the degree of social harm.⁵⁵

For a penalty of this sort to have an effect there must be some potential for breach because otherwise the court will not be able to learn about the contract. When the seller has fixed costs of performance, the penalty will have no consequence because the seller will perform and, accordingly, the penalty will not be imposed. But when the costs of performing a contract vary, there is a chance for the seller to consider breach, which would result in the potential assessment of the penalty. Of course, a penalty that exactly equals the social harm will not be effective because the chance of the penalty's assessment will be less than 100 percent unless the parties expect breach with certainty, in which case the contract would not form.

An example demonstrates this effect and also suggests a potential solution. Imagine the buyer values a deal at 400 and the seller expects the cost of performance to be 0 with 50 percent probability and 1,000 with 50 percent probability. The social harm from the deal is 250. Under the expecta-

⁵³ There also may be an expressive externality in that enforcing (or partially enforcing) a contract may undermine the ability to deter contracts going forward. *See infra* Part II.A.

⁵⁴ This is the conventional economic rationale for the proper remedy in tort. *See COOTER & ULEN, supra* note 37, at 347.

⁵⁵ As a general matter, doctrine and theory discourage use of extra-compensatory remedies in contract law because, among other reasons, they can discourage efficient breach. *See, e.g.,* Richard Craswell, *Contract Remedies, Renegotiation, and the Theory of Efficient Breach*, 61 S. CAL. L. REV. 629, 665-68 (1988); Richard Craswell, *Deterrence and Damages: The Multiplier Principle and Its Alternatives*, 97 MICH. L. REV. 2185, 2230 (1999) (arguing that extra-compensatory remedies can lead transactors to price this risk into the contract).

tion rule, the seller will perform in the low-cost condition and breach in the high-cost condition, so the seller's expected costs are the expected costs of performance plus the expected damages payment or $(.5 \cdot 0 + .5 \cdot 400)$ 200. In this case, the contract produces negative net social effects because the surplus from exchange, 200, is less than the social harm created by the deal, 250. But a penalty equal to 250 will allow the parties to enter into the contract because the expected cost for the seller, $(.5 \cdot 0 + .5 \cdot 650)$ ⁵⁶ 325, will be less than the buyer's reservation price of 400.

A court could try to set the penalty at the proper level by setting the penalty equal to the amount of social harm multiplied by the inverse of the chance of paying damages.⁵⁷ In this example, that penalty would be the inverse of .5, or 2, multiplied by the harm of 250 for a penalty of 500. The seller's expected costs with this penalty would be $(.5 \cdot 0 + .5 \cdot 900)$ or 450. In this case, the difference between the buyer's valuation and the seller's costs $(400 - 450)$, or -50, is exactly equal to the social cost that the exchange would create—the surplus from exchange of 200 minus the 250 social cost. Because the rule aligns the parties' costs with the social costs it creates, in theory, efficient incentives.

But a penalty of this sort is not a panacea for at least two reasons. First, the penalty will not be relevant to the seller's decision-making if the costs of performance are low enough. Altering the previous example shows how this effect can operate: imagine that the cost of performance in the previous example had a 50 percent probability of being 700 rather than 1,000. In this case, the seller, when faced with the high-cost condition, would prefer to perform rather than pay the 900 in damages, which includes the expectation measure (400) and the penalty (500). The contract would form because, while the expected costs of 350 would be less than the buyer's reservation price, the seller would rather perform inefficiently at a cost of 700 than pay damages of 900.⁵⁸

Second, and more importantly, the seller is unlikely to pay an actual penalty to the court or the government because it would be preferable to negotiate a settlement price that falls between the buyer's valuation and the penalty. So, in the example where the damages are 900 and the cost of performance in the high-cost scenario is 1,000, the buyer and seller should settle the case at a price between the 900 in damages and the buyer's valuation of 400. A penalty, thus, would not deter deals at a level that is equal to the amount of the penalty. Instead, this ability to settle the case can enable

⁵⁶ The 650 figure is the sum of expectation damages: 400, plus the penalty of 250.

⁵⁷ This approach is the typical prescription from legal economists on how to calibrate the level of punitive damages for difficult to detect torts. See A. Mitchell Polinsky & Steven Shavell, *Punitive Damages: An Economic Analysis*, 111 HARV. L. REV. 869, 874-76 (1998).

⁵⁸ One should not expect the seller to actually perform the contract inefficiently. Instead, the buyer and seller should negotiate a settlement between the buyer's valuation and the cost of the inefficient performance.

the parties to bargain around the penalty and obviate the penalty's ability to deter. Under the assumptions of the example, there is a 50 percent probability of the low-cost condition that will prompt the seller to perform, thereby allowing the buyer to receive the value of performance (400). There is also a 50 percent chance of the high-cost condition occurring, which would result in the buyer receiving a settlement between 400 and 900. Supposing that the buyer and seller expect to split the difference in dividing the gain from the settlement, the buyer would expect a settlement of 650 in the breach condition.⁵⁹ So the buyer would be willing to pay $(.5 \cdot 400 + .5 \cdot 650)$ 525. The seller expects to perform for zero half the time and to pay 650 with 50 percent probability for a total of 325. These figures produce a surplus of $(525 - 325)$ 200—the same surplus one observes under the expectation rule. Therefore, one would expect the contract to form despite the inefficiency caused by it doing more harm than good (-250).

Converting the penalty to a bounty would, unsurprisingly, have the same effect of undermining the ability to deter. A bounty would allow the buyer to receive the payment that measures the social harm as opposed to a penalty, which would be paid to the court or the government. So, in the example above, a bounty would entitle the buyer to a payment of 900 rather than just the expectation measure of 400. As a consequence, the buyer would not be willing to settle for a payment between 400 and 900 because going to court would assure a payment of 900. As with a penalty, however, the parties would be able to evade a bounty's deterrent effect by pricing it into the contract. In the working example, the buyer would expect to receive 400 half of the time and 900 half of the time for an expected value of $(200 + 450)$ 650. The seller expects performance to cost nothing with a 50 percent probability and expects to pay 900 50 percent of the time for an expected cost of 450. The surplus—as with the expectation measure and the penalty—remains 200.

These exercises show that penalties and bounties cannot be a reliable deterrent to socially harmful contracting because the parties can price the potential for these payments into their contracts. This effect occurs even if one assumes that courts could easily observe the buyer's valuation of a good or service, the seller's reservation price, and the social harm of the transaction, despite the reality that these values are likely to be costly to ascertain and prone to error.⁶⁰ It is worth pointing out, however, that under

⁵⁹ The surplus from exchange is $900 - 400 = 500$, half of which is 250. Adding the 250 to the damage award of 400 provides the 650 figure.

⁶⁰ Cf. Nathan B. Oman, *The Failure of Economic Interpretations of the Law of Contract Damages*, 64 WASH. & LEE L. REV. 829, 855 (2007) (noting that a proposed economic solution to the problem of overreliance is not feasible because it would place enormous information demands on courts); Melvin A. Eisenberg, *Actual and Virtual Specific Performance, the Theory of Efficient Breach, and the Indifference Principle in Contract Law*, 93 CAL. L. REV. 975, 986-87 (2005) (arguing against taking the promisee's valuation of the promisor's investment in precaution into account due to the difficulty of measuring this value).

the threat of a fine or a bounty, the prices that the parties will be willing to pay to make privately profitable deals that impose large social costs will rise, sometimes dramatically. As a consequence, wealth constraints may allow penalties and bounties to exert a marginal deterrent effect. Likewise, as the magnitude of a probabilistic fine increases, risk aversion⁶¹ on the part of the seller may eliminate any surplus when the potential bounty to be paid is high. But even with these caveats, the general ineffectiveness of penalties and bounties suggests that, without a criminal or civil sanction levied by a third party, the non-enforcement remedy represents the most promising means for deterring illegal contracts.

C. *Ambiguously Illegal Contracts*

An additional concern with the stern remedy of non-enforcement is that it will overdeter those contracts that have only a slight chance of being illegal. Professor Arthur Corbin expressed this sentiment in his treatise by explaining:

There is a zone within which even a well-meaning citizen can not determine whether a proposed bargain is or is not "illegal." The careful and timorous may avoid worries and possible penalties by following such advice as "When the matter is doubtful; don't"; but such a practice may unnecessarily limit his own enterprise and the general prosperity as well.⁶²

More recently, Judge Posner hinted at this problem in a case on the legality of forward contracts by noting that "if the legality of a contract cannot easily be determined in advance, that might be a factor rebutting the presumption noted earlier that illegal contracts are unenforceable."⁶³ There is something to this worry, but it may not be as dire as it appears at first blush.

The key analytic difference between the obviously illegal contract and the arguably illegal contract is the likelihood of the transaction being subject to the remedy for an illegal contract. Unlike the obviously illegal contract, the arguably illegal contract will have a less than 100 percent chance that a court will apply the non-enforcement remedy. This complicates the decision calculus because the parties must account for the probability that the non-enforcement remedy would apply if the contract were found to be illegal—call this x —and the probability that expectation damages would apply if the contract were held to be legal—given by $(1 - x)$. The decision

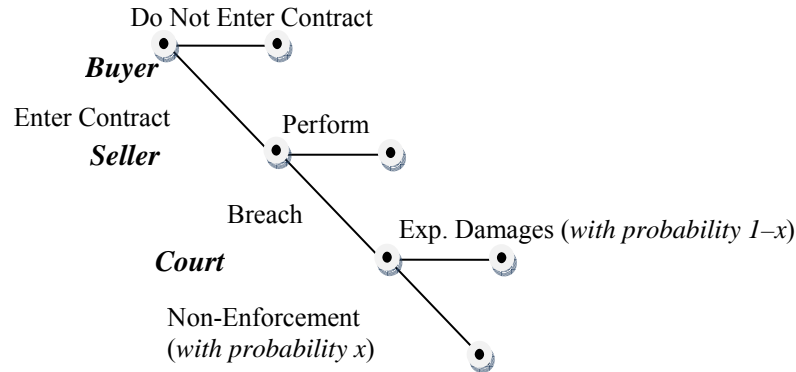
⁶¹ See Yair Listokin & Kenneth Ayotte, *Protecting Future Claimants in Mass Tort Bankruptcies*, 98 NW. U. L. REV. 1435, 1462-63 (2004) (explaining that risk aversion is the notion that people will be willing to pay money to avoid a high-risk endeavor in favor of a low-risk proposition even if the two possibilities have the same expected value). In the context of a probabilistic penalty, this desire to avoid risk can deter a party from a proposition that has positive expected value. *Id.*

⁶² ARTHUR LINTON CORBIN, CORBIN ON CONTRACTS §1374 (1952).

⁶³ *Nagel v. ADM Investor Servs., Inc.*, 217 F.3d 436, 441 (7th Cir. 2000).

tree in Figure 1 below depicts contract formation, the seller's decision to perform or breach, and the outcome in litigation. Assuming complete knowledge, the buyer can infer payoffs using expected value calculations and use backward induction to decide whether to contract in the first instance.⁶⁴

Figure 1



The next two subsections analyze how the introduction of this variable is likely to affect parties' decision-making. The discussion begins with an assumption of fixed costs of performance and then turns to an examination of how variable performance costs can change the analysis.⁶⁵

1. Contracts with Fixed Costs of Performance

By using examples with fixed costs of performance, this subsection puts into relief how the lure of appropriating an upfront payment or benefit operates in the context of contracts where doubt may exist about the legality of the bargain. The introduction of the probability that a contract is deemed to be illegal, and the corresponding probability that the contract is held to be legal, should lead the parties to think about payoffs in expected value

⁶⁴ See DOUGLAS G. BAIRD, ROBERT H. GERTNER & RANDALL C. PICKER, *GAME THEORY AND THE LAW* 53-57 (1994) (explaining how backwards induction works in the context of contract enforceability).

