

The University of Chicago Law Review

Volume 83

Spring 2016

Number 2

© 2016 by The University of Chicago

ARTICLES

Contrived Threats versus Uncontrived Warnings: A General Solution to the Puzzles of Contractual Duress, Unconstitutional Conditions, and Blackmail

Einer Elhauge[†]

Contractual duress, unconstitutional conditions, and blackmail have long been puzzling. The puzzle is why these doctrines sometimes condemn threatening lawful action to induce agreements but sometimes do not. This Article provides a general solution to this puzzle. Such threats are (and should be) deemed unlawfully coercive only when they are contrived, meaning that the threatened action would not have occurred if no threat could have been made. I show that such contrived threats can be credible because making the threat strongly influences whether the threatened action occurs. When such threats are uncontrived warnings, meaning that the threatened action would have occurred even if no threat could have been made, they are not coercive and can only benefit the agreeing parties. However, sometimes (as with blackmail) agreements produced by uncontrived warnings are also unlawful on the ground that they harm third parties. The contrived-threat test explains why the Medicaid-defunding threat in the Affordable Care Act was

[†] Carroll and Milton Petrie Professor of Law, Harvard Law School. I am grateful for funding from the Petrie-Flom Center and Harvard Law School and for helpful comments from Michael Abramowicz, Jonathan Adler, Albert Alschuler, Scott Altman, Ian Ayres, William Baude, Adam Cox, Elizabeth Emens, Richard Fallon, Joe Farrell, Brian Fitzpatrick, Charles Fried, John Goldberg, Wendy Gordon, Rick Hills, Bert Huang, Daryl Levinson, John Manning, Eric Rasmusen, Chris Robertson, Louis Michael Seidman, Christopher Serkin, Steven Shavell, Suzanna Sherry, Sonja Starr, Matt Stephenson, Cass Sunstein, Mark Tushnet, Adrian Vermeule, Abe Wickelgren, and participants in the Harvard Law Faculty Workshop, the Harvard Law and Economics Workshop, the Vanderbilt Law Faculty Workshop, and the 2014 Yale conference on Medicare and Medicaid.

unconstitutional. It also explains why the recent King v Burwell conclusion—that the Affordable Care Act does not withhold tax credits from states that do not create insurance exchanges—would have been constitutionally required even if it had not been required by the statutory text.

INTRODUCTION	505
I. THREATS THAT INDUCE CONTRACT MODIFICATIONS.....	510
A. Why Contract Standards Suggest a Contrived-Threat Test	511
1. The special rule for modifications without mutual consideration.....	511
2. The general duress rule.....	512
B. How a Contrived-Threat Test Explains the Case Law.....	513
1. Cases of modifications without mutual consideration.	513
2. Cases of modifications with mutual consideration.....	518
C. A General Proof for the Contrived-Threat Test.....	522
1. Uncontrived warnings.....	523
2. Contrived threats.....	523
D. Other Legal Elements.....	525
1. No reasonable alternative.	525
2. Lack of objectively demonstrable reason or significant benefit.....	528
3. Unfair modification terms.....	529
E. Why Not Instead Focus on Whether the Threat Is Credible?.....	530
1. Why contrived threats can be credible.....	531
2. Why contrived threats should be deterred even if they are not credible.....	536
3. The credibility test does not fit contract law.	537
F. Comparing Other Proposed Tests	538
1. Professors Nozick and Schelling on threats versus warnings. ..	538
2. Changed circumstances.....	541
II. UNCONSTITUTIONAL CONDITIONS	544
A. Unconstitutional Conditions Affecting States	544
1. Findings showing a contrived threat.....	545
2. The lack of a reasonable alternative.....	549
3. The critiques.....	550
4. <i>King v Burwell</i>	553
B. Unconstitutional Conditions Affecting Private Persons	557
1. A general proof.....	557
2. The fit with case law.....	563
3. Comparing other proposed tests.....	572
III. THREATS THAT INDUCE NEW PRIVATE CONTRACTS.....	575
A. Threats to Exercise Unrelated Rights.....	575
B. Threats Not to Enter into the Challenged New Contract.....	576

2016]	<i>Contrived Threats versus Uncontrived Warnings</i>	505
1.	Threats to withhold desired goods unless the market price is paid.	576
2.	Threats to withhold desired goods when parties have been cut off from the market.....	578
3.	Bilateral negotiation to split extramarket surplus.	580
C.	Blackmail	581
1.	Contrived blackmail threats are always coercive and invalid... ..	582
2.	Uncontrived blackmail threats benefit contracting parties but harm third-party interests.....	582
3.	Prior literature.....	584
CONCLUSION		585

INTRODUCTION

The Supreme Court upheld the constitutionality of the Patient Protection and Affordable Care Act¹ (ACA) with one notable exception: it struck down the provision that threatened to remove preexisting federal Medicaid funding from states that refused to accept an expansion of Medicaid, on the ground that this threat “crossed the line distinguishing encouragement from coercion.”² This constitutional standard has been condemned as amorphous and meaningless.³ Because the standard explicitly incorporates contract-law concepts,⁴ we might hope to find clear guidance in the underlying contract law. But contract law makes the legality of threats that induce contractual agreements turn on whether the threats are “improper” or in “bad faith,” conclusory labels that themselves have been deemed incoherent and meaningless.⁵ Nor could the Court find useful guidance from the general unconstitutional-conditions doctrine because it has been deemed conclusory, incoherent, and “infamously inadequate.”⁶

¹ Pub L No 111-148, 124 Stat 119 (2010).

² *National Federation of Independent Business v Sebelius*, 132 S Ct 2566, 2603, 2606 (2012) (Roberts, joined by Breyer and Kagan) (plurality) (“NFIB”); id at 2661–62 (Scalia, Kennedy, Thomas, and Alito dissenting).

³ See id at 2641 (Ginsburg concurring in part and dissenting in part, joined by Sotomayor); Nicole Huberfeld, Elizabeth Weeks Leonard, and Kevin Outterson, *Plunging into Endless Difficulties: Medicaid and Coercion in National Federation of Independent Business v. Sebelius*, 93 BU L Rev 1, 88 (2013).

⁴ See *NFIB*, 132 S Ct at 2602 (Roberts, joined by Breyer and Kagan) (plurality); id at 2659–60 (Scalia, Kennedy, Thomas, and Alito dissenting).

⁵ See Oren Bar-Gill and Omri Ben-Shahar, *Credible Coercion*, 83 Tex L Rev 717, 752–53, 779 (2005) (finding that contract law has failed to “produc[e] a coherent jurisprudence of coercion”); Robert A. Hillman, *Policing Contract Modifications under the UCC: Good Faith and the Doctrine of Economic Duress*, 64 Iowa L Rev 849, 862, 875–78 (1979).

⁶ Daryl J. Levinson, *Framing Transactions in Constitutional Law*, 111 Yale L J 1311, 1350 (2002). See also Kathleen M. Sullivan, *Unconstitutional Conditions*, 102 Harv

Indeed, the Supreme Court seems to have given up even trying to define a coherent doctrine, openly admitting both in the Medicaid-defunding-threat case and a subsequent unconstitutional-conditions case that it could not articulate the line dividing constitutional conditions from unconstitutional ones, although it was nonetheless—somehow—confident that the conditions it was considering were over the line.⁷ This Article solves these doctrinal puzzles with a coherent theory that is normatively attractive and descriptively fits current legal results.

For analytical clarity, it is important to exclude some cases that lie outside these puzzles. If a threatened action by a government or private party were independently unconstitutional or illegal, there would be no difficulty condemning the threat. Likewise, sometimes even an uncoerced agreement to a condition or contract would be independently illegal or unconstitutional, usually because it would harm third parties or violate equal protection norms. In these cases, the independent doctrine of illegality or unconstitutionality does all the work.

Other times, whether the agreement was coerced is irrelevant because the threatener has a power to compel without any agreement, such as when the government has a power to order the relevant action because the individual or state has no constitutional right against such compulsion. Whenever direct coercion is permissible, that fact moots the issue of when threats of otherwise-lawful action should be deemed coercive, because they would be permissible either way. But in these cases the work is done by the independent legal doctrines that create a power to coerce.

The cases of interest involve situations in which a threat to engage in otherwise-lawful action (like a termination of funding) induces an agreement that is otherwise lawful, but the threat is nonetheless deemed too coercive to enforce the induced agreement. Those threats are the ones for which we could use some coherent principle to explain why they are sometimes condemned and sometimes allowed.

This Article provides a simple principle for resolving this puzzle. It begins, in Part I, by resolving the underlying confusion

L Rev 1413, 1428 (1989) (finding that “the Court’s unconstitutional conditions rulings display serious inconsistencies in their account of coercion”).

⁷ See *NFIB*, 132 S Ct at 2606 (Roberts, joined by Breyer and Kagan) (plurality); id at 2662 (Scalia, Kennedy, Thomas, and Alito dissenting); *Agency for International Development v Alliance for Open Society International, Inc*, 133 S Ct 2321, 2330 (2013).

in contract law about duress and contract modifications. I submit that this confusion can be resolved with the following simple principle: a threat to engage in otherwise-lawful action that induces contract modification is unlawfully coercive only when the threat is contrived, meaning that the threatened action would not have occurred if no threat could have been made.

The essence of the normative justification can easily be summarized. When a threat is not contrived, any legal prohibition that prevented the party from communicating the threat would, by definition, result in the threatened action. Moreover, because the threat, if made, would have induced the contract modification, both sides must, by definition, have thought that they were better off with the contract modification than with the threatened action. Deterring these threats would harm both parties. In such cases, communicating the planned action is less a threat than an uncontrived warning of what is coming, which gives the parties an opportunity to agree on an alternative that makes them both better off.

In contrast, when the threat is contrived, preventing the threat would not result in the threatened action, and the parties would stick to the initial contract. Preventing such a contrived threat would clearly leave the threat recipient better off because he prefers the preexisting contract. Ex post, preventing the contrived threat would leave the threatener worse off because she prefers the modification that the threat could have induced. But if the modification raises total value for both parties, a threat is unnecessary to achieve that modification because it could be achieved with a bonus that shares the joint gain. If, in contrast, the modification would reduce total value, then the prospect of such welfare-reducing threats would (if they were allowed) be priced into the contract and reduce reliance on the contract in a way that would harm both parties. Moreover, sometimes a contrived threat will not succeed in inducing contract modification and will then be carried out, which will leave both parties worse off than if the preexisting contract had continued. An ability to make such threats might also result in the other side making welfare-reducing threats of its own. Thus, ex ante, both parties would be better off preventing contrived threats.

The distinction between contrived threats and uncontrived warnings turns only on whether, in a no-threat world, the threatened action would have occurred. It does not turn on the wording of the communication. Nor does it require inquiry into

the threatener's purpose. The threatener may have the benign purpose of making the threat recipient better off, but the issue here presumes that the threat recipient begins with a legal autonomy right to make his own choices free of coercion by others, however well-intentioned that coercion may be. A threatening government may not even have a collective purpose, but such a purpose is unnecessary because what matters is what the government would have actually done without the threat. To be sure, the lack of any reason to take the threatened action in a no-threat world provides powerful evidence that the threatened action would not have occurred in that world. In that sense, purpose may sometimes be relevant to determining what would have happened, but the ultimate test is solely what would have happened without the threat, which need not require inquiry into purpose.

The contrived nature of a threat should also be distinguished from the threat's credibility. A threat is credible if, *post-threat*, the threatener would carry out the threatened action if the threat recipient did not agree. A threat is uncontrived if, in a *no-threat* world, the threatener would have taken that action anyway. The two differ because making a threat can strongly influence whether the threatened action occurs. As detailed below, a contrived threat to engage in unprofitable action can be credible when done in stages, as in the usual contracts case in which performance is withheld day by day, because the victim's refusal to modify at each stage is itself not credible, so that the threat is expected to result in profitable modification. Making a contrived threat can also change the emotional calculus or create reputational effects that make persons carry out the threat, even though they would not have taken the action in a no-threat world. Indeed, a credibility test has circularity problems when such reputational effects exist, because whether a threat is credible can depend on whether the induced agreement is enforceable. Persons or governments may also take steps to make their contrived threats more credible, such as incurring commitments, debts, or costs that make carrying out the threatened action more likely. When those steps would not have been taken absent an ability to communicate the threat, the credible threat remains contrived.

A contrived threat is thus not the same as a bluff, because contrived threats are often credible. Nor are contrived threats harmless when they are bluffs. Bluffs can be very effective in

coercing agreements when the victim is mistaken or even just unsure about the credibility of the threat. Indeed, if the bluff induced the victim to accept an adverse modification, it must have appeared credible enough to the victim. Nor is harm avoided when the victim rejects a modification. If a victim does so because he wrongly perceived that the threat was not credible, the threat will be carried out, harming both parties. If a victim does so because he correctly perceived that the threat was not credible, the threat will not be carried out but will still have created unnecessary conflict and negotiation costs. It is thus always better to deter a contrived threat from being made (by eliminating any possible gain from it), regardless of how credible the threat might have been or seemed.

My distinction between contrived threats and uncontrived warnings is related to, but different from, distinctions between threats and warnings that have been drawn by Professor Robert Nozick, Professor Thomas Schelling, and others.⁸ Whereas my definition turns on a pure but-for prediction, Nozick and other philosophers define threats in a way that combines moral and prediction baselines, which, as I show below, not only muddles the results but also perversely leads to more immoral conduct. Schelling and others use a prediction baseline—but one that is based on a pre-threat baseline rather than my no-threat baseline—and Schelling draws no normative conclusions about threats but rather focuses on analyzing their credibility.

Legally, an important benefit of using a contrived-threat test is not only that it is normatively desirable but also that, as Part I shows, the contrived-threat test descriptively explains many features and conclusions of contract law, whereas scholars who advocate a credibility test acknowledge that their test bears no relationship to existing contract law. Under contract law, a modification without mutual consideration is unenforceable if induced by a contrived threat, and a modification with mutual consideration is unenforceable if induced by a contrived threat to take action that would leave the victim significantly worse off.

I then show, in Part II, that the contrived-threat test also explains the Medicaid-defunding-threat case and provides a general, desirable solution to the problem of unconstitutional conditions. When a threat is contrived, the government benefit would have been provided in the but-for world without that

⁸ See Part I.F.1.

condition, and thus the threat to withhold the benefit penalizes the exercise of a constitutional right. When the threat is uncontrived, the government benefit would not have been provided in that but-for world, and thus withholding the benefit imposes no penalty. Some constitutional law scholarship going back to Professor Seth Kreimer has, following Nozick, included a similar prediction baseline but (like Nozick) combined it with other baselines in a way that I show both confuses and worsens the results.⁹ My contrived-threat test also shows that even if it had been textually plausible to interpret the ACA to deny federal tax credits to states that declined to create health insurance exchanges, such an interpretation should have been rejected because that would be a contrived threat and thus unconstitutional.

Finally, Part III shows that the contrived-threat test also explains and justifies current legal treatment of threats that induce private parties to enter into new contracts. It also solves the blackmail puzzle, showing that all blackmail contracts are necessarily produced by either (1) contrived threats, which create inefficiencies for the contracting parties without affecting information disclosure, or (2) uncontrived warnings, which suppress the disclosure of information to third parties whose interests often trump the interests of the contracting parties. The combination is what explains and justifies the full scope of blackmail law.