⁶⁵ The decision tree in Figure 1 obviously simplifies the processes at each of the stages represented. Whether to enter a contract can involve multiple decisions about how much more to negotiate. The "perform or breach" choice often requires sequential investment decisions and the litigation process involves many stages of motions, discovery, potential settlement, and appeal. The representation in the figure places the focus on how the remedy for a finding of illegality can affect earlier decisions, so, for clarity, I omit some of the complexity involved.

terms.⁶⁶ Imagine, for example, that the parties in the example above both believe the likelihood of the contract being illegal to be 50 percent, which implies that the likelihood of the contract being legal is also 50 percent.⁶⁷

Under this assumption the seller's expected value of breach would incorporate the relevant probability and the associated remedy. The seller would breach whenever the expected value of expectation damages is less than the cost of performance. Using the same numbers as the example above, 400 for the buyer and 200 for the seller, the seller would compare the expected value of expectation damages ($.5 \cdot 400 = 200$) to the cost of performance (200). The values are equal so the seller will be indifferent between performing or breaching. Once the probability of illegality exceeds 50 percent, however, the incentive to perform disappears because the expected cost of breach is less than the cost of performance.⁶⁸ Note also that breach in this example is inefficient if there is no third-party harm because the buyer's valuation exceeds the seller's cost.

Looking at this expected value calculation more abstractly reveals that the ratio of the seller's valuation of performance to the buyer's valuation of performance is a key variable for determining whether the non-enforcement rule will lead to breach.⁶⁹ When this ratio exceeds the probability that the contract will be held legal (i.e., the chance that expectation damages will apply), the seller will breach. This result makes intuitive sense; this ratio is high when the surplus is small relative to the stakes of a transaction, which means the seller stands to gain little through performance, but much through breach. One might expect to see this structure in contracts for the sale of goods, where margins can be small. Likewise, service agreements that involve keeping valuable assets embargoed for a small fee, like escrow accounts, would also tempt breach because of the large gain and small profit.

One can observe this effect by increasing the valuations of the buyer and seller to 1,000 and 800, respectively, which keeps the surplus fixed at 200. Again, the seller would compare the expected value of expectation damages ($.5 \cdot 1,000 = 500$) to the cost of performance (800). If the expected legality of the contract is 50 percent, the seller will prefer breach because

⁶⁶ See John E. Calfee & Richard Craswell, *Some Effects of Uncertainty on Compliance with Legal Standards*, 70 VA. L. REV. 965, 970 (1984) (noting that defendants faced with an imprecise legal standard will tend to think about potential liability in probabilistic terms).

⁶⁷ The remainder of this section assumes that the probability of a contract being deemed illegal is synonymous with the probability of the non-enforcement remedy applying to a finding of illegality.

⁶⁸ If the chance of the contract being illegal were 60 percent, the cost of breach would be 160 ($.4 \cdot 400$) and the cost of performance would remain 200. Hence the seller would breach.

⁶⁹ Let the value of performance to the buyer be V and the cost of performance to the seller be C with the assumption that $V \geq C$. The chance of the contract being held illegal is x , which is bounded between zero and one. With price P , the seller's payoff from performance will be $P - C$. The seller will breach if the expected value of paying expectation damages is greater than the payoff of performance, or $P - (1 - x)V > P - C$. This means that breach will occur whenever $C/V > (1 - x)$.

the expected value of renegeing—assuming that the parties split the surplus equally and set the upfront price at 900—would be $(.5 \cdot 900)$ 450 and the expected expectation damages would be $(.5 \cdot -100)$ -50 for a total of 400, while the gain from performance is only 100. As the equation in the footnote suggests, the point of indifference between breach and performance in this case occurs when there is an 80 percent chance of the contract being legal.⁷⁰

There is, however, an important caveat to this analysis.⁷¹ The high level of deterrence observed in this example is, in large measure, a consequence of the fixed cost of performance. Once one allows the likely costs of performance to vary, the parties may be able to price the contract in a manner that keeps the level of surplus intact. The next subsection describes the conditions under which this outcome could occur. The parties, however, could even accomplish this goal under the assumption of fixed costs of performance by adjusting the level of damages. If a contract stipulated damages equal to the cost of performance divided by the likely chance of illegality, the contract would still form. So, imagine a buyer's valuation of 1,000, a seller's cost of 800, and a chance of illegality of 50 percent. Using the expectation remedy, the seller would breach and the contract would be deterred, but if the parties set damages equal to $(800/.5 = 1,600)$, the seller would be indifferent between breach $(.5 \cdot 1,600 = 800)$ and performance (800). This remedy is unlikely to be permitted, however, under most articulations of the penalty rule because the stipulated damages must be equivalent to the ex ante assessment of actual damages in the case of breach.⁷²

The restitution rule may diminish the deterrence of desirable contracting that the non-enforcement regime can produce. As noted earlier, this remedy allows buyers to recoup the upfront payment through a lawsuit if the seller breaches.⁷³ Restitution thus alters the decision calculus by elimi-

⁷⁰ *Id.*

⁷¹ The caveats identified above about the actual costs of deterrence also apply. As noted, parties may be able to work around the non-enforcement remedy by restructuring deals by, for example, using staged payments.

⁷² See RESTATEMENT (SECOND) OF CONTRACTS § 356(1) (1981) (stipulating that parties may specify damages but “only at an amount that is reasonable in the light of the anticipated or actual loss caused by the breach and the difficulties of proof of loss. A term fixing unreasonably large liquidated damages is unenforceable on grounds of public policy as a penalty.”). This rule has been the subject of much criticism from legal economists. See, e.g., Aaron S. Edlin & Alan Schwartz, *Optimal Penalties in Contracts*, 78 CHI.-KENT L. REV. 33, 54 (2003) (arguing “that [Uniform Commercial Code] § 2-718 and RESTATEMENT (SECOND) OF CONTRACTS § 356 should be repealed”); Alan Schwartz, *The Myth that Promisees Prefer Supra-compensatory Remedies: An Analysis of Contracting for Damage Measures*, 100 YALE L.J. 369, 370-71 (1990) (arguing that the justifications for the penalty doctrine are unsound); Charles J. Goetz & Robert E. Scott, *Liquidated Damages, Penalties and the Just Compensation Principle: Some Notes on an Enforcement Model and a Theory of Efficient Breach*, 77 COLUM. L. REV. 554, 594 (1977) (“In sum, contemporary cost-benefit analysis suggests that the traditional penalty rule is anachronistic . . .”).

⁷³ See *supra* Part I.A.

nating any gain to the seller when the contract is held to be illegal. The seller would decide whether to breach by comparing the chance of paying expectation damages (i.e., the chance that the contract would be held to be legal) against the chance that restitution would provide zero net gain (i.e., the chance that the contract would be held to be illegal). In this example, the restitution remedy will always produce performance because the expected damages can go no lower than zero—which will be the case when there is a 100 percent chance of illegality—and, under the assumption of a fixed cost of performance, the seller will always stand to gain by following through on the promise.

2. Contracts with Varying Costs of Performance

Contrary to the assumption in the previous section, the costs of contract performance often vary and, when a seller's cost of carrying through on a promise spikes, it can create an incentive to breach. The possibility of varying costs alters the analysis substantially because it can create a welfare loss in the situation where the expectation remedy would result in performance, but the current non-enforcement remedy would produce breach. An important complication for the prevention of harm is that the uncertainty associated with enforcement and costs can allow parties to price a potentially illegal contract without any loss of surplus.

A key variable for the analysis of contracts with varying costs of performance is the expected damages the seller would pay in the event of breach. This expected value is the chance of paying expectation damages multiplied by the level of expectation damages. Assuming a model where the buyer and seller enter into a contract and the seller then finds out what the cost of performance will be, the seller will make the perform-or-breach decision by comparing the cost of performance to the expected level of damages.

An example demonstrates this concept; assume again that the buyer's valuation is 400 and the chance of illegality is 50 percent. But now the seller's costs vary and there is a 50 percent chance that performing will cost 100 and a 50 percent chance that performance will cost 300. The expected damages for a seller's breach are the chance of illegality multiplied by the buyer's valuation: $(.5 \cdot 400)$ 200. If performance costs 100, the seller will perform because this cost is less than the expected damage award, but if the cost to perform turns out to be 300, the seller will breach because the expected damage award is less than the cost. Under an award of expectation damages, of course, the seller would perform in both costs scenarios because both of these costs are less than the amount of expectation damages (400).

This prospect of breach created by the non-enforcement rule should be reflected in the price that the buyer and seller are willing to pay to enter into the contract. The reservation price for the seller would be the expected costs

under the performance and breach scenarios multiplied by their respective probabilities: $(.5 \cdot 100 + .5 \cdot 200)$ 150. The buyer's reservation price would be the expected gain under the performance and breach outcomes: $(.5 \cdot 400 + .5 \cdot 200)$ 300.

This example shows two features of the non-enforcement rule. First, if the chance of illegality stays constant, an increase in the variance of the seller's costs can allow deals to occur when they would not have occurred under constant costs. In the example in the last section, the buyer's valuation of 400 and the seller's constant cost of 200 created a situation where the parties were indifferent between the contract forming or not forming.⁷⁴ In the instant example, where the buyer's valuation is 400 and the seller's expected cost is $(.5 \cdot 300 + .5 \cdot 100)$ 200, the variation creates a positive difference between the respective reservation prices and, thus, allows for exchange.

Second, notice that the bargaining range under the non-enforcement rule, 150, is narrower than the range of 200 under the expectation rule. This narrower range has the possibility to prevent the buyer and seller from making a deal, because as transaction costs increase they may eliminate the diminished surplus.

As the variance associated with the costs of performance increases, however, the bargaining ranges align for the expectation rule and non-enforcement rule. For instance, imagine that the seller expects performance to cost 50 with a 90 percent probability and to cost 1,600 with a 10 percent probability. In this case, the seller will breach in the high cost scenario regardless of the damages rule because the cost of 1,600 exceeds the expected damage award under the expectation rule (400) and the expected damages under the non-enforcement rule (200).

In situations where the seller will breach regardless of the rule, the benefit from exchange only comes from the ability to perform at low cost. The damages rule will affect the pricing of the deal, but the bargaining range will remain the same, as the previous example shows. Under the expectation rule, the buyer expects to receive the full benefit of the bargain, 400, either through performance or through a damage award. The seller's reservation price is the expected cost of performance in the low cost scenario $(.9 \cdot 50 = 45)$ plus the expected damage award in the high cost scenario $(.1 \cdot 400 = 40)$ for a total expected cost of 85. The bargaining range in this case is $(400 - 85)$ or 315.

Under the non-enforcement rule, the buyer's expected gain under the performance scenario is $(.9 \cdot 400)$ 360 and under the breach scenario $(.1 \cdot 200)$ 20, which makes for a reservation price of 380. When costs turn out to be low, the seller will perform, which creates expected costs of $(.9 \cdot 50)$ 45 and the seller will breach when the costs are high $(.1 \cdot 200 = 20)$, so the

⁷⁴ See *supra* Part I.C.1.

total expected cost for the seller is 65. This provides the same bargaining range as the expectation rule ($380 - 65 = 315$). So, while the damages rule for illegal contracts can affect prices in high cost variance scenarios, it does not affect the size of bargaining ranges.

Unlike the case where performance costs are fixed, the restitution remedy can induce breach when the costs of performance vary. For breach to occur, the seller's cost of performance must at least rise above the contract price—the amount that the seller would have to pay in restitution—because at that point the seller would rather breach and pay this amount than perform. As with expectation damages, the degree of variance in expected performance costs drives the ex ante assessment of likely breach rather than the expected average cost of performance. The possibility of breach, however, is unlikely to deter formation of the contract. In contrast to the non-enforcement remedy, which can subject the buyer to large losses through appropriation of a conferred benefit, the worst outcome under the restitution rule is a return to the status quo for the buyer (excluding the transaction costs necessary to get there).⁷⁵ As long as there is some chance that the seller will perform rather than breach the contract, this condition should be sufficient to support the exchange because the expected gain will be positive (i.e., greater than 0 percent).⁷⁶ Of course, breach by the seller will not necessarily be efficient due to the familiar result that under-compensatory damages, such as restitution and reliance-based damage measures, can induce inefficient breach.⁷⁷

II. A FRAMEWORK FOR ANALYSIS

The analysis in the previous Part suggests three variables that are central to the choice of remedies for illegal contracts: (1) the magnitude of external harms created by the contract; (2) the presence of ex ante ambiguity as to the contract's illegality; and (3) the ability of a remedy to deter future buyers and sellers. As noted at the outset, the current doctrinal framework for illegal contracts does not expressly direct an inquiry into external harm, which can lead to the problems of overdeterrence and underdeterrence. Current law does, however, allow for the accommodation of the second and

⁷⁵ See RESTATEMENT (SECOND) OF CONTRACTS § 371 (1981).