I. THREATS THAT INDUCE CONTRACT MODIFICATIONS

Whether a contract modification is enforceable turns on whether the threat that induced it is improper or in bad faith. Part I.A shows that, although those standards appear quite conclusory, the guidance about how to apply them is consistent with deeming threats improper or in bad faith only when they are contrived. Part I.B shows that a contrived-threat test also explains the pattern of case results in a normatively attractive way. Part I.C provides a more general proof about the normative desirability of a contrived-threat test. Part I.D shows that other legal elements for making contract modifications unenforceable can also be explained under a contrived-threat approach. Finally, Part I.E explains why a contrived-threat test is preferable to the credible-threat test that is proposed by some leading scholars,

⁹ See notes 223–29.

and Part I.F explains why a contrived-threat test is also preferable to various other tests suggested by scholars.

A. Why Contract Standards Suggest a Contrived-Threat Test

For threats of otherwise-lawful action that induce contract modifications, contract law has both a special rule and a more general rule. The special rule applies only to modifications without mutual consideration, which increase the obligations of one party without providing any fresh consideration for that change. The more general duress rule applies to any threat that induces contractual agreement. Although the general legal standards for both are conclusory, the official guidance on them suggests a contrived-threat test.

1. The special rule for modifications without mutual consideration.

A threat to breach a contract that induces a modification without mutual consideration is not binding unless the modification is (for a goods contract) in “good faith”¹⁰ or (for a nongoods contract) “fair and equitable in view of circumstances not anticipated by the parties when the contract was made.”¹¹ Although terms like “good faith” and “fair and equitable” are conclusory, the law does provide guidance on the meaning of these terms.

For nongoods contracts, this guidance provides: “The limitation to a modification which is ‘fair and equitable’ . . . requires an objectively demonstrable reason for seeking a modification.”¹² An objectively demonstrable reason provides an independent reason to breach if no modification were possible, thus suggesting an uncontrived warning. The lack of any such reason would instead suggest a likely contrived threat.

The unanticipated-circumstances element is often confused with unforeseen circumstances, but that is not the correct test; the Restatement (Second) of Contracts makes clear that the relevant unanticipated circumstances could be established by such foreseeable events as a person being offered a higher price or incurring a 17 percent cost increase.¹³ This is consistent with a contrived-threat test because such changes, though foreseeable,

¹⁰ UCC § 2-209, comment 2 (ALI 2014).

¹¹ Restatement (Second) of Contracts § 89(a) (1981).

¹² Restatement (Second) of Contracts § 89, comment b (1981).

¹³ Restatement (Second) of Contracts § 89, comment b, illustrations 3–4 (1981).

can often be the basis for an uncontrived warning if they make breach profitable despite contract remedies. The Restatement indicates that what it means by an unanticipated circumstance is one that “was not adequately covered” by the contract, “even though it was foreseen as a remote possibility.”¹⁴

“Unanticipated” thus means not specifically addressed by the contract. This factor is quite relevant to judging whether a threat is contrived, because if any changed circumstances were already specifically addressed in the contract, then it seems unlikely that the change would have provoked a breach in a world where no modification were possible. For example, suppose that costs have increased by 17 percent, but the contract specifically provides that if costs increase above 10 percent, the contract price will increase by the cost increase plus a specified markup. In that case, the parties have already determined what suffices to induce performance under those circumstances, so a claim that those circumstances would inevitably provoke breach seems contrived. Threatening to breach because of a circumstance specifically anticipated by the contract thus seems likely to be a contrived threat.

For goods contracts, the Uniform Commercial Code comments flesh out the “good faith” test by stating:

[T]he extortion of a “modification” without legitimate commercial reason is ineffective as a violation of the duty of good faith. . . . The test of “good faith” . . . may in some situations require an objectively demonstrable reason for seeking a modification. But such matters as a market shift which makes performance come to involve a loss may provide such a reason even though there is no such unforeseen difficulty as would make out a legal excuse from performance.¹⁵

The reference to an “objectively demonstrable reason” again suggests an uncontrived warning, and the last sentence confirms that such a reason can exist even because of foreseen changes in circumstances.

2. The general duress rule.

The general duress rule provides that any contract (including a new contract or modification with mutual consideration) is

¹⁴ Restatement (Second) of Contracts § 89, comment b (1981).

¹⁵ UCC § 2-209, comment 2 (ALI 2014).

voidable when induced by (1) “an improper threat” that (2) “leaves the victim no reasonable alternative” but to agree.¹⁶

At one time, the no-reasonable-alternative element referred to conclusory, and somewhat mystical, notions of whether the victim’s “free will” was overborne.¹⁷ However, contract law has long rejected such notions and concluded that this element simply requires evidence that carrying out the threat would have made the victim significantly worse off.¹⁸

The key normative work is thus done by the impropriety element, which standing alone is conclusory. But Restatement § 176 provides an extended definition of when a threat is improper. Some of the definition simply refers to threatened actions that are independently unlawful, which does not help with the current issue.¹⁹ But other parts of the definition say that a threat is improper if it is in bad faith, and the accompanying comments indicate that is so when the threat lacks an “objectively demonstrable reason,” which, as the above shows, suggests a contrived threat.²⁰ Even more helpfully, the Restatement provides, “A threat is improper if the resulting exchange is not on fair terms, and . . . the threatened act would harm the recipient and would not significantly benefit the party making the threat.”²¹ If a threatened act would not significantly benefit the threatener, the party likely would not have taken the action if unable to make the threat, thus indicating a contrived threat.

B. How a Contrived-Threat Test Explains the Case Law

1. Cases of modifications without mutual consideration.

Consider first the renowned case of *Alaska Packers’ Association v Domenico*.²² A group of fishermen agreed in San Francisco to take a boat to Alaska and fish in exchange for a company paying

¹⁶ Restatement (Second) of Contracts § 175 (1981).

¹⁷ Restatement (Second) of Contracts § 175, comment b (1981).

¹⁸ See Restatement (Second) of Contracts § 175, comment b (1981) (rejecting the test that “the threat must arouse such fear as precludes a party from exercising free will and judgment or that it must be such as would induce assent on the part of a brave man or a man of ordinary firmness . . . because of [the test’s] vagueness and impracticability,” and providing a series of examples making clear that no reasonable alternative exists whenever the alternative would make the victim significantly worse off).

¹⁹ Restatement (Second) of Contracts § 176(1)(a)–(c) (1981).

²⁰ Restatement (Second) of Contracts § 176(1)(d) & comment e (1981).

²¹ Restatement (Second) of Contracts § 176(2) (1981).

²² 117 F 99 (9th Cir 1902).

each fisherman \$50 for the season plus 2 cents for each salmon he caught.²³ Once the fishermen were in Alaska, they threatened to stop work entirely and return to San Francisco unless their fixed pay was increased to \$100.²⁴ Although the fishermen claimed that they threatened this because their nets were defective, the district court found otherwise, and the Ninth Circuit determined that the threat to stop work was “without any valid cause” and timed to take advantage of the fact that the company could not get replacement fishermen to such a remote location in time for the short fishing season.²⁵ The company, which had also invested \$150,000 in a salmon cannery in Alaska, agreed to increase the fixed pay to \$100.²⁶ The court held that, under these circumstances, this modification was unenforceable.²⁷

Given these findings,²⁸ the fishermen’s threat was clearly contrived. The fishermen had no valid reason for threatening to stop work, and doing so would not only deprive them of any wages for the season but also require them to fund their travel back to San Francisco and subject themselves to a risk of damages. There was thus no good reason to think that, if unable to make this threat, the fishermen would have stopped working. If contract law deters such threats from being made (by eliminating any gains from the modifications they induce), then the fishermen would have performed under the existing contract.

The lower court had upheld the modification because it concluded that it was mutually beneficial based on the company’s consent, which indicated that the company must have decided the modification was better than suing for breach of contract—

²³ Id at 100.

²⁴ Id at 100–01.

²⁵ Id at 101–02 (stressing that the fishermen “willfully and arbitrarily broke” their contractual obligation and noting the parallel to another case in which the threat to breach was “unjustifiable”).

²⁶ *Alaska Packers*, 117 F at 100–01.