⁷⁶ The expected gain for the buyer under a restitution rule will be the probability of performance multiplied by the expected gain plus the probability of breach times the gain or loss associated with breach. Because restitution restores the status quo, the expected gain associated with breach is zero, so the total expected gain reduces to the probability of performance multiplied by the gain from performance.

⁷⁷ See Thomas S. Ulen, *The Efficiency of Specific Performance: Toward a Unified Theory of Contract Remedies*, 83 MICH. L. REV. 341, 356-57 (1984) (explaining how restitution can induce inefficient breach).

third concerns, albeit imperfectly.⁷⁸ This Part discusses how each of these variables does, or does not, fit into the present approach to remedying illegal contracts. Furthermore, the Article proposes changes to the prevailing framework that could both limit the problem of overdeterrence and harness the ability of non-enforcement to deter in appropriate cases.

A. *The Magnitude of External Harms Created by the Contract*

The analysis in Part I shows that the non-enforcement remedy for illegal contracts can sometimes exert a powerful deterrent effect and that this deterrent effect can persist when there is even slight doubt concerning the legality of a contract. These results suggest that courts should assess an illegal contract's expected amount of harm as part of its remedy calculus. When this external harm is negligible or low, non-enforcement should be avoided because its deterrent effect will reduce little harm and can deter much beneficial contracting. In contrast, when arguably illegal contracts have the potential to create significant harm, use of the non-enforcement remedy—and its powerful deterrent effect—can be beneficial.

The *Restatement*, unfortunately, does not provide a ready mechanism for incorporating this precise concept into the law of illegal contracts. It urges courts to consider “any special public interest in the enforcement of the [illegal contract] term”⁷⁹ and “the seriousness of any misconduct involved”⁸⁰ in deciding on an appropriate remedy. While this language could plausibly incorporate the concept of harm, the phrasing is so general that it can accommodate a range of concerns that have little to do with external harm. Indeed, the application of this language tends not to focus on a direct analysis of the third-party harm a contract might cause. Rather, the “seriousness of any misconduct” element, for example, often results in an inquiry into individual culpability rather than third-party harm,⁸¹ while several courts applying the “special public interest” element cite it as a reason in favor of the enforcement of certain contracts when that public interest is lacking.⁸²

⁷⁸ See RESTATEMENT (SECOND) OF CONTRACTS § 178 (1981).

⁷⁹ *Id.* § 178(2)(c).

⁸⁰ *Id.* § 178(3)(c).

⁸¹ The full text of the element shows how the focus is on an assessment of an individual's conduct by providing that courts should determine “the seriousness of any misconduct involved and the extent to which it was deliberate.” *Id.* The focus on the deliberateness of the conduct demonstrates both how it is an individual's conduct that courts should weigh and that the extent or potential of third-party harm is not part of the inquiry that the standard invites. See *id.* cmt e (explaining the intended operation of this element).

⁸² See, e.g., *Bradley Grain Co. v. Peterson*, 267 N.W.2d 836, 839 (S.D. 1978) (contending that a lack of a “special public interest” weighed in favor of enforcement).

Despite the awkward fit with the *Restatement*, some courts have noted the importance of harm in guiding the choice of remedies for illegal contracts.⁸³ In fact, some of these courts limit the use of non-enforcement when the contract does not appear to harm the public welfare, and sometimes go as far as to fully enforce contracts despite a finding of illegality.⁸⁴ But recognition of these effects has been relatively rare and there is no sufficient analytical or doctrinal framework for urging courts to think about illegal contract remedies in this way. The occasional use of this type of analysis is not sufficient to solve the overdeterrence of desirable contracting because, from an *ex ante* perspective, transactors will struggle to predict their chances of avoiding the non-enforcement rule.

Correcting the potential problems posed by overdeterrence requires a stronger commitment by courts to limit the use of non-enforcement to those contracts that create substantial third-party harm. Judges have already showed some willingness and ability to do this analysis⁸⁵ so it does not appear to put much pressure on the judiciary's institutional competence to incorporate the analysis of harm into the scrutiny of legal contracts. There is also an analogue in the express incorporation of externality analysis in the assessment of stipulated damages clauses insofar as courts have stated a reluctance to enforce stipulated damages when the clause has the potential to inflict third-party harm.⁸⁶

One avenue for cultivating the confidence that could diminish transactors' non-enforcement fears is a strong doctrinal presumption against non-enforcement. This presumption, if adopted by statute or incorporated into the *Restatement*, would only allow the use of the non-enforcement rule where the party asserting the illegality could show that the contract has the potential to create discernable third-party harm. Such a rule would help to limit the deterrent effect associated with non-enforcement to those contracts that economic analysis suggests should be discouraged, while at the same time limiting the deterrence of contracts that can be expected to be wealth-producing.

A presumption of this sort would also help to condense the present multi-factor test in the *Restatement*. The seven elements in the prevailing

⁸³ See *Mincks Agri Ctr., Inc. v. Bell Farms, Inc.*, 611 N.W.2d 270, 292 (Iowa 2000).

⁸⁴ Courts have used this reasoning in cases where an unlicensed service provider, such as a plumber or contractor, performs a job and the recipient of the service attempts to avoid payment by claiming that the lack of a license renders the contract illegal. See, e.g., *Duncan v. Cannon*, 561 N.E.2d 1147, 1153 (Ill. App. Ct. 1990) (enforcing a contract despite a showing that a contractor did not have the required license to perform the work).

⁸⁵ See *Watts v. Watts*, 405 N.W.2d 303, 313-16 (Wis. 1987) (refusing to hold that a contract for the division of assets between unmarried cohabitants violated public policy due, in part, to the parties' legitimate expectations that it would be enforced).

⁸⁶ See, e.g., *Smelkinson Sysco v. Harrell*, 875 A.2d 188, 196 n.5 (Md. Ct. Spec. App. 2005) ("When parties are sophisticated and externalities are absent, courts do not review the parties' contractual choices for reasonableness.").

standard—most of which use vague terminology—make it difficult to predict an outcome in any particular case.⁸⁷ A presumption against the use of the non-enforcement remedy would lend more certainty to a party's prediction about likely remedies. Insofar as this presumption should decrease the probability that parties assign to the potential use of the non-enforcement remedy, this increased certainty should do much of the work to decrease overdeterrence. Furthermore, there is room for additional benefit if courts are able to develop more dependable tests for determining whether contracts cause third-party harm and when they do not.

Absent a presumption against non-enforcement, the *Restatement* should at least incorporate a more express recognition of the role that external harm should play in the enforcement of illegal contracts. Courts have already recognized external harm as a consideration in the choice of remedies for illegal contracts so this approach would not introduce an unfamiliar element to the existing test.⁸⁸ Adding this element to the test should help encourage more courts to make the connection between externalities created by a contract and the incentives created by different remedies.

One could object that the use of a presumption or element that requires a showing of third-party harm to use the most powerful deterrent available would significantly diminish the force of express restrictions on contract and the public policy doctrine more generally. After all, why would anyone adhere to contract restrictions when violating them carries no meaningful sanction? This concern is, however, not as strong as it may initially appear. The express prohibitions are a creature of statute and there is nothing to prevent a legislature from specifying a statutory remedy. Just as the current version of the *Restatement* provides, specific legislative enactments should trump a more general common-law presumption.⁸⁹

A related objection might contend that enforcing an obviously illegal bargain would create some degree of third-party harm by conveying a message that the state tolerates violation of positive law. This argument resonates with the language that some courts use to condemn entry into illegal bargains.⁹⁰ The tension between the lack of harm posed by a particular con-

⁸⁷ See RESTATEMENT (SECOND) OF CONTRACTS § 178 (1981); see also *infra* Part II.B.

⁸⁸ See, e.g., *Asdourian v. Araj*, 696 P.2d 95, 106 (Cal. 1985) (enforcing a clearly illegal agreement, *inter alia*, because there was no evident harm to the public from the contract).

⁸⁹ See RESTATEMENT (SECOND) OF CONTRACTS § 178 cmt. a (1981).

⁹⁰ Here is one rather overwrought expression of this sentiment:

It is an ancient doctrine of the common law that no court should lend its aid to enforce a contract to do an act that is illegal, or which is inconsistent with sound morals or public policy, or which tends to corrupt or contaminate by improper influences the integrity of our social or political institutions. The State, in the exercise of the police power and in the interest of the public welfare, has the undoubted right to regulate and limit the right of contract.

Howard Sports Daily, Inc. v. Weller, 18 A.2d 210, 215 (Md. 1941) (citation omitted) (denying an attempt by a firm to compel the Maryland Public Service Commission to order Western Union to provide telegraph service to a horse race information provider).

tract and the aversion to failing to punish illegal conduct sometimes arises in doctrine.

Compare, for example, two cases involving a similar problem—breach of a contract to deliver grain—where both defendants asserted illegality defenses unrelated to the formation or content of the contract. In *Mincks Agri Center, Inc. v. Bell Farms, Inc.*,⁹¹ the defendant contended that the contract was illegal because the plaintiff had relinquished its license to deal in grain prior to delivery.⁹² In *Bradley Grain Co. v. Peterson*,⁹³ the defendant sought to escape the contractual obligation via a statute that rendered illegal any contract for the sale of grain that had not been reduced to writing.⁹⁴ In both cases, there was no question that the statutes had been violated and there was no assertion in either situation that enforcement of the contract would create any harm, but the courts came to different conclusions in their application of section 178 of the *Restatement*. The *Mincks* court reasoned that allowing an unlicensed dealer to enforce a contract would create harm by undermining the ability of grain dealers to rely on the presence of a license.⁹⁵ The *Bradley* court did not share this concern; it perceived no harm to public policy in allowing enforcement of the contract.⁹⁶

While these two cases can be distinguished on other grounds,⁹⁷ each court's views on the harm created by enforcing clearly illegal contracts implies that judges and policymakers may place different weights on the expressive harm this outcome can create.⁹⁸ Given the general, inchoate nature of the perceived harm and the difficulty of empirically verifying the magnitude of any such harm, parties can reasonably disagree on the prob-

⁹¹ 611 N.W.2d 270 (Iowa 2000).

⁹² *Id.* at 271.

⁹³ 267 N.W.2d 836 (S.D. 1978).

⁹⁴ *Id.* at 836-37.

⁹⁵ *Mincks*, 611 N.W.2d at 279 (stating that to allow an exception in one case would be tantamount to holding “that a producer must deliver grain to an unlicensed dealer”).

⁹⁶ *Bradley*, 267 N.W.2d at 839 (“We further feel that there is little likelihood that a refusal to enforce the contracts will further the legislative policy.”).

⁹⁷ The most prominent distinction between the cases involves the application of all the factors set forth in the *Restatement (Second) of Contracts* section 178. Compare *Mincks*, 611 N.W.2d at 279-81 (the court's application of every factor in the test weighed against the plaintiff), with *Bradley*, 267 N.W.2d at 839 (the court's application of the test was, at best, mixed).

⁹⁸ Expressive harm has been defined alternatively as the consequentialist claim that legal expression can cause behavioral change and the non-consequentialist claim that the law takes on a moral value based on the message that it expresses. See Richard H. McAdams, *An Attitudinal Theory of Expressive Law*, 79 OR. L. REV. 339, 339 n.2 (2000). Both these definitions can accommodate the effect that I refer to here. Judges and policymakers may refuse to allow a remedy for a clearly illegal contract because—outside of any party-specific concerns—they believe that permitting a remedy would encourage illegal conduct in other contexts or they believe that the message expressed by allowing a remedy is incompatible with their view of the appropriate role for law.

lem created by enforcing a clearly illegal contract.⁹⁹ Insofar as courts view this harm as minor, a stronger showing of tangible harm would be appropriate before using the non-enforcement remedy.

But for those courts that perceive enforcement of contracts that obviously violate a statute as harmful, there are other considerations that may warrant infrequent use of non-enforcement. First, there is a plausible argument that the degree of harm from not punishing violations of contract restrictions scales with the clarity of the violation. Compare situations like *Mincks* and *Bradley* where, from an ex ante perspective, there is little doubt that the agreements violate a contractual restriction to more ambiguous situations where it can be quite difficult to tell whether a contract is illegal or not. To the degree that courts tie the harm in enforcing illegal bargains to the message that it is permissible to violate positive law, the situations that have ex ante ambiguity do not squarely implicate this concern. By applying the non-enforcement remedy in these situations, courts tell parties that when they are faced with a situation that *might* violate a restriction, harm can be done by encouraging a party to make that choice. This position takes a stronger view on the harm caused by clearly illegal contracts because it suggests that in interstitial cases the prospect for violation of the law should trump the lack of clarity that this harm will actually arise. Even courts that view the enforcement of clearly illegal contracts as harmful may believe that extension of this principle to more ambiguous contract restrictions weights this harm too strongly.

Second, there are remedies that can express displeasure with a litigant's conduct while falling short of non-enforcement. Restitution, which often serves as an alternative remedy to non-enforcement, denies the recipient the full value of the promise, which is a worse position relative to the full enforcement of a promise that expectation damages allows.¹⁰⁰ Restitution can also expose a party to a higher threat of breach because the breaching party will not have to pay as high a damage award.¹⁰¹ These adverse effects of restitution may accomplish some of the expressive effect sought by these courts without obtaining the overdeterrence that can come with a harsher remedy.¹⁰²

⁹⁹ See generally *Cevern, Inc. v. Ferbish*, 666 A.2d 17 (D.C. 1995) (discussing these issues in the context of a licensing case).

¹⁰⁰ See COOTER & ULEN, *supra* note 37, at 248-49.

¹⁰¹ Given the possibility for renegotiation, the inefficient breach threatened by the restitution remedy may not occur. Instead, the parties may negotiate a contract modification that would specify performance in return for a price that is more favorable to the party that threatens breach. This type of renegotiation, while unlikely to affect the efficient outcome, would probably alter the distribution of the surplus by shifting the threat points in a way that reduces the payoff for the party who is subject to restitution.

¹⁰² See Kevin E. Davis, *Civil Remedies for Corruption in Government Contracting: Zero Tolerance versus Proportional Liability* 34-35 (Inst. for Int'l Law & Justice Working Paper 2009/4, Apr. 17, 2009), available at <http://www.ijl.org/publications/documents/2009-4.Davis.pdf> (arguing that enforcing contracts procured through bribery may not compromise legal integrity in the manner often assumed

Note, however, that restitution will not always be a meaningful option. When a party makes a promise to make a payment or render a service in the future and the counterparty breaches before any payment has been made or a service rendered, any forfeiture will be slight. Both *Mincks* and *Bradley* involved promises for the delivery of grain on a future date at a specified price.¹⁰³ In both cases, however, the seller refused to perform, which resulted in lawsuits for breach. Because the grain had not been delivered, the buyers would not forfeit anything beyond any reliance expenditures and other incidental costs. In situations like these, restitution is not much of an option because there is nothing to return to a buyer. Thus, the ability of the restitution option to mediate between full enforcement and non-enforcement may be curtailed by the structural features of a contract. Nevertheless, there may be other options—such as compensating money spent in reliance—that convey an appropriate expressive message without overdetering at the level that the non-enforcement approach does.

B. *Ex Ante Ambiguity*

Legitimate ambiguity about a contract's illegality matters because the threat of non-enforcement can deter contracts that are likely to be lawful. When there is little chance that a contract is legal, the correspondingly small potential for the overdeterrence of lawful contracts diminishes. But as the ex ante ambiguity about the legality of a contract increases, so does the potential for undesirable deterrence.

The current doctrinal approach, at least as embodied by the *Restatement*, provides a mechanism for incorporating this concern.¹⁰⁴ The *Restatement* specifies seven elements for courts to consider when deciding if an illegal term should be enforced, including “the parties’ justified expectations.”¹⁰⁵ The associated commentary clarifies this phrase by explaining that courts should not “frustrate a party’s legitimate expectations unless there is a corresponding benefit to be gained in deterring misconduct or avoiding an inappropriate use of the judicial process.”¹⁰⁶ The *Restatement* illustrates the concept through an example of an agreement between *A* and *B* for *A* to cover *B*’s legal expenses if *B* will go onto the land of *C* to test a disputed right

because there is a distinction between holding a contract enforceable and holding that there is a legal duty to perform it).

¹⁰³ *Mincks Agri Ctr., Inc. v. Bell Farms, Inc.*, 611 N.W.2d 270 (Iowa 2000); *Bradley Grain Co. v. Peterson*, 267 N.W.2d 836 (S.D. 1978).

¹⁰⁴ Some courts have, independent of the *Restatement*, identified the need to show that illegality was “obvious or palpable” in order for non-enforcement to be appropriate. *See, e.g.*, *Transfair Int’l, Inc. v. United States*, 54 Fed. Cl. 78, 87 (2002). This requirement assesses ex ante ambiguity more directly than the *Restatement*’s inquiry into the parties’ justified expectations.

¹⁰⁵ RESTATEMENT (SECOND) OF CONTRACTS § 178(2)(a) (1981).

¹⁰⁶ *Id.* cmt e.

of way.¹⁰⁷ The commentary contends that this contract to commit this potentially legal act should be enforced even if *B*'s actions constitute a trespass because the clarification provided by *B*'s action would deter subsequent misconduct.¹⁰⁸

Courts applying this language in the *Restatement* take two distinct approaches. The first approach decouples a finding of illegality from the enforcement of the bargain. In these cases, courts acknowledge that the object of the agreement is illegal, but nevertheless permit enforcement of the contract term when it furthers the interest of protecting legitimate expectations about an agreement.¹⁰⁹ The second approach hinges the determination whether the contract is illegal (or violates public policy) on the legitimate expectations of the parties. When a contract furthers that interest, courts will conclude that enforcement of the contract would not be contrary to public policy.¹¹⁰

Either of these two approaches can incorporate the parties' ex ante beliefs about the legality of the contract in deciding on an appropriate remedy. The first approach would begin with an analysis of the technical merits of the illegality allegation followed by a determination of the appropriate remedy.¹¹¹ This latter inquiry could be informed by an assessment of the parties' belief about whether the contract was likely to be legal or not. When parties have a legitimate belief that a contract is likely to be legal, the case for non-enforcement is not as strong as the alternative because of the deterrent effect it could have on legitimate contracting. But when this belief is not as strong, the argument against non-enforcement weakens because the deterrent effect will operate on those contracts that are unlikely to be legal.

The second approach would collapse the two inquiries and have courts evaluate the parties' beliefs about legality as part of the determination whether the contract violates public policy and is, thus, enforceable or not. In this framework, a court's assessment of the parties' beliefs about legality could be used as part of the determination whether the contract violates public policy. Applying the model above, a finding that the parties could reasonably believe that the contract was legal should, all other things being equal, contribute to the conclusion that a contract does not violate public policy. But when courts determine that the parties knew, or should have

¹⁰⁷ *Id.* illus. 17.

¹⁰⁸ *Id.*

¹⁰⁹ *See, e.g.,* *Bradley Grain Co. v. Peterson*, 267 N.W.2d 836, 838-39 (S.D. 1978) (holding that a contract violated a statute that required grain contracts to be in writing, but enforcing it nevertheless due to the parties' defensible expectations that the contract would be performed).

¹¹⁰ *See, e.g.,* *Watts v. Watts*, 405 N.W.2d 303, 313-16 (Wis. 1987) (refusing to hold that a contract for the division of assets between unmarried cohabitants violated public policy due, in part, to the parties' legitimate expectations that it would be enforced).

¹¹¹ The Minnesota Supreme Court adopted this approach. *See* *Isles Wellness, Inc. v. Progressive N. Ins. Co.*, 725 N.W.2d 90, 92 (2006) ("Not every illegal contract must be voided in order to protect public policy.").

known, that the contract was likely to be illegal, that determination should weigh in favor of holding that the contract violates public policy.

One might question whether a frank acknowledgment of the law being ambiguous or that a case poses a hard issue would be desirable or plausible coming from a common law court. The second approach accommodates this concern more easily than the first because the acknowledgment of the ambiguity counts as a reason why the contract does not violate public policy. And, indeed, courts have long noticed that the concept of public policy is vague and have cited that uncertainty as a reason against its use.¹¹²

But the first approach might struggle more with this tension because courts could be put in the position of acknowledging a contract's illegality while allowing for its enforcement. This unease triggers the expressive concerns discussed in Part II.A. Again, the distinction between clearly illegal contracts and ambiguously illegal contracts may allow courts to limit use of the non-enforcement remedy based on the perceived expressive effect of enforcing the contract. Likewise, the ability to choose restitution or reliance-based damages provides a middle ground between the extremes of full enforcement and non-enforcement in a manner that conveys disapproval while attenuating the deterrent effect of the remedy for minimally harmful, but illegal, contracts.

And, finally, the practice of *stare decisis* helps to ensure that parties will not make excessive use of arguments based on *ex ante* ambiguity. Determining whether a contract is legal or not may be a difficult, indeterminate endeavor when the question is one of first impression, but once such a ruling has been made it will be a precedent in that jurisdiction. This means that the contract is no longer arguably illegal; rather, the precedent should allow other parties to recognize whether a similar contemplated bargain is illegal or not. For those that fall on the illegal side of the line, since the parties can no longer claim that there was a legitimate question whether the contract was legal, the possibility of full enforcement through expectation damages should not be an option.

C. *The Ability to Deter Contract Formation*

The analysis in this Part provided a series of predictions about the effects of non-enforcement on contract formation. When contracts clearly violate a restriction, the application of the non-enforcement remedy can be

¹¹² Richardson v. Mellish, (1824) 130 Eng. Rep. 294, 303 (Ex.) (noting that public policy is “a very unruly horse, and when once you get astride it you never know where it will carry you. It may lead you from the sound law. It is never argued at all but when other points fail.”). Courts continue to use this dictum as guidance. *See, e.g.*, Aurora Nat'l Life Assurance Co. v. Harrison, 462 F. Supp. 2d 951, 971-72 (S.D. Iowa 2006), *aff'd sub nom.* Aurora Nat'l Life Assurance Co. v. Ewing, 527 F.3d 1358 (8th Cir. 2008).

expected to exert a strong deterrent effect. In these situations, parties can be expected to either refrain from contracting or, if possible, restructure their transactions to accommodate the inability to rely on state enforcement of their bargains. The reason for this deterrence is the risk of forfeiture that the non-enforcement remedy entails. The *Restatement* provides a mechanism for accommodating this concern by specifying that courts should consider “any forfeiture that would result if enforcement were denied.”¹¹³ Some courts have even tied this forfeiture concern to the tension between overdetering beneficial contracts and underdetering harmful contracts.¹¹⁴

The analysis in the previous Part, however, suggests that courts may be able to do more to analyze the interaction of forfeiture, deterrence, and ambiguity. The analysis shows that, all other things being equal, high-stakes, low-margin contracts can be quite easily deterred, including in situations where the parties believe that the chances of the contract being held illegal are low. Courts should, accordingly, be especially wary of applying the non-enforcement remedy in these situations, to the degree that they can recognize them.¹¹⁵ Transactions that involve large payments and low profits—like some types of cost-plus contracts, escrow arrangements, and the like—could be quite susceptible to deterrence.

When the costs of performance vary, however, deterrence concerns may not be at their height.¹¹⁶ The model developed above shows that the combination of ambiguity as to enforcement and variance in costs will allow parties to price those risks into the contract.¹¹⁷ When this happens, the remedy chosen can have little effect on deterrence going forward. Therefore, before courts can treat remedies as irrelevant to future behavior, they should inquire into the knowledge that both parties have about the illegality of the contract. A disparity in knowledge can arise due to the fact that a

¹¹³ RESTATEMENT (SECOND) CONTRACTS § 178(2)(b) (1981).

¹¹⁴ See *Transfair Int'l, Inc. v. United States*, 54 Fed. Cl. 78, 85 n.8 (2002) (noting that a per se rule against enforcement of illegal bargains could overdeter contracting). The court in *Transfair* credits the *Restatement* with recognition of the concerns about overdeterrence and underdeterrence. *Id.* at 85. The *Restatement*, however, does not discuss these issues in any detail, so any contemplation of these policy issues is, at most, implicit. See *infra* Part IV.B (discussing why *Transfair* may be incorrect due to its failure to conduct a harm-based analysis).