²⁷ Id at 102–03. Although some read this case to adopt a flat rule that modifications without mutual consideration are never enforceable, the court stressed that its holding that mutual consideration was absent applied only “under such circumstances” and also depended on the conclusion that any mutual waiver of the original contract was not “voluntary,” because the circumstances were like another case in which the threatener took “an unjustifiable advantage of the necessities of the other party” by threatening nonperformance and thus “coerce[d]” the other party to increase the consideration it provided. Id at 102, quoting *King v Duluth, M. & N. Ry. Co.*, 63 NW 1105, 1106 (Minn 1895).

²⁸ Some have argued that the true facts were otherwise. See, for example, Bar-Gill and Ben-Shahar, 83 Tex L Rev at 757 (cited in note 5). But what matters for assessing the legal standard is the conclusion that the found facts made the modification unenforceable.

probably because the fishermen could not pay damages.²⁹ The problem with the lower court's logic is that it wrongly assumes that, absent any communicated threat, the fishermen would have walked off the job. Instead, the findings indicate that they would not have done so, and that their threat to do so was thus contrived to extort the modification. A contract doctrine that deters such contrived threats from ever being communicated will thus result in performance of the original contract. A contract doctrine that instead made modifications produced by contrived threats enforceable would encourage such contrived threats, which would clearly make the company worse off because it prefers original contract performance to the modification.

Moreover, a contract doctrine that made modifications produced by contrived threats enforceable would also be bad for fishermen *ex ante*. Switching to such a doctrine would encourage contrived threats and modifications that would make the fishermen's original promise to perform less valuable to the fish company, and thus would make the fish company unwilling to pay as much or to enter into the contract at all. To make itself less vulnerable to such contrived threats, the company might also be less willing to make investments in reliance on the contract (like \$150,000 for a cannery or spending to charter a boat), which is *ex ante* bad for the fishermen because reducing those investments reduces the productive value of their labor and thus reduces what they can get paid.

In contrast, suppose that the facts had been different. Suppose that the fishermen threatened to stop work unless their pay were increased to \$100 because, once they got to Alaska, another firm offered to pay them \$90. This would not be a contrived threat because, if unable to make this threat, the fishermen would (given their assumed inability to pay contract damages) walk off the job for better pay, which would clearly leave the fish company worse off than with the modification. It would instead be an uncontrived warning of what would inevitably happen if no modification were possible. Deterring the modification induced by this uncontrived warning would be harmful to both sides—without it, the fishermen would get \$90 instead of \$100, and the fish company would be without fishermen.

But doesn't this uncontrived warning still diminish the willingness of the fish company to pay as much in the original contract

²⁹ *Alaska Packers'*, 117 F at 102.

or to make investments that rely on it? Actually, no. The uncontrived warning allows a modification that makes the fish company better off than it would be if the fishermen walked off the job, which is what would happen if no warning were given. Thus, the possibility that uncontrived warnings would produce such modifications actually *increases* the value of the contract and encourages greater investment in relying on it. To be sure, the prospect that the fishermen might independently want to walk off the job (in a way that a contract damages action would not fully compensate) *does* diminish the value of the contract and reliance on it. But if the warning is uncontrived, that prospect will arise regardless; the ability to modify the contract when it arises thus can only reduce the losses from that prospect.

Consistent with this conclusion, the Restatement indicates that a modification increasing an employee's pay for already-contracted performance would be binding if it came after an employee told his employer that he was abandoning the contract because another firm had offered him more money.³⁰ The *Alaska Packers'* court's own reasoning likewise suggests that its result would have been different if the threat to breach the contract were made for a "valid cause," which switching to a higher-paying firm would presumably supply.³¹ Indeed, *Alaska Packers'* indicated no doubts about the enforceability of a prior modification without mutual consideration that increased the fixed pay of many of the fishermen from \$50 to \$60 *before* they left San Francisco.³² The fishermen would not have been able to induce that change if the fish company could secure substitute fishermen for less than \$60, which suggests that the market rate had risen to \$60 for these fishermen and that they would thus have declined to get on the boat for Alaska if that modification were not possible.

Likewise, if a party's contract-performance costs have increased above expected costs by an amount sufficient to make performance unprofitable, then modifications that pay it more

³⁰ Restatement (Second) of Contracts § 89, illustration 3 (1981). See also *Schwartzreich v Bauman-Basch, Inc.*, 131 NE 887, 890 (NY 1921). Although a threat to leave for higher pay could be contrived if the pay increase were lower than the expected contract damages, the employee in *Schwartzreich* had simply announced he was leaving for higher pay without demanding any modification, thus indicating that he would have left if no modification were possible. *Id.* at 888.

³¹ *Alaska Packers'*, 117 F at 102.

³² *Id.* at 100.

for the same performance have generally been deemed binding.³³ In such cases, threats not to perform are usually uncontrived warnings because the party would likely walk away from the unprofitable contract if unable to communicate any threat to do so. This is not always so, because paying contract damages might be even more unprofitable than performance. But the cases finding the modifications enforceable generally indicate that the threatener was unwilling to perform at the contract price (perhaps because the threatener rationally found the performance cost worse than the risk of paying contract damages),³⁴ which would make the threat uncontrived.

If contract remedies left victims of breach just as well off as they would be with contract performance (as expectation damages aspire to do), then threats to breach contracts could never get threatened parties to agree to modifications that left them worse off than with contract performance; they would instead reject the modifications and take the contract remedies.³⁵ Undercompensatory contract remedies are thus necessary for a threat to breach to lead to harmful modifications. But modifications without fresh consideration by definition leave threatened parties worse off than they would be with contract performance. Accordingly, such nonmutual modifications indicate that contract remedies must undercompensate threatened parties in some way. This could be true for various reasons, including that expectation damages cannot be collected or that actual contract remedies often provide less than full expectation damages. But direct proof that contract remedies are undercompensatory is unnecessary because one can infer undercompensatory remedies from the nonmutual modification itself. So it is not surprising that courts require no proof of undercompensatory remedies in cases in which threats to breach induced nonmutual modifications. However, while undercompensatory contract remedies are necessary for a threat to breach to lead to harmful modifications, they are not sufficient: the threat to breach must also be contrived. Indeed, undercompensatory contract remedies are often what lead to uncontrived warnings that a breach is coming. When

³³ See Restatement (Second) of Contracts § 89, illustrations 1–2, 4 (1981).

³⁴ See, for example, Restatement (Second) of Contracts § 89, illustration 1 (1981); *Brian Construction and Development Co v Brighenti*, 405 A2d 72, 74 (Conn 1978); *Siebring Manufacturing Co v Carlson Hybrid Corn Co*, 70 NW2d 149, 151 (Iowa 1955); *Goebel v Linn*, 11 NW 284, 284–85 (Mich 1882).

³⁵ See Part I.D.1.

that is the case, communicating the uncontrived warning can lead to modifications that make both parties better off than the breach that would occur if no warning could be communicated.

2. Cases of modifications with mutual consideration.

Consider now the famous case of *Wolf v Marlton Corp.*³⁶ After contracting to buy a house in a housing development, the buyers decided to divorce and wanted to escape the contract.³⁷ They threatened that, unless the seller agreed to cancel the contract and return their security deposit, they would go ahead with the purchase and deliberately resell the house to an undesirable purchaser, which would reduce property values in the neighboring tracts owned by the seller.³⁸ The court held that this threat of economic harm overcame the free will of the seller but could constitute duress only if the threat were also “wrongful.”³⁹ The court acknowledged that the threatened action was perfectly lawful because the buyers had the legal right to resell to whomever they wished, regardless of the economic effect on the seller.⁴⁰ However, the court concluded that it was nonetheless wrongful when “a party for purely malicious and unconscionable motives threatens to resell such a home to a purchaser, specially selected because he would be undesirable, for the sole purpose of injuring the builder’s business.”⁴¹ The Restatement likewise makes clear that a threat to resell property to a purchaser that will have adverse effects on surrounding land is improper only if the threat is made “solely” to induce the contract modification.⁴²

The court’s conclusion fits well with the theory that threats are wrongful when they are contrived. If the threat to resell to a purchaser specially selected to be undesirable was “purely malicious” and with the “sole purpose” of harming the seller’s business and inducing a modification,⁴³ then there was no good reason to think that (if unable to communicate the threat) the buyers would have limited their resale options to undesirable purchasers. Limiting the set of potential purchasers would

³⁶ 154 A2d 625 (NJ Super App Div 1959).