¹¹⁵ Observing the stakes of a transaction should be simple enough for courts, but it may be more difficult to observe the margin that each party expects to receive. The gain a party makes on a transaction typically derives from the subjective cost to a party for providing that good or service. Piecing together these subjective costs from wide swathes of evidence may be quite burdensome and, moreover, parties may not be entirely forthcoming about these costs. See *supra* Part I.A. And insofar as courts are aware of the incentives provided by non-enforcement, parties may attempt to skew the evidence to create the appearance of a transaction that leads to the use of their preferred remedy. So sellers, who would probably favor non-enforcement because it allows them to keep upfront payments, might try and show that the transaction involved high margins because that result would provide an argument against use of a compensatory remedy.

¹¹⁶ See *id.*

¹¹⁷ *Id.*

party may not know the facts or is unfamiliar with the law. The factual case is relatively easy; when only one party is aware that a contract may be illegal, that party should bear any loss.¹¹⁸

The more difficult question exists when there is a difference in legal knowledge about a contemplated contract. When a repeat player has a stronger grasp on a contract's potential for illegality than another party—take, for example, an insurance company and a potential customer—the ability to price the illegality can induce over-contracting. An insurance company that knows that a life insurance contract is at risk of being declared illegal might be willing to accept a lower price than it would if the contract were clearly legal because the non-enforcement rule would provide an opportunity to avoid paying out on the contract. But customers might not have this knowledge and the lower price could induce those customers into buying a policy that they would not otherwise purchase at the higher price a fully-legal contract would warrant. The potential for these undesirable bargains may justify the *Restatement's* rule that provides for restitution when a party is “excusably ignorant . . . of legislation of a minor character.”¹¹⁹ Although, this standard does not provide much guidance for deciding what type of ignorance is excusable or how to tell whether legislation is minor. Furthermore, for those contracts that involve the low probability of a high payout, such as insurance contracts, restitution is not that helpful of a remedy because it offers a far lower award as compared to full enforcement of the contract.¹²⁰

Finally, where it appears that parties have the ability to account for the expected remedy through pricing, two conclusions follow. First, to the degree that these bargains cause harm, the difficulty in deterring them suggests that these types of contracts would be most appropriate using civil and criminal prosecutions rather than to rely on the incentive effects of illegal contract remedies. Second, because these situations do not appear to affect the ultimate allocation of contracts, courts should use a freer hand in making decisions based on the equities, if they are so inclined.

III. APPLYING THE FRAMEWORK

This Part returns to the examples in the Introduction and uses the framework developed above to assess courts' treatment of these cases.

¹¹⁸ One frequently cited case of factual ignorance is that of a woman being induced into marriage without knowledge that the man already has a spouse. Courts generally do not apply the non-enforcement rule in these circumstances. Kostritsky, *supra* note 1, at 135-37 (explaining the factual defense and citing *Ashley v. Dalton*, 81 So. 488, 488-89 (Miss. 1919)).

¹¹⁹ RESTATEMENT (SECOND) OF CONTRACTS § 198(a).

¹²⁰ Restitution would only provide a return of the premium payments, which, in the face of the insured event, are unlikely to cover the necessary costs and expenses.

These examples involve contracts where transactors could have at least some *ex ante* concern about the legality of their contract. But the expected third-party effects in the types of contracts analyzed in this section vary, sometimes substantially. The analysis below suggests how courts could address the question of third-party harm in each of these areas, while also taking account the likely ability of a remedy to affect contracting choices given the roles that surplus and variance play in affecting deterrence.

A. *The Corporate Practice of Medicine*

A wide swath of rules and regulations govern the business arrangements of professionals such as lawyers, accountants, and doctors.¹²¹ These rules typically vary by state and, given the relatively small number of professionals, there is sometimes a dearth of guidance in the form of published cases or regulatory opinions.¹²² This lack of direction can create legal uncertainty when professionals seek to create innovative ways to structure their relationships.¹²³ One doctrine in medical regulation—the restriction on the corporate practice of medicine—has given rise to a set of particularly thorny problems involving illegal contracts. This section analyzes these problems using the framework developed above.

The corporate practice of medicine doctrine curtails the contractual freedom of doctors and other medical professionals under the aegis of insulating professional judgment from a pure profit motive.¹²⁴ The borders between permissible and impermissible practices, however, are not always clear, and where courts have found violations of these agreements they have diverged in their choices of remedies. The relevant statutes generally prohibit lay-owned entities from controlling or employing medical professionals, or sharing in revenues produced by medical professionals.¹²⁵ While statutes alter some of these corporate practice restrictions in order to accommodate health maintenance organizations (“HMOs”) and employment of doctors by

¹²¹ See generally Nancy J. Moore, *Entrepreneurial Doctors and Lawyers: Regulating Business Activities in the Medical and Legal Professions*, in CONFLICTS OF INTEREST IN CLINICAL PRACTICE AND RESEARCH 171 (Roy G. Spece, Jr., et al. eds., 1996).

¹²² See, e.g., MARY H. MICHAL ET AL., CTR. TO ADVANCE PALLIATIVE CARE, CORPORATE PRACTICE OF MEDICINE 50 STATE SURVEY SUMMARY (2006), available at <http://www.capc.org/tools-for-palliative-care-programs/tools-for-hospice-hospital-relationships/legal-issues/corporate-practice.pdf>.

¹²³ See *Penny v. Orthalliance, Inc.*, 255 F. Supp. 2d 579, 581-82 (N.D. Tex. 2003) (holding illegal a business structure that had the effect of making the defendant corporation practice dentistry without a license and interpreting the meaning of a Texas restriction on dentistry without the benefit of any “legislative history, case law, or attorney general opinions”).

¹²⁴ Fichter, *supra* note 11, at 4.

¹²⁵ *Id.* at 7-8.

hospitals, many of these limitations remain on the books.¹²⁶ One commentator has called the current state of these laws “multiform, . . . porous, and . . . erratically enforced.”¹²⁷ And, given the variable outcomes that courts have come to in applying these statutes, with good reason.¹²⁸

In the medical context, many of the cases involve attempts to structure practices in a way that gives health professionals access to capital and management expertise from lay sources without running afoul of ethical prohibitions.¹²⁹ These innovations can be attractive to health professionals because devising an exit strategy is difficult when one can only sell a practice to another licensed professional or group of licensed professionals.¹³⁰ To add liquidity and management expertise to these markets, a number of lay firms have attempted to use a relatively complex set of transactions to purchase some of the assets and revenue streams from medical practices.¹³¹ The typical arrangement involves a lay firm that offers to purchase the assets and goodwill from a medical practice for cash and (sometimes) stock in the lay firm.¹³² In addition to the upfront payment, the lay firm will usually agree to handle a variety of administrative tasks, like billing and procurement, freeing the medical professional—who is now a salaried employee—to concentrate on the provision of care.¹³³ Some states expressly permit these arrangements, while others have not provided guidance as to whether or not these structures are lawful.¹³⁴

Sometimes these transactions do not work out as planned. One common problem is that the stock in the lay firm declines substantially, leaving the health professionals feeling that they gave away their practice on the cheap.¹³⁵ And, given the availability of a plausible illegality defense, it may not take a severe souring of affairs to tempt the professionals to appropriate the upfront payment. In the illegality cases that address the transactions between lay firms and doctors, orthodontists, and chiropractors, courts have reached varying conclusions and have applied varying remedies.¹³⁶ Judges

¹²⁶ Note that federally-qualified HMOs are exempt from much state regulation, including the “corporate practice of medicine” doctrine. *See* 42 U.S.C. § 300e-10(a) (2000).

¹²⁷ Fichter, *supra* note 11, at 9.

¹²⁸ *Id.*

¹²⁹ *See id.* at 10-11 (describing the benefits of the Physician Practice Management Company, one of the first wide-scale attempts to use this practice model).

¹³⁰ *Id.* at 10.

¹³¹ *See id.* at 14.

¹³² *See* Theodore N. McDowell, Jr. & Martin D. Brown, *Physician Management Companies: Just What the Doctor Ordered?*, in *HEALTH LAW HANDBOOK* 3, 4-5 (Alice G. Gosfield ed., 3d ed. 1996).

¹³³ *Id.* at 3-5.

¹³⁴ *See* Fichter *supra* note 11, at 12 (finding that ambiguities in state law can create additional regulatory hurdles for the lay firms, which they sometimes try and contract around through the formation of a professional corporation that is captive to the lay firm).

¹³⁵ *See, e.g.*, *Clower v. Orthalliance, Inc.*, 337 F. Supp. 2d 1322, 1328 (N.D. Ga. 2004).

¹³⁶ The spate of Orthalliance litigation shows the different outcomes that can occur under what are, more or less, the same set of contracts. Orthalliance signed agreements with orthodontists and other

are sometimes reticent to find these arrangements illegal.¹³⁷ But in other cases, courts have struck down the arrangements *and* have applied the non-enforcement rule, introducing substantial uncertainty and risk for the parties that might contemplate this type of transaction in the future.¹³⁸

Unlike some questions of illegality, where there may be a legitimate claim that the line between legality and illegality approximates the divide between contracts that do and do not produce wealth, it is more difficult to make that argument here. Different courts have held the same practice arrangements to be legal in some jurisdictions and illegal in others.¹³⁹ Unless one believes the unlikely scenario that the exact same practice structure can be wealth-producing in one area but not in another, it would seem that one of these jurisdictions is making an inferior choice. To be sure, this answer is not clear. While lay management offers lower cost of capital, administrative expertise, network effects, and other operating efficiencies, the sharper focus on the bottom line may encourage the provision of unnecessary care. The debate between the so-called contractual and trust-based models of optimal medical care is a long-raging one that I do not propose to answer here.¹⁴⁰

Regardless, the profusion of hard cases and divergent results suggests that there is at least some merit to the view that innovation in this area is wealth-producing—or, at least, does not create substantial harm.¹⁴¹ By applying the non-enforcement rule to arrangements that have been found to be illegal, courts create a major roadblock to potential innovation because this

dental professionals all over the country that, among other consideration, granted substantial equity stakes to individual professionals. *Id.* at 1329. When those equity stakes declined precipitously in value, a number of illegality claims ensued. *Compare id.* at 1329-30 (determining the contracts to be legal under Georgia law), *and* *Orthodontic Affiliates, P.C. v. OrthAlliance, Inc.*, 210 F. Supp. 2d 1054, 1058-61 (N.D. Ind. 2002) (determining the contracts to be legal under Indiana law), *with* *Penny v. OrthAlliance, Inc.*, 255 F. Supp. 2d 579, 582-83 (N.D. Tex. 2003) (declaring the arrangement illegal under Texas law).

¹³⁷ See *Clower*, 337 F. Supp. 2d at 1329-30; *Orthodontics Affiliates*, 210 F. Supp. 2d at 1058-61.

¹³⁸ See, e.g., *Morelli v. Ehsan*, 756 P.2d 129, 133 (Wash. 1988) (denying an accounting of an illegal partnership between a physician and a non-physician after the non-physician had expended capital to help build a medical practice); *Gerald F. Moore, P.C. v. Orthodontic Ctrs. of Am., Inc.*, No. D035808, 2002 WL 32351, at *8-10 (Cal. Dist. Ct. App. Jan. 11, 2002) (holding a business arrangement between an orthodontist and a lay firm partially illegal and unenforceable after the firm had incurred costs and expenses related to the arrangement).

¹³⁹ See *supra* text accompanying note 136.

¹⁴⁰ See generally Mark A. Hall, *Law, Medicine, and Trust*, 55 STAN. L. REV. 463 (2002) (attempting to reconcile conceptual models of medical care).

¹⁴¹ *Isles Wellness, Inc. v. Progressive N. Ins. Co.*, 725 N.W.2d 90, 95 (Minn. 2006) (stating that attempts by insurance companies to avoid payment promises by asserting that a service amounted to the illegal practice of medicine “does little to protect patients from the ‘specter of lay control over professional judgment.’” (quoting *Isles Wellness, Inc. v. Progressive N. Ins. Co.*, 703 N.W.2d 513, 517 (Minn. 2005))); see also *Cal. Physicians’ Serv. v. Aoki Diabetes Research Inst.*, 78 Cal. Rptr. 3d 646, 654 (Cal. Ct. App. 2008) (citing *Isles Wellness*, 725 N.W.2d at 95).

remedy poses a very strong deterrent to lay firms that may otherwise propose novel structures for medical practice structures.¹⁴² One of the chief promises that the lay firms get in exchange for the large upfront payments is a share of the future revenues from the practice.¹⁴³ When courts apply the non-enforcement rule after finding the contract illegal, the firm loses the upfront payment and holds an unenforceable entitlement to those future revenue streams. The prospect of this loss will obviously inhibit innovation and—to the degree firms face uncertainty about the prospects of the non-enforcement rule being applied in any particular jurisdiction—this deterrent will not be limited solely to the jurisdiction that applies the rule.

The framework set out in Part II suggests that there is a strong argument in these cases to at least provide for restitution as a remedy under certain circumstances. In many instances, the regulatory contours are vague¹⁴⁴ so parties would be justified in the belief that there is a less-than-certain probability that the contract would be deemed illegal. From an *ex ante* perspective, the transactions appear to fit the high-stakes, low-margin model in that they involve a large upfront payment, which allows for small cost savings over time. Additionally, the variance that might allow parties to price the risk in the contracts does not appear to be present insofar as medical practices tend to be stable enterprises with predictable revenue.¹⁴⁵ Accordingly, these types of investments may be imperiled by even a small chance that the non-enforcement rule will occasion one of the parties with a significant loss. Insofar as the use of restitution, or even expectation damages, can encourage these investments, it should spur innovation, which, in at least some cases, will turn out to be perfectly legal. Finally, in those cases where courts find professional arrangements illegal, these decisions will clarify the law for future parties who can no longer make a credible claim that the law was ambiguous.

¹⁴² See Lesley H. Curtis & Kevin A. Schulman, *Overregulation of Health Care: Musings on Disruptive Innovation Theory*, L. & CONTEMP. PROBS., Autumn 2006, at 195, 202 (“The corporate practice of medicine and restrictive state laws precluded the evolution of truly new models of service delivery in many states.”).

¹⁴³ See *Orthodontic Affiliates, P.C. v. OrthAlliance, Inc.*, 210 F. Supp. 2d 1054, 1056 (N.D. Ind. 2002) (explaining how the structure of the arrangement involved a large upfront payment that would be recouped by the lay firm over years of contractual rights to revenue streams).

¹⁴⁴ See *supra* text accompanying note 136.

¹⁴⁵ Note that in some cases, the transactions involve equity interests that would, of course, be subject to high variance. It is not clear from the cases examined in this Article, however, whether the parties did indeed price this risk.

B. *Trade with Rogue Nations*

A variety of statutes, regulations, and executive orders prohibit American trade and contracting with specified rogue nations.¹⁴⁶ A number of these prohibitions specifically target Iran, an effort that began in the wake of the hostage crisis of the late 1970s and continued as Iran has been perceived as a broad state-sponsor of terrorism.¹⁴⁷ Perhaps the most expansive prohibition is Executive Order 13,059, issued by President Clinton, which prohibits a wide variety of trading activity with Iran.¹⁴⁸ These restrictions on contracting and other trading stem from the express belief that “the actions and policies of the Government of Iran” present an “unusual and extraordinary threat to the national security, foreign policy, and economy of the United States.”¹⁴⁹ Or, put in the context of the framework developed in this Article, these prohibitions seek to prevent contracts because allowing transactions with Iranian interests has the potential to create or facilitate third-party harm.

These trading prohibitions with Iran have been extended to government contractors and any subcontractors they use to perform services.¹⁵⁰ When contractors violate these prohibitions the U.S. government has sought non-enforcement of the contract on the basis of illegality, as occurred in *Transfair International, Inc. v. United States*.¹⁵¹ Transfair contracted with the U.S. Agency for International Development (“USAID”) to deliver relief supplies to Eritrea.¹⁵² The company, in turn, subcontracted with a British company to perform the services.¹⁵³ The British company apparently intended to use a Ukrainian carrier to fly the supplies to Eritrea, but due to an alleged restriction imposed by the government of the Ukraine, an Iranian-flagged carrier ended up delivering the supplies over three trips.¹⁵⁴ USAID contacted Transfair after the first and the second trips to inquire into the flag of the carrier and, on both occasions, Transfair indicated to the gov-

¹⁴⁶ The statutory authority for these regulatory prohibitions includes the Trading with the Enemy Act of 1917, 50 U.S.C. app. §§ 1-44 (1917) and the International Emergency Economic Powers Act, 50 U.S.C. §§ 1701-1707 (1977). For a general discussion of these prohibitions, see Stanley J. Marcuss, *Grist for the Litigation Mill in U.S. Economic Sanctions Programs*, 30 LAW & POL’Y INT’L BUS. 501 (1999).

¹⁴⁷ See, e.g., Exec. Order No. 12,957, 60 Fed. Reg. 14,615 (Mar. 15, 1995) (prohibiting U.S. involvement in Iranian oil production); Exec. Order No. 12,959, 60 Fed. Reg. 24,757 (May 6, 1995) (prohibiting importation of most Iranian goods).

¹⁴⁸ Exec. Order No. 13,059, 62 Fed. Reg. 44,531 (Aug. 19, 1997).

¹⁴⁹ *Id.*

¹⁵⁰ Restrictions on Acquisitions of Supplies or Services from Prohibited Sources, 48 C.F.R. § 25.701 (2008); 31 C.F.R. § 560.701 (2008); 22 C.F.R. § 228.03 (2008).

¹⁵¹ 54 Fed. Cl. 78 (2002).

¹⁵² *Id.* at 79.

¹⁵³ *Id.*

¹⁵⁴ *Id.*

ernment that a Ukrainian carrier had been used.¹⁵⁵ After the third flight left for delivery, however, Transfair informed USAID that, in fact, the carrier had been Iranian on all three trips.¹⁵⁶ Transfair subsequently billed USAID for the agreed amount of \$258,720, but USAID refused to pay.¹⁵⁷ Transfair sued for breach of contract and the government defended on the grounds of illegality.¹⁵⁸

The government filed a motion to dismiss, which asserted that the non-enforcement remedy should apply.¹⁵⁹ The court used section 178 of the *Restatement* to guide its analysis¹⁶⁰ and, in doing so, demonstrated how that framework can lead courts to focus on issues that are collateral to the policy aims of contract restrictions. After reviewing the elements of section 178 and its commentary, the court identified three factors that should guide the analysis: (1) the relative culpability of the parties; (2) whether allowing forfeiture would aid deterrence; and (3) the proportionality of the use of non-enforcement to the illegality.¹⁶¹ Since the assessment of each party's culpability would require the development of the factual record, the court denied the motion to dismiss.¹⁶²

Given that section 178 does not expressly invite a harm-based rationale, it is not surprising that the court's synthesis of the test does not focus on this factor. Instead, by elevating the concern about culpability, the court complicates an analysis that could be made simpler and more effective by looking to deterrence concerns. Culpability—in the sense of knowing whether a contract restriction was being violated¹⁶³—is not a helpful inquiry because the remedy affects the incentives to invest in compliance with the contract restrictions. A remedy like non-enforcement provides much stronger motivation to ensure that subcontractors do not use Iranian carriers as compared to a restitution-like remedy because a transactor stands to lose much more through non-enforcement's imposed forfeiture. Using culpability to determine the remedy gets the analysis backwards; the remedy will tend to determine the precaution that parties invest in compliance with contract restrictions. That remedy should depend, at least in part, on the potential for violation of a restriction to cause third-party harm.¹⁶⁴ In the case of

¹⁵⁵ *Id.*

¹⁵⁶ *Id.*

¹⁵⁷ *Transfair Int'l*, 54 Fed. Cl. at 79.

¹⁵⁸ *Id.* at 80-81.

¹⁵⁹ *Id.*

¹⁶⁰ *Id.* at 84.

¹⁶¹ *Id.* at 85.

¹⁶² *Id.* at 87.

¹⁶³ *Transfair Int'l*, 54 Fed. Cl. at 85 (as the court explained its proposed inquiry, the elements include “the culpability of the promisee, including what it knew about the illegality . . . [and] the corresponding culpability of the promisor, including whether it knew about the illegality prior to the completion of performance”).

¹⁶⁴ *See supra* Part I.B.1.

trading prohibitions, the express motivation is to prevent the harm to third-party interests that trading with rogue nations can have.¹⁶⁵ This is the type of situation that, all other things being equal, calls for use of the non-enforcement remedy. By allowing this fact-based culpability inquiry to preclude a motion to dismiss, the use of section 178 dilutes the deterrent power of non-enforcement and frustrates the goals of the prohibition by allowing for underdeterrence of the perceived harm.

The other elements identified by the court are less objectionable. The inquiry as to whether or not forfeiture would aid deterrence mirrors the third element identified in Part II.¹⁶⁶ Since the court concluded that the culpability analysis would require fact finding,¹⁶⁷ it did not examine this element in any detail. But it is straightforward enough to tell that the non-enforcement remedy should act as a substantial deterrent because it would inflict a large loss on a contractor that provided valuable services. In contrast to some of the situations modeled in Part I, the government is unlikely to try and bargain around the deterrent measure because, at least in part, it has an interest in avoiding this third-party harm. Thus, the re-pricing concerns are not relevant to the analysis here.

The court's proportionality inquiry could be potentially troublesome, but the court couched this analysis as a comparison to the civil and criminal penalties available for violations of the restriction. To the degree that these penalties stand as a proxy for an activity's expected external harm, they may be a useful gauge for the harm-based analysis that I propose. Policy-makers, however, may not always perceive criminal and civil sanctions as this sort of substitute, particularly when the use of the non-enforcement remedy is likely to prove the most effective deterrent. In any event, a direct inquiry into harm would be preferable to the potential imprecision invited by a proportionality analysis.

Other courts have reached the proper result in cases concerning trade with rogue nations, but the reasoning has diverged from the focus on harm, as a relatively recent Ninth Circuit case, *Bassidji v. Goe*,¹⁶⁸ illustrates. The case involved a deal between a Hong Kong company and an Iranian company to arrange for the harvesting of brine shrimp eggs from a lake in Iran.¹⁶⁹ Securing the necessary licenses and permits for the harvesting required significant payments to the Iranian government and a representative of the Iranian company, Mr. Arshian, executed several guarantee checks

¹⁶⁵ One may, of course, question whether bans on trade effectively accomplish the goal of deterring perceived third-party harm. But insofar as legislative and executive commentary on restrictions voices that motivation, this expression should be sufficient for courts, who—as a general institutional matter—are not well equipped to do the evidence gathering that would be necessary to refute these claims. *See, e.g.*, Tex. Natural Res. Conservation Comm'n v. IT-Davy, 74 S.W.3d 849, 854 (Tex. 2002).

¹⁶⁶ *See supra* Part II.C.

¹⁶⁷ *Transfair Int'l*, 54 Fed. Cl. at 87.

¹⁶⁸ 413 F.3d 928 (9th Cir. 2005).

¹⁶⁹ *Id.* at 930-31.

related to these fees.¹⁷⁰ An American affiliated with the Hong Kong company, Goe, promised to reimburse Arshian up to an amount of almost \$1.9 million for funds spent on the required fees.¹⁷¹ In reliance on Goe's guarantees, Arshian made some of the required payments to the Iranian government and subsequently requested reimbursement from Goe.¹⁷² Goe refused to pay and, according to Arshian, this prevented the Iranian company from raising sufficient capital to obtain the licenses.¹⁷³ A lawsuit followed and Goe defended on the ground that the contract to reimburse Arshian was illegal due to Executive Order 12,959 (superseded by 13,059).¹⁷⁴

Executive Order 13,059's restrictions include "any transaction or dealing by a United States person, wherever located, including purchasing, selling, transporting, swapping, brokering, approving, financing, facilitating, or guaranteeing, in or related to . . . goods or services of Iranian origin or owned or controlled by the Government of Iran."¹⁷⁵ Because Goe's guarantee related to the harvesting of Iranian goods,¹⁷⁶ the court held that the contract was illegal and determined that the non-enforcement remedy was appropriate.¹⁷⁷

The rationale for the choice of the non-enforcement remedy, however, was orthogonal to any harm-related concerns. The court read a governing case, *Kaiser Steel Corp. v. Mullins*,¹⁷⁸ for the proposition that courts may order relief that "does not seek directly to order illegal activity,"¹⁷⁹ but may not order a party to commit an illegal act. The court reasoned that ordering Goe to pay Arshian would be tantamount to enforcing an illegal guarantee.¹⁸⁰ But reading an award of money damages arising out of an illegal contract as the equivalent of ordering an illegal act limits, quite sharply, the ability of courts to provide any remedy for illegal contracts.