³⁷ Id at 626–27.

³⁸ Id.

³⁹ Id at 629, quoting *Rubenstein v Rubenstein*, 120 A2d 11, 15 (NJ 1956).

⁴⁰ *Wolf*, 154 A2d at 630.

⁴¹ Id.

⁴² Restatement (Second) of Contracts § 176, comment e, illustration 10 (1981).

⁴³ *Wolf*, 154 A2d at 630.

naturally tend to lower the expected resale price, which would harm the threatening buyers. Deterring the threat would thus result in the original buyers making no special effort to seek out undesirable purchasers. The threat was therefore contrived and, accordingly, wrongful.

The *Wolf* court's reasoning indicates that its holding would have been different if the threat had not made for the sole purpose of harming the seller but was instead warning about coming action that would be independently beneficial for the threatening buyers. Suppose, for example, that the buyers had told the seller they planned to resell to a registered child molester because he was willing to pay 10 percent over market price. In that case, the communication would be an uncontrived warning because, if unable to make this communication in order to negotiate a modification, the buyers would in fact resell to the child molester, who was willing to pay the most. The seller would prefer to receive this warning because it enables the seller to negotiate a modification that makes the seller better off than he would be without the warning.

The *Wolf* decision shows that threats can be contrived, and thus improper, even if there is some probability that the threatened action would have occurred in a no-threat world. After all, even if the home buyers made no special effort to resell the house to an undesirable purchaser, there are presumably some odds (call it X percent) that this would happen just from reselling to the highest bidder. Nonetheless, the threat to resell *only* to an undesirable purchaser means the home buyers are threatening to increase those odds to 100 percent, which is still a contrived threat because the threatened action would not have occurred $100 - X$ percent of the time in a no-threat world. Accordingly, a threat stated with certainty (as most threats are) is contrived when the threatened action is actually a probabilistic event.⁴⁴ More generally, a threat that overstates the likelihood of the threatened action is contrived.

In contrast, accurately stating the odds that the threatened action would occur in the no-threat world would be an uncontrived warning. Suppose, for example, that the home buyers were trying to resell their house and the bidding came down to two bidders, one of whom was a registered child molester, and

⁴⁴ Most threatened actions involve acts over which the threatener has control, so most threatened actions are not probabilistic in this sense.

that if no threat were communicated, the odds would be fifty-fifty which bidder would win. If the home buyers were to accurately communicate those odds to the seller, then that would be an uncontrived warning because the communication would not overstate the odds that the threatened action would have occurred anyway. If the seller preferred the modification to those fifty-fifty odds, then the seller would be better off knowing about those odds and agreeing to the modification. If the communication were deterred, then the seller would be worse off because he would suffer fifty-fifty odds that he would deem worse than the modification.

The *Wolf* case also shows that threats that induce contract modifications can be contrived, and thus improper, even though no threat to breach the contract was made. In *Wolf*, the threat was actually to perform the contract, but then to act in a particular manner afterward that was contrived.

Other cases involve threats to breach a contract that lead to modifications with mutual consideration. A leading case is *Austin Instrument, Inc v Loral Corp.*⁴⁵ Austin had a contract to supply gear parts to Loral to make radar sets for the Navy.⁴⁶ When Loral got a second Navy contract for radar sets, Loral told Austin it would get a second contract only on those gear parts for which it was the lowest bidder.⁴⁷ Austin immediately said that it was opposed to supplying less than all gear parts on the second contract, and the next day Austin threatened to cease deliveries under the first contract unless Loral not only ordered all gear parts from Austin on the second contract but also gave Austin substantial price increases on gear parts under the first contract, whether or not those gear parts were already delivered.⁴⁸ Although the court focused on whether Loral had any reasonable alternative, because that was the issue disputed on appeal, the court made clear that duress also required a showing that the threat was “wrongful.”⁴⁹

Here the threat was properly deemed wrongful because the threat to cease delivery was contrived.⁵⁰ True, the lower court

⁴⁵ 272 NE2d 533 (NY 1971).

⁴⁶ *Id.* at 534.

⁴⁷ *Id.*

⁴⁸ *Id.*

⁴⁹ *Austin*, 272 NE2d at 535.

⁵⁰ Dicta in the opinion suggest that any threat to breach a contract is wrongful. *Id.* The actual holding, however, is limited to the court’s factual findings, and the Restatement makes it crystal clear that “[a] threat by a party to a contract not to perform his

noted that Austin claimed it was motivated by cost increases that made supplying gear parts unprofitable,⁵¹ but the high court's statement of facts omitted this claim in a way that implicitly rejected it. Further, such a cost-based motive seemed implausible given the facts that (1) the timing indicated that the threat was motivated by a desire to supply all gear parts on the second contract, rather than by cost increases; (2) cost increases that make supplying gear parts unprofitable are not a likely motive for insisting on supplying more gear parts; and (3) a need to cover costs on future supplies cannot explain the demand for price increases on already-delivered gear parts. Thus, the facts suggested that Austin would not have naturally ceased delivery under the first contract, and that it threatened to do so only because making that threat would induce the desired contract modification.

When cost increases do make contract performance highly unprofitable, the Restatement concludes that it is permissible to threaten to breach unless the price is increased.⁵² This conclusion is consistent with a contrived-threat test because such facts suggest that the party would not perform even if there were no prospect of inducing a modification, which would make the threat an uncontrived warning.

When a threat to breach induces a nonmutual modification, we can (as noted above) infer that contract remedies must be undercompensatory, because otherwise the threatened party would not have agreed to the modification. In contrast, we cannot make the same inference for modifications with mutual consideration, because it is possible that the return consideration made the threatened party better off at the time of the modification. When threats to breach induce mutual modifications, courts like the one in *Austin* often do require evidence that the contract remedy is significantly undercompensatory.⁵³

contractual duty is not, of itself, improper." Restatement (Second) of Contracts § 176, comment e (1981).

⁵¹ *Austin Instrument, Inc v Loral Corp*, 316 NYS2d 528, 530 (NY App 1970). The lower court did not confirm whether Austin's costs had actually risen enough to make performance unprofitable, but rather it simply noted that cost increases and renegotiations were common at the time. *Id.*

⁵² Restatement (Second) of Contracts § 176, illustration 8 (1981).

⁵³ See *Austin*, 272 NE2d at 535.

C. A General Proof for the Contrived-Threat Test

Assume a threat is made to induce a contract modification. Call the value of the initial contract Ct for the threatener and Cr for the recipient of the threat. Call the value of the modified contract Mt for the threatener and Mr for the recipient. And call the value of the situation if the threatened action were taken Tt for the threatener and Tr for recipient.

We know that $Mr > Tr$, because otherwise the threat would not induce modification. For modifications without mutual consideration, by definition $Cr > Mr$ because the modification worsens the deal for the recipient with zero offsetting benefits. Therefore, $Cr > Mr > Tr$. For modifications with mutual consideration, it may not be clear that $Cr > Mr$ at the time of modification.⁵⁴ But the no-reasonable-alternative element of duress means that the threatened action would leave the victim significantly worse off than under the contract and thus means that $Cr > Tr$.⁵⁵ Such a threat would be necessary to induce modification only if the recipient preferred the contract to the modification, because if $Mr > Cr$, the recipient would accept the offered modification without any threat. So we can infer from such a threat that $Cr > Mr$. Alternatively, direct proof might exist that $Cr > Mr$. Either way, contract-doctrine elements that are separate from the “improperness” or “bad faith” of the threat establish that $Cr > Mr > Tr$.

The threatener would not have sought the modification unless the threatener preferred the modification both to the contract (so $Mt > Ct$) and to the threatened action (so $Mt > Tt$). However, further ranking has two possibilities that differ depending on whether the threatener prefers the contract to taking the threatened action. One possibility is that $Mt > Tt > Ct$. In this case, the threat is uncontrived because taking the threatened action would make the threatener better off than she would be under the contract. The other possibility is that $Mt > Ct > Tt$. In this case, the threat is contrived because taking the threatened action would make the threatener worse off than she would be under the contract.

⁵⁴ When a threat recipient opposes modification enforcement, we know that $Cr > Mr$ at the time of litigation, but that valuation could reflect changes in costs or market prices since the time of modification.

⁵⁵ See Part I.D.1.

1. Uncontrived warnings.

If $Mt > Tt > Ct$ and the law prevents the threat and thus the modification, then the threatened action will occur because it makes the threatener better off than sticking to the contract. Preventing the threat and modification would thus make both parties worse off because they would both prefer the modification to the situation with the threatened action (that is, $Mr > Tr$ and $Mt > Tt$). The threat is thus an uncontrived warning of what is inevitably coming unless the parties agree to a mutually beneficial alternative.

True, such an uncontrived threat will induce a modification that is less valuable to the recipient than the contract. Thus, relative to contract performance, the prospect of modifications might seem to reduce the contract's initial value to the recipient. But contract performance is the wrong baseline because, without the uncontrived threat and modification, the victim would instead suffer the threatened action, the prospect of which would reduce the contract's initial value to the recipient *even more*. Accordingly, relative to the correct but-for baseline of the threatened action occurring, the prospect of allowing modifications when uncontrived warnings are made will raise the contract's initial value to the threat recipient, making the recipient willing to give contract terms that are more favorable to the threatener and to make more relationship-specific investments, both of which will also raise the contract's initial value to the threatener. Both parties are thus better off if such uncontrived warnings are allowed to induce contract modifications.

2. Contrived threats.

If $Mt > Ct > Tt$ and the law prevents the threat and the modification, then the threatened action would not occur, because taking the threatened action would leave the threatener worse off than under the contract. The original contract will thus be performed. Preventing such a contrived threat would thus clearly leave the recipient better off because $Cr > Mr$. That may suffice if the normative goal of duress/modification doctrine is simply to protect the recipient.

Further, *ex ante*, preventing contrived threats will also be preferable to the threatener, and thus both parties would benefit from a contract doctrine that prevented contrived threats. True, *ex post*, preventing such a contrived threat would leave the threatener worse off because $Mt > Ct$. But *ex ante*, the prospect

of contrived threats will be priced into the contract and affect the extent of reliance on that contract. To evaluate those effects, consider the following three possible consequences of a contrived threat.

a) Contrived threats that induce modifications that reduce joint welfare. One variation is that a contrived threat might induce a modification that lowers the total combined value for the parties, which means that $Cr + Ct > Mr + Mt$. In that case, not only would the modification lower joint welfare, but also the prospect of such modifications would lower the contract's initial value to the recipient, leading the recipient to demand better terms and make fewer relationship-specific investments, both of which lower the contract's initial value to both the threatener and the threat recipient. Further, the potential threatener would know that the recipient could make his own threats of this nature, which would leave both the threatener and the recipient worse off. In short, such welfare-reducing threats will be priced into the contract, deter reliance on the contract in a way that harms both parties, and be used against both sides, so that ex ante both parties would be better off preventing such threats.

b) Contrived threats that induce modifications that increase joint welfare. The other variation is that the modification increases total joint value because the modification's benefit to the threatener exceeds its harm to the recipient. In that case, $Mt - Ct > Cr - Mr$. But then the threatener does not need to make a threat to induce the modification. The threatener could instead obtain the same modification by offering a bonus B that exceeds the recipient's harm ($Cr - Mr$) but is lower than the threatener's benefit from the modification ($Mt - Ct$), so that $Mt - Ct > B > Cr - Mr$. In short, if the modification increases joint welfare, it can be achieved without making any threat (of $Tr < Mr < Cr$) by instead offering a bonus that shares the joint gains between the two parties so that $Mr + B > Cr$. Thus, preventing contrived threats is never necessary for threats that enhance joint welfare.

c) Contrived threats that induce no modification. The final possibility is that a contrived threat fails to induce a modification, perhaps because the recipient did not perceive it to be credible. If the recipient is mistaken in his perception, then the threat will be carried out, which makes both sides worse off because $Cr > Tr$ and $Ct > Tt$. If the recipient is correct in his

perception, then the contract will continue, but the parties will have incurred unnecessary costs of conflict and negotiation.

d) Overall effect of banning all contrived threats. In short, preventing all contrived threats will prevent some modifications that reduce joint welfare, will never prevent modifications that increase joint welfare, and will prevent some contrived threats from being carried out to the detriment of both parties. Thus, banning all contrived threats leaves both parties better off *ex ante*, without any need to examine the welfare effects of the modification.

One might wonder whether we should also prevent uncontrived warnings because they, too, can never be necessary for modifications that enhance joint welfare relative to contract performance. The reason we should not is that, although uncontrived warnings can indeed produce modifications that lower joint welfare relative to contract performance, without the uncontrived warnings we would not get contract performance. Instead, we would get the threatened action, and the induced modification does enhance welfare for both parties relative to the situation with that threatened action.

In short, to assess the effects of a threat/warning, we must compare those effects to a but-for baseline that reflects the effects that would exist in the world *but for* the making of that threat/warning. For a contrived threat, the effects are negative because the but-for baseline is contract performance, which gives higher joint value than any modification that requires such a threat. For an uncontrived warning, the effects are positive because the but-for baseline is a threatened act that gives both parties lower value than the modification does.

D. Other Legal Elements

1. No reasonable alternative.

The above analysis establishes that whether a threat is legally improper should turn on whether it is contrived. For any threat to induce modification, it must be the case that $Mr > Tr$. For modifications without mutual consideration, by definition $Cr > Mr$, and thus we know $Cr > Mr > Tr$. For modifications with mutual consideration, duress doctrine also requires proving that the threat recipient had “no reasonable alternative,” which means that the threatened action would leave the recipient significantly worse off than under the contract and thus means

that $Cr > Tr$.⁵⁶ Because such a threat could be necessary only if $Cr > Mr$, this element establishes that $Cr > Mr > Tr$. As I note in Part I.C, this value ranking is a requisite factual premise to show that the situation involved a true threat rather than a bonus. Alternatively, one might have direct evidence that $Cr > Mr$, in which case we can infer that $Cr > Tr$ because $Mr > Tr$.⁵⁷ Direct evidence that Cr is significantly greater than Mr should thus suffice to satisfy the no-reasonable-alternative test.

Administrative grounds explain why the law requires a significant difference even though a small difference could result in some small harm. In a world with costless, perfectly accurate adjudication, all contrived threats would be condemned even if the harm were small. But we do not live in such a world. Sometimes a court will mistakenly treat an uncontrived warning like a contrived threat. A significance threshold assures that the potential harm is large enough to exceed both litigation costs and the overdeterrence effects that can result from inevitable errors in adjudicative fact-finding. The significance threshold thus reduces the overdeterrence of uncontrived warnings that might be mistaken for contrived threats, but it does so at a cost of underdetering contrived threats with small harms.

For modifications without mutual consideration, contract law does *not* require proving the no-reasonable-alternative test. This makes sense because, as the proof above shows, any modification without mutual consideration necessarily satisfies the condition that $Cr > Mr > Tr$. Because this automatically follows, litigation costs or errors in establishing this condition are likely to be far lower than for modifications with mutual consideration.

If contract-law remedies fully satisfied the expectation damages goal of putting the recipient in the same position as contract performance would put him, a threat to breach a contract would mean $Cr = Tr$, and thus the necessary premise that $Cr > Tr$ could not be satisfied and a recipient would never agree to a modification for which $Cr > Mr$. Accordingly, threats to breach a contract can cause the relevant problem only when contract remedies are undercompensatory—which alas is often the

⁵⁶ Restatement (Second) of Contracts § 175(1) (1981).

⁵⁷ Some argue that the no-reasonable-alternative test is unnecessary because the recipient would never agree to the modification unless he had no reasonable alternative. See, for example, Oren Bar-Gill and Omri Ben-Shahar, *The Law of Duress and the Economics of Credible Threats*, 33 J Legal Stud 391, 423 (2004). But that assumes the modification was adverse to the recipient when it was made, which is not necessarily true for modifications with mutual consideration. See note 54.

case given litigation costs and delays, uncertain adjudication, limited defendant assets, and doctrines that sometimes provide damages below expectation levels. For modifications without mutual consideration, we know that $Cr > Mr$, so we can infer that legal remedies must not have been fully compensatory, or else the recipient would not have agreed to the modification. For modifications with mutual consideration, courts instead often require evidence that the contract remedy is significantly undercompensatory.⁵⁸

One might think that the underlying problem of inadequate contract remedies indicates that the solution should be to increase contract remedies rather than to have doctrines that police modifications. However, that solution would not help when, as in the *Wolf* case, the contrived threat is not a threat to breach the contract. Moreover, even when the contrived threat is a threat to breach, that solution may be undesirable or unattainable. Optimal contract remedies can only minimize the sum of underdeterrence and overdeterrence of breaches, thus making undercompensatory remedies inevitable in some cases. Nor will increasing contract remedies help in cases in which the threatener lacks sufficient assets to pay the damages. In any event, to the extent that actual contract remedies leave $Cr > Tr$, it remains desirable to prevent modifications caused by contrived threats to breach contracts. The fact that different contract remedies might have deterred the threats to breach does not alter the desirability of voiding modifications induced by contrived threats that are made, given actual contract remedies.

The fact that the problem of contrived threats to breach contracts arises only when there are inadequate contract remedies also explains why, if preventing contrived threats is good for both parties, the parties' initial contracts do not simply prohibit contrived threats. The answer is that the same inadequate contract remedies would also fail to deter a breach of any promise not to make contrived threats. The only solution to that problem is to make modifications produced by contrived threats unenforceable. This is a solution that only the courts can supply because contract law provides that contracting parties can always agree to modify any contractual prohibition on modifications, thus making contractual prohibitions on future modifications unenforceable.⁵⁹

⁵⁸ See, for example, *Austin*, 272 NE2d at 535.

⁵⁹ See Restatement (Second) of Contracts § 311, comment a (1981).

2. Lack of objectively demonstrable reason or significant benefit.

The legal standards indicate that, in general, a threat is improper and in bad faith unless there is an “objectively demonstrable reason for seeking a modification”⁶⁰ or a “significant[] benefit”⁶¹ from the threatened action.⁶² While having a demonstrable reason or motive for the threatened action suggests that the threat was uncontrived, one might wonder about the requirement that the reason must be “objective” or must produce some concrete benefit. Why shouldn’t a purely subjective reason, like anger or spite, count if that motive would have dictated the threatened action even if no threat had been made?

Suppose the *Alaska Packers’* fishermen and the *Wolf* divorcing home buyers were so angry about their situations that (even if no modifications were possible) they would have taken the threatened actions despite the fact that doing so would clearly harm themselves financially. If so, refusing to enforce such modifications because those reasons were subjective would actually harm the recipients if we were sure both that (a) these were the facts and (b) the anger itself was not contrived.

What explains the rule requiring an objective reason or concrete benefit is that both those premises are untrue. First, claims about anger are subjective and easy to fake. Nor would it suffice to ascertain whether the anger was real; one would need to quantify its subjective effect to determine whether it would really outweigh the financial harm of taking the threatened action. This creates a proof problem not only in hard cases, but also as a systemic matter, because the proof problem could be artificially generated in every case.

Second, even if anger would genuinely motivate the threatened action, anger might itself be endogenous to the desire to create a commitment or reputation that can make contrived threats more credible. Persons might work themselves into a lather because their anger makes it credible that they will carry out their threats, but their anger itself might be contrived. Or they might want a reputation for acting out of anger when they are crossed, so that their contrived threats will be more effective. Either way, anger itself can be contrived, and refusing to enforce modifications

⁶⁰ UCC § 2-209, comment 2 (ALI 2014).

⁶¹ Restatement (Second) of Contracts § 176(2)(a) (1981).

⁶² See Part I.A.

induced by threats that were produced by contrived anger might thus deter those threats from ever being made.

In short, in a world with costless, perfectly accurate adjudication, uncontrived warnings should be permissible even when based on nonpecuniary motives. However, in the actual world, it is hard to detect anger that is faked or contrived, so admitting subjective motives would lead to significant underdeterrence of contrived threats. To reduce this underdeterrence, contract law thus has sound reasons to deem a threat improper (that is, contrived) unless the threatened action is supported by some objective reason or concrete benefit other than inducing the modification, although the necessary cost is to overdeter some genuine, emotional uncontrived warnings.

3. Unfair modification terms.

Another doctrinal issue is why contract law makes the fairness of the induced modification relevant to whether it should be enforceable.⁶³ Given the proof above, one might think that any contrived threat necessarily induces harmful modification and that any uncontrived warning induces beneficial modification, so that any separate inquiry into the fairness of the resulting modification seems unnecessary.

However, there is a sound reason for this requirement. Even though the refusal to perform a contract may be an uncontrived warning, in that the threatener would prefer nonperformance if *no* modification were possible, a threat not to perform unless a *particular* modification is granted may be contrived because a lesser modification would have sufficed to induce performance.⁶⁴ For example, a cost increase might make a party unwilling to perform at the original contract price, but a threat not to perform unless given a price increase that vastly exceeds that cost increase would still be a contrived threat because a smaller price increase would suffice. Thus, to limit the threatener to the zone of uncontrived warnings, the law must also police the fairness of the demanded modification. This explanation provides concrete content to what “fairness” means here—it means the modification

⁶³ See Restatement (Second) of Contracts §§ 89, 176 (1981).

⁶⁴ Consistent with this conclusion, the Restatement provides: “Bad faith may be shown by proving that the person making the threat did not believe there was a reasonable basis for the threatened process . . . or that he realized the demand he made was exorbitant.” Restatement (Second) of Contracts § 176, comment d (1981) (emphasis added).

should not exceed the minimum that is necessary for the threatener to be willing to perform the original contract.

For example, a Restatement illustration indicates that despite a demonstrable increase in the market price for metal needed to make castings, it would be improper to threaten to withhold contractual supply of those castings unless the contract price were increased from 50 cents to 75 cents, in a case in which rivals sold the same castings for 55 cents but with several months' delay.⁶⁵ Given the demonstrable cost increase, the threat not to perform at the original contract price of 50 cents could well be uncontrived, and other Restatement provisions seem to treat such threats as proper.⁶⁶ But the problem here is that the threat was not to perform unless the price were increased to 75 cents, which does seem contrived because rivals were willing to supply the castings at 55 cents despite the same increase in market costs for metal. Thus, although a threat not to perform unless the contract price were increased to 55 cents could be an uncontrived warning, a threat not to perform unless the contract price were increased to 75 cents would be a contrived threat.

E. Why Not Instead Focus on Whether the Threat Is Credible?

Professors Oren Bar-Gill and Omri Ben-Shahar argue that contract modifications should always be enforced whenever they are induced by credible threats or by noncredible threats that are accurately perceived, limiting potential nonenforceability to noncredible threats that are inaccurately perceived.⁶⁷ Other scholars reach a similar conclusion.⁶⁸ Although their work is highly illuminating, such a credible-threat test differs from a contrived-threat test in important ways that make the latter more desirable.

Whether a threat is contrived turns on a *no-threat* baseline: If no threat could have been communicated, what would the threatener have done? If she would have taken the threatened action, then the threat is uncontrived; if she would not have, then the threat is contrived. Whether a threat is credible instead

⁶⁵ Restatement (Second) of Contracts § 89, comment b, illustration 5 (1981).

⁶⁶ Restatement (Second) of Contracts § 89, comment b, illustrations 1–2, 4 (1981).

⁶⁷ Bar-Gill and Ben-Shahar, 33 J Legal Stud at 392–94, 417 (cited in note 57); Bar-Gill and Ben-Shahar, 83 Tex L Rev at 717, 720, 727–28 (cited in note 5).

⁶⁸ See Daniel A. Graham and Ellen R. Peirce, *Contract Modification: An Economic Analysis of the Hold-Up Game*, 52 L & Contemp Probs 9, 11–12, 23 (1989).

turns on a *post-threat* baseline: After the threat is communicated, would the threatener actually carry it out if the modification were rejected?⁶⁹ If she would, then her threat is credible; if she would not, then her threat is not credible.

This is a critical difference because the contrived-threat test is designed to prevent such threats from being made in the first place. A post-threat baseline instead assumes that the threat would be made either way.⁷⁰ The tests reach different results in cases when, if no threat could be made, the threatener would not take the threatened action, but if a threat were made, the threatener would carry it out. In such cases, the threat is contrived but credible, and a credibility test would thus wrongly enforce any modification that it induced. In contrast, my contrived-threat test would not enforce the induced modification, because denying threateners any benefits from making such threats would desirably deter such threats from ever being made.

1. Why contrived threats can be credible.

A contrived threat can be credible whenever the act of making the threat makes the threatener more likely to take the threatened action if no modification is given. Making a threat can have this effect for many reasons.

a) A contrived threat can be credible when staged. A contrived threat not to perform a contract is generally carried out in stages that make it credible because the victim's refusal to modify at each stage is itself not credible.⁷¹ Suppose, for example, that withholding performance is unprofitable standing alone, so that a onetime threat to never perform might not seem credible.

⁶⁹ See Bar-Gill and Ben-Shahar, 83 Tex L Rev at 722 (cited in note 5) ("Credibility is evaluated with an eye to the hypothetical temporal moment when the threat fails to induce the threatened party to surrender."); id at 742 ("Our analysis suggests that the correct baseline . . . is not the position of the threatened party prior to the threat, but rather the position that she would be in if she were to reject the threat."); id at 742 n 51 (stressing that Bar-Gill and Ben-Shahar are using a "post-threat baseline").

⁷⁰ See id at 742 (observing that Bar-Gill and Ben-Shahar's post-threat "position takes the existence of a threat to be part of the unfortunate but relevant reality in which the dilemma has to be resolved").

⁷¹ Professor Lucian Arye Bebchuk established a similar point for threats to sue. See Lucian Arye Bebchuk, *A New Theory concerning the Credibility and Success of Threats to Sue*, 25 J Legal Stud 1, 4–25 (1996). Threats of contract nonperformance are even more credible than litigation threats because, while in Bebchuk's model both sides incur equal costs at each litigation stage, in the contract situation usually the victim immediately incurs large costs from nonperformance while the threatener incurs costs only later if the victim sues or does not pay.

Nonetheless, a threatener usually withholds contract performance only one day at a time. The threat to continue doing so each day is credible when a victim has no reasonable alternative, because each day the victim's refusal to modify is not credible, given that the victim is harmed each day he refuses to modify. Moreover, the threatener usually suffers harm only later if the victim sues or does not pay—possibilities that can be discounted because they would require the victim to follow through with a noncredible threat to refuse modification. Thus, continuing to withhold performance, even though unprofitable standing alone, can become profitable (and thus credible) because it is expected to result in a profitable modification. In a sense, the threat is credible each day because it creates a profitable reputation for credibility on following days.⁷²

Bar-Gill and Ben-Shahar acknowledge that such a staged breach would be credible even though each stage is costly to the threatener, but because they employ a credibility test, they reach the mistaken conclusion that modifications induced by such breaches should be enforceable.⁷³ Their conclusion is mistaken because enforcing the modification encourages such contrived threats even though the prospect of them makes both parties worse off. A contrived-threat test instead correctly denies enforcement of the modification because no stage of the threatened action would have occurred without the prospect of such a modification, given that each stage of the threatened action is unprofitable but for its ability to induce modification.

b) A contrived threat can be credible because it creates a reputation for carrying out future threats. Once a threat is made, failing to carry it out can have reputational effects for future threats. It can thus be credible that a threatener will carry out a contrived threat (even though the threatened action is directly unprofitable) because the threatener profits from enhancing the credibility of her future threats.

⁷² The same theory works for more-violent threats. The loan shark need not resort immediately to killing the nonpaying customer. Taking one finger off at a time can be more effective. Even though carrying out each threat to amputate a finger is unprofitable standing alone, it becomes profitable (and thus credible) because carrying out such a threat is likely to convince the victim that other fingers will follow and thus induce him to pay up.

⁷³ Bar-Gill and Ben-Shahar, 33 J Legal Stud at 417, 420–21 (cited in note 57) (concluding that any modification induced by a credible threat should be enforced and that threatening such a staged breach is credible).

Bar-Gill and Ben-Shahar acknowledge not only that this factor can make unprofitable threats credible, but also that it creates an endogeneity problem for their approach, because it means that whether a threat is credible can depend on whether contract law enforces modifications induced by such threats.⁷⁴ Suppose that carrying out a threat will reduce profits unless it enhances a reputation for carrying out similar threats, which will lead to future profitable modifications. If so, then the credibility of the threat turns entirely on whether contract law will enforce such modifications. If contract law does, then the threat is credible and the credibility test concludes that contract law correctly enforces such modifications. If contract law does not, then the threat is not credible and the credibility test concludes that contract law correctly refuses to enforce such modifications. Either conclusion is consistent with their test, which thus offers no basis for choosing between them. In other words, because their credibility test depends on whether contract law *does* enforce the threat, it cannot provide independent grounds to decide whether contract law *should* enforce the threat.

Although Bar-Gill and Ben-Shahar correctly conclude that analysis should take this endogeneity problem into account, their credibility test offers no answer to the question of which of the endogenous solutions one should choose. The answer instead depends on whether deterring the threat is desirable, which a credibility test does not answer because credible threats can be either contrived (and thus harmful) or uncontrived (and thus beneficial).

A contrived-threat test, in contrast, avoids any endogeneity problem. Contrived threats should make modifications unenforceable because in the no-threat world the threatened actions would not occur. Thus, no reputation could be developed, and profits from similar future modifications would be unavailable. The contrived-threat test accordingly always provides a clear conclusion that contract law should void modifications induced by such contrived threats, no matter how credible they may be.

c) A contrived threat can be credible because making the threat creates emotional benefits to carrying it out. Given human nature, the mere act of making a threat can alter the psychic benefits of carrying it out. A person who fails to carry out a

⁷⁴ Bar-Gill and Ben-Shahar, 83 Tex L Rev at 725–26 (cited in note 5) (noting that “credibility is endogenous” because “the legal definition of duress is one of the factors that can affect the credibility of the threat”).

threat loses face and may feel obliged to live up to her threat. Indeed, a threatener may develop a character for carrying out her threats precisely because it makes her threats more credible in ways that benefit her. Developing such a character can give threateners psychic benefits from carrying out threats that cause them to do so even when there are no future reputational benefits.⁷⁵

Bar-Gill and Ben-Shahar acknowledge that emotional factors can make it credible to carry out an unprofitable threat.⁷⁶ However, they do not consider the possibility that making the threat can itself change the emotional calculus. A threat would be contrived if, absent making any threat, the psychic benefits would not suffice to take the unprofitable action. But making this contrived threat can increase the psychic benefits from taking that action in a way that offsets the lost profits and thus makes the threat credible. Under the credibility test, this would make modifications induced by such threats enforceable, which would encourage