¹⁷⁰ *Id.* at 931.

¹⁷¹ *Id.*

¹⁷² *Id.*

¹⁷³ *Id.*

¹⁷⁴ *Bassidji*, 413 F.3d at 931. Arshian sold his rights under the guarantees to the named plaintiff. *Id.* According to the respondent's opening brief, Arshian was apparently imprisoned for his failure to pay all of the required fees. *See* Brief of Plaintiff-Respondent, *Bassidji v. Goe*, 413 F.3d 928 (9th Cir. 2002), 2002 WL 32290829.

¹⁷⁵ Exec. Order No. 13,059, 62 Fed. Reg. 44,532 § 2(b)(ii)(d) (Aug. 21, 1997).

¹⁷⁶ This was a straightforward application of the language of the order, but some of the other provisions might present more difficult problems that would implicate ambiguity concerns. So, for example, the prohibitions on trading "indirectly" with Iranian interests might prove difficult to define. *See id.* § 2(a)(i), (b)(i), (d)(ii); *see also* Marcuss, *supra* note 146, at 504-06 (describing some of the difficult problems of interpretation that trading prohibitions present). Of course, in this instance, the deterrent power of the combination of ambiguity and non-enforcement may be a desirable feature.

¹⁷⁷ *Bassidji*, 413 F.3d at 939-40.

¹⁷⁸ 455 U.S. 72 (1982).

¹⁷⁹ *Bassidji*, 413 F.3d at 936.

¹⁸⁰ *Id.* at 939.

The *Restatement's* approach, which contemplates recovery even in some situations where a contract is illegal,¹⁸¹ and the cases that fully enforce awards where contracts have been held to be illegal,¹⁸² suggest that other courts do not read an award of money damages as broadly. And this is probably correct; there is a difference between commanding an illegal act through specific enforcement and issuing a judgment that one party has a monetary obligation to another. The money judgment is an obligation to pay another person, which is not an illegal act, unlike an order of specific enforcement, which would obligate the subject of the order to act illegally. The rule applied in *Bassidji*, accordingly, runs the risk of unnecessarily limiting the use of non-enforcement. While this effect may not matter in cases like *Bassidji*, where there is a strong harm-based rationale for non-enforcement,¹⁸³ extending this rule to cases where the harm is not as apparent runs a strong risk of overdetering potentially illegal contracts.

C. *Wagering Contracts and Insurable Interests*

The insurable interest requirement has roots in a 1774 statute passed by the British Parliament.¹⁸⁴ That statute, according to some histories, was a response to the unsavory practice of taking out life insurance policies on those being tried for capital crimes and on elderly celebrities—which until the statute, was a state-sanctioned form of wagering.¹⁸⁵ American common law courts picked up this trend in the nineteenth-century and required that the beneficiary of the contract have an “insurable interest” in the insured life—a requirement most states have now incorporated into their insurance codes.¹⁸⁶ Insurance contracts lacking the requisite “insurable interest” are illegal and, in many cases, unenforceable.¹⁸⁷ This regulatory approach is thought to further two primary goals. First, it limits the ability to engage in

¹⁸¹ See RESTATEMENT (SECOND) OF CONTRACTS § 178 illus. 4, 5 (1981).

¹⁸² See, e.g., *Asdourian v. Araj*, 696 P.2d 95, 107 (Cal. 1985).

¹⁸³ Note that *Bassidji* is probably a case of a low-variance contract that involves high-stakes and relatively low margins insofar as securing the permit would assure operation of the project without much risk. This situation creates a strong potential for the non-enforcement remedy to deter because Goe may have feared appropriation if he made good on the guarantee. See *Bassidji*, 413 F.3d at 935. Other cases, however, involve more speculative investments where it may be more difficult to deter insofar as parties can price the risk of the project and of non-enforcement. See *Kashani v. Tsann Kuen China Enter. Co.*, 118 Cal. App. 4th 531 (Ct. App. 2004) (refusing to enforce obligations under a contract to invest in a manufacturing plant in Iran because Executive Order 13,059 rendered the contract illegal and unenforceable).

¹⁸⁴ See Peter Nash Swisher, *The Insurable Interest Requirement for Life Insurance: A Critical Reassessment*, 53 DRAKE L. REV. 477, 481-82 (2005) (recounting the history of the insurable interest requirement).

¹⁸⁵ *Id.*

¹⁸⁶ *Id.* at 482-83.

¹⁸⁷ *Id.* at 483.

speculation, which, as evidenced by its historic roots, has been viewed as undesirable.¹⁸⁸ Second, this regulatory approach prevents some degree of moral hazard by allowing parties to insure only against those contingencies that they have an independent reason for preventing.¹⁸⁹

While these goals may be clear, the legal definition of “insurable interest” is a notorious tangle. The Supreme Court identified the problem long ago in *Warnock v. Davis*,¹⁹⁰ stating: “It is not easy to define with precision what will in all cases constitute an insurable interest, so as to take the contract out of the class of wager policies.”¹⁹¹ And there has not been much progress on this front in the nearly 130 years since this observation. More recent evaluations maintain that the application of the doctrine is “erratic, ambiguous, and inconsistent.”¹⁹² And, to return to the stylized example set out in the introduction, commentators have observed that much current law is unclear on whether or not domestic partners have an insurable interest in each other’s lives.¹⁹³

It is not surprising that courts struggle to apply the insurable interest doctrine. The prevailing tests ask them to assess whether there are reasonable grounds to believe that the insured has an interest “either pecuniary or of blood or affinity” in the well-being of the insured life or property.¹⁹⁴ Judgments about the reasonableness of a pecuniary interest, the proximity of a blood relation, or the depth of affinity obligate courts to draw lines without clear guideposts. This task understandably leads to ex ante uncertainty about whether a given insurance contract would be legal or not.

This uncertainty, when coupled with the application of the non-enforcement rule, can serve as a significant deterrent to the formation of a contract. This deterrent’s power derives, in large measure, from the high variance associated with insurance contracts. The application of the non-enforcement rule means that when the stated contingency takes place, the beneficiary will not receive any compensation. Regardless of whether the insured event takes place, all paid premiums would be for naught because the beneficiary would not receive anything at all. Restitution—a return of all premiums paid—presents a similarly disinvolving proposition. All the paid premiums will be lost if the stated event does not pass and, if the event does occur, the beneficiary will get the premiums back, but will still have to

¹⁸⁸ *Id.*

¹⁸⁹ See ERIC MILLS HOLMES & JOHN ALAN APPLEMAN, APPLEMAN ON INSURANCE LAW AND PRACTICE § 3.6 (2d ed. 2009).

¹⁹⁰ 104 U.S. 775 (1881).

¹⁹¹ *Id.* at 779.

¹⁹² Jacob Loshin, Note, *Insurance Law’s Hapless Busybody: A Case Against the Insurable Interest Requirement*, 117 YALE L.J. 474, 487 (2007).

¹⁹³ *Id.* at 507-08 (questioning whether domestic partners qualify as an insurable interest).

¹⁹⁴ *Warnock*, 104 U.S. at 779. There is a dispute in the debate over insurable interests in property about whether or not one must have a legal interest in property to insure it or merely a factual interest. See Loshin, *supra* note 192, at 486.

shoulder the loss for the putatively insured event. This is an attractive situation for the insurance provider, who gets an option on the legality of the underlying contract, because it is, ironically, the insured who must bear the risk of this legality determination.¹⁹⁵ Given the likely asymmetry in knowledge between the parties, the insurer may be able to price into the contract the potential for illegality without the buyer's knowledge that the reduced price reflects a chance that forfeiture will occur.¹⁹⁶

The perverse incentives created by the insurable interest doctrine have led commentators to heap criticism upon it and even call for its abolition.¹⁹⁷ The analysis of remedies for arguably illegal contracts suggests an approach that commentators have thus far ignored—enforcing contracts where there is legitimate ambiguity about the insurable interest requirement with full expectation damages. This approach appears to be warranted under the framework developed above. The fact that a buyer wants to purchase insurance for a particular contingency and an insurance provider is willing to enter into a contract strongly suggests that the risk shifting provided by the contract creates wealth—at least for both parties to the contract. Of the policy concerns that motivate the doctrine—preventing speculation and limiting moral hazard¹⁹⁸—the wagering rationale does little to justify the strong medicine of non-enforcement or restitution. At worst, this sort of speculation is slightly wealth-reducing if one assumes that there is a small dead-weight loss necessary to administer transactions that have an expected payoff of zero. But the speculation rationale has long been derided by courts and commentators as a misguided attempt to purge risk and uncertainty from contract¹⁹⁹—a goal that undermines one of the fundamental motivations for the use of enforceable contracts.²⁰⁰

¹⁹⁵ The cost of this option is an open question. To the degree both parties are knowledgeable of the legal indeterminacy underlying the contract, the premiums should be priced accordingly. But there may be information asymmetries that interfere with the identification and pricing of this option, which may reduce the option's price for the insurer. Loshin, *supra* note 192, at 495-96 (characterizing the insurable interest as "an option to breach without paying damages" and speculating that these information asymmetries probably do exist).

¹⁹⁶ One potential source of inefficiency is entry into insurance contracts by customers who would not purchase a policy if they were aware of the likely benefits once those benefits have been discounted by the chance for illegality.

¹⁹⁷ See Loshin, *supra* note 192, at 495-96 (calling for the abolition of the insurable interest doctrine).

¹⁹⁸ HOLMES & APPLEMAN, *supra* note 189, § 3.6.

¹⁹⁹ As early as 1905, Justice Holmes recognized that attempting to regulate troublesome speculation was likely more trouble than it was worth, observing that:

People will endeavor to forecast the future, and to make agreements according to their prophecy. Speculation of this kind by competent men is the self-adjustment of society to the probable. Its value is well known as a means of avoiding or mitigating catastrophes, equalizing prices, and providing for periods of want. It is true that the success of the strong induces imitation by the weak, and that incompetent persons bring themselves to ruin by undertaking to speculate in their turn. But legislatures and courts generally have recognized that the natural evolutions of a complex society are to be touched only with a very cautious hand, and that

That leaves the moral hazard rationale. By reducing the cost of a loss that the insured would bear if harm came to the person or property specified in the policy, the attendant loss of precaution imparts an undeniable negative externality.²⁰¹ But it is unlikely, for several reasons, that the harm created by ambiguous contracts eclipses the diminished contracting that results from the potential application of under-compensatory remedies to these contracts. First, the hypothesized harms that underlie the insurable interest requirement tend toward the sinister, such as murder and arson, which the criminal law already deters with significant force. The marginal gain in deterrence added by the threat of non-enforcement is unlikely to figure significantly in the mind of a potential wrongdoer given that the criminal penalties tend to be much more severe.²⁰² One may retort that moral hazard could result, not in overt criminal acts, but in reduced investments in precaution to protect the insured person. One example is a corporation that takes out life insurance policies on its employees and invests less in safety as a consequence.²⁰³ But these types of responses carry collateral consequences as well; tort liability and administrative fines can naturally follow decisions to invest less precaution in accident prevention. Given this other liability, there is reason to doubt that the insurable interest requirement shifts the precaution incentives closer to their optimal point.

Second, the ambiguous cases tend to involve a relationship between the insured and the object of the insurance contract that involves *some* connection between the two. Indeed, the gravamen of the issue in these cases is often whether there is enough affinity, blood, or interest in that connection to warrant the finding of an insurable interest.²⁰⁴ The existence of at least some relationship between the insured and the object of the contract should diminish the moral hazard problem that worries courts.

The lack of a compelling rationale for the harsher remedies of non-enforcement and restitution in the marginal insurable interest cases provides

such coarse attempts at a remedy for the waste incident to every social function as a simple prohibition and laws to stop its being are harmful and vain.

Bd. of Trade of Chi. v. Christie Grain & Stock Co., 198 U.S. 236, 247-48 (1905).

²⁰⁰ See Roy Kreitner, *Speculations of Contract, or How Contract Law Stopped Worrying and Learned to Love Risk*, 100 COLUM. L. REV. 1096, 1111-13 (2000) (explaining contract law's shift away from the regulation of speculation).

²⁰¹ HOLMES & APPLEMAN, *supra* note 189, § 3.6.

²⁰² Note that in this example, the parties do not have a shared interest in the secrecy of the contract, so the potential for strong collateral consequences will not give both the parties a strong incentive to hide the contract's existence.

²⁰³ This phenomenon has given rise to commentary defending the use of insurable interest doctrine to prevent corporations from purchasing life insurance on its employees. See, e.g., Michael J. Henke, *Corporate-Owned Life Insurance Meets the Texas Insurable Interest Requirement: A Train Wreck in Progress*, 55 BAYLOR L. REV. 51, 77 (2003); George Steven Swan, *Law and Economics of Company-Owned Life Insurance (COLI): Winn-Dixie Stores, Inc. v. Commissioner of Internal Revenue*, 27 S. ILL. U. L.J. 357, 357 (2003).

²⁰⁴ See Loshin, *supra* note 192, at 484.

additional support for the use of expectation damages. Indeed, courts may be tempted to fudge the analysis in close cases so that they can award expectation damages when it appears warranted despite the lack of cognizable insurable interest. But why not acknowledge the illegality through analysis that maintains fidelity to statutory language and precedent and award expectation damages anyway?²⁰⁵ This approach preserves the clarity of the underlying doctrine, supports the formation of marginal contracts that have a substantial probability of being held legal with hardly any credible fear of negative externalities.

IV. OBJECTIONS

This Part engages two important objections before concluding. The first considers whether or not the reorientation of the framework I propose here goes far enough. One might argue that it would be preferable to eliminate the public policy doctrine altogether or that the interpretation of illegality doctrines should be driven by economic analysis instead of just the approach to remedies. The second objection discusses the role of knowledge of illegality and whether the role of culpability that I downplay should remain an important consideration.

A. *Why Target Damages?*

An alternative to changing the damages framework would be to factor harm and ambiguity determinations into the initial assessment of whether or not a contract is illegal. But this approach is likely to be a complement to, rather than a replacement for, altering the damages regime.

Using harm and ambiguity in the illegality determination is most attractive in the context of the rule prohibiting contracts that are against public policy. As the discussion of that rule in Part II notes, some courts treat the application of the determination of whether the public policy rule applies as a test of whether the non-enforcement rule should apply.²⁰⁶ This approach collapses the illegality determination and the remedy determination into a single inquiry. As long as this analysis takes into account the

²⁰⁵ Some states have obtained a similar result through the use of statutory non-contestability rules, which prohibit an insurer from contesting the legality of an insurance policy after a specified time period of collecting premiums. New York, for example, requires life insurance policies to include a prohibition on contesting the policy after the policy has been in effect for two years. *See* N.Y. INS. LAW § 3203(a)(3) (2008); *see also* *New England Mut. Life Ins. Co. v. Caruso*, 535 N.E.2d 270, 273 (N.Y. 1989) (enforcing the non-contestability provision in a life insurance policy to bar a claim that the insured lacked an insurable interest).

²⁰⁶ *Watts v. Watts*, 405 N.W.2d 303, 316 (Wis. 1987) (uniting the merits determination and the remedy determination).

harm and ambiguity concerns that can distort contracting incentives, it is interchangeable with the framework proposed here.

But matters are more complicated when the source of the illegality is a statute or a regulation. In these instances, it may be more difficult to accommodate harm and ambiguity concerns into the analysis of whether the contract is illegal because constraints on interpretation may preclude taking these concerns into account. These constraints pose more of a problem as the relevant language becomes more specific—or, put another way, when the source of law is more rule-like than standard-like. When the contract restriction uses a vague standard rather than a narrow rule, there is more room for purposive interpretation by courts.²⁰⁷ This additional latitude could allow courts to incorporate harm-based and ambiguity-based concerns. Indeed some courts resolved ambiguous questions about medical practice restrictions to infer a harm-prevention rationale.²⁰⁸ But when the contract restriction does not allow for as much room in interpretation, such as the specific trading prohibitions in Executive Order 13,059 or the requirement that a contractor have a license, it may be more difficult for a court to incorporate harm-based and ambiguity-based concerns because the reasonable bounds of the language do not permit these readings.²⁰⁹ When the question is whether a transaction constitutes a guarantee or not, an inquiry into the harm produced by the transaction has little bearing on the application of the word transaction to the underlying behavior.²¹⁰ These types of contract restrictions could drive a court to the conclusion that a contract is illegal even though there may be very strong reasons not to apply the non-enforcement rule. In these situations, the expressed willingness to impose a lesser punishment via the remedy is critical for obtaining the proper result and for supplying proper contracting incentives.

Using the remedy rather than the initial determination to incorporate incentive concerns also allows for more flexibility and fine-tuning, which facilitates more accurate calibration of incentives. The analogy here is to

²⁰⁷ See John F. Manning, *Textualism and the Equity of the Statute*, 101 COLUM. L. REV. 1, 107-08 (2001) (arguing that law-making bodies can provide for “flexible, purposive interpretation” by using standards rather than rules).

²⁰⁸ See *Berlin v. Sarah Bush Lincoln Health Ctr.*, 688 N.E.2d 106, 113 (Ill. 1997) (refusing to invalidate employment agreements under the corporate practice of medicine doctrine where the statute did not contain an express rule against corporate practice and where “concern for lay control over professional judgment is alleviated” by structure of the contract).

²⁰⁹ See Gregory E. Maggs, *Karl Llewellyn’s Fading Imprint on the Jurisprudence of the Uniform Commercial Code*, 71 U. COLO. L. REV. 541, 553 (2000) (“Although disagreement exists over the difference between rules and standards, commentators typically distinguish them in the following manner. Rules generally define the permitted or prohibited conduct with precision, leaving the courts to determine only what happened. Standards, by contrast, usually require courts to decide not only what happened, but also to some extent what the law should permit and what it should not.” (footnote omitted)).

²¹⁰ Precedent could be an additional constraint of this type. Prior case law could, for example, limit the types of reasons that can be used in determining whether a contract is illegal.

criminal law where the split between the liability determination and the choice of the penalty is uncontroversial.²¹¹ Separating these two inquiries allows a court more latitude to adjust the incentive effects of the penalty because the penalty phase usually provides a range of punishments that vary in their severity. Uniting the illegality determination with the remedy determination would limit the ability of courts to act in the flexible manner that comes with the separation of the two. If a court has serious concerns about the legality of a contract but, based on a concern for incentives, determines that the contract is legal, the court can do nothing to adjust the remedy because a determination of legality locks in either the default remedy or the remedy specified by the parties to the contract. In contrast, decoupling the two remedies would allow the court to accommodate the incentive concerns by determining the contract is illegal and adjusting the remedy accordingly.

B. *Knowledge and the Role of Culpability*

As part of my critique of the current approach to illegal contract regulation, I question the role that culpability plays in the analysis. For example, I argue that the culpability inquiry in *Transfair* was a distraction from the role that the potential for third-party harm should have played in that case.²¹² But it is worth asking what role, if any, culpability should play in the analysis of illegal contracts because this inquiry often focuses on whether a party knew or should have known that a contract was illegal. Some of these questions are straightforward, such as in cases where a party has unilateral knowledge of the illegality. In these cases, courts should allocate the loss to a party who failed to disclose this knowledge because those disclosures can change a party's view on whether the contract is desirable.²¹³ Questions about how to treat parties that have roughly equal knowledge, however, can be more difficult to analyze because the desire of parties to contract will depend on the likely remedy.

Juliet Kostritsky has examined the role that knowledge of illegality plays in contracting and her framework provides a useful foil.²¹⁴ She advocates for an approach that will “minimize the costs from the nonenforcement of contracts” and she carefully works through different scenarios to

²¹¹ See Robert Weisberg, *Guideline Sentencing, Traditional Defenses, and the Evolution of Substantive Criminal Law Doctrine*, 7 FED. SENT'G REP. 168, 168-69 (1995) (noting the increasing migration of substantive elements of criminal law into the sentencing phase, but concluding that the general approach of separating the determination of liability from the determination of the penalty is not subject to significant criticism).

²¹² See *supra* Part III.B.

²¹³ Recall the example of bigamy—once a cad discloses his other marriage, the woman is unlikely to have any interest in the proposal. See *Ashley v. Dalton*, 81 So. 488, 489 (Miss. 1919).

²¹⁴ See Kostritsky, *supra* note 1, at 134-35.

see how this principle applies in a variety of illegal contract cases.²¹⁵ Kostritsky expects that parties to an illegal contract who are aware of the risk and cost of non-enforcement will allocate that loss to the cheapest cost-avoider and, in the absence of this type of express bargain, she argues that courts should attempt to identify the party that can avoid these costs most easily.²¹⁶ The largest difference between our approaches is that she seeks to minimize the social costs associated with deterring the formation of illegal contracts while my focus is on maximizing the social welfare associated with the activity that illegal contracts target.

In many cases, these different approaches will yield similar results. Kostritsky argues that courts should grant relief to the parties that are ignorant of the factual or legal reasons that render a contract unlawful.²¹⁷ She bases this argument on the fact that the search costs for those parties to discover the relevant facts would be high relative to the low cost of requiring disclosure from the knowledgeable party.²¹⁸ I agree with this outcome, although I do so on the basis that disclosure will tend to promote social welfare because disclosure of these material facts will prevent parties from entering into contracts that they do not view as beneficial. As I discuss in the example of the insurable interest,²¹⁹ if an insurer knows that a contract may be illegal and a potential insured does not, the insurer may price a policy at a lower level to account for the chance of non-enforcement. An uninformed potential customer may purchase the insurance because the price is attractive, but would not do so if the discount for potential illegality were removed. Purchases on an uninformed basis may, accordingly, be inefficient and that is why this type of conduct should not be rewarded.

But in some cases the framework I develop incorporates additional concerns from those of Kostritsky, concerns which can produce different outcomes in similar cases. She argues, for example, that when there is parity in knowledge of a contract's illegality, the non-enforcement rule should apply because this approach will minimize the costs of deterrence.²²⁰ In order for this argument to be correct, both parties must know that the contract is illegal and there must not be any potential positive externalities from the litigation. If the first requirement has not been met, the relationship between the incentive to invest and the remedy should be considered before a court applies non-enforcement as a remedy. When a restriction is ambiguous, the parties may be able to learn better whether a contemplated deal is illegal by investing in the acquisition of knowledge. This investment is a

²¹⁵ *Id.* at 123.

²¹⁶ *Id.* at 123-24.

²¹⁷ *Id.* at 135.

²¹⁸ *Id.* at 137-38.

²¹⁹ *See supra* Part III.B.

²²⁰ *See* Kostritsky, *supra* note 1, at 138 ("When the parties' knowledge of the underlying illegality appears equal, the courts should reject or restrict both parties' claims for judicial relief.").

social cost that will vary based on the remedy. Non-enforcement should produce a stronger incentive to invest because the threat of forfeiture imposes a larger loss on a party than a remedy like restitution. If, however, the legality of the contract remains ambiguous even after a party has made significant investment in determining its likely status, the social costs devoted to finding out this information may exceed any potential harm under the contract. In these situations, courts should be cognizant of both the harm and ambiguity dimensions of the problem because, even where parties are equally situated with respect to knowledge about the problem, the application of the non-enforcement rule can lead to over-investment in determining the legality of the contract.

CONCLUSION

Courts have, for several generations, justified the use of the non-enforcement remedy on the basis of its ability to deter. This Article argues that the deterrent ability of non-enforcement should be calibrated to the incentives it provides to contract. In instances where contracts pose the threat of harm, the culpability analysis invited by the *Restatement* can interfere with the quick application of the non-enforcement rule and, accordingly, dilute the ability of the rule to prevent that harm. Rote application of the non-enforcement rule, however, can overdeter beneficial contracting as well. Where parties have difficulties determining whether a contract is legal or not, the specter of non-enforcement can prevent potentially beneficial contracts from forming. There is a simple response to these concerns: courts should expressly incorporate inquiries into the harm that contracts of this sort are likely to produce and the amount of ex ante ambiguity faced by the parties. Requiring parties to show harm and the absence of ambiguity before applying non-enforcement would attenuate the current regime's overdeterrence and underdeterrence concerns while restoring the incentive to innovate that the threat of non-enforcement can threaten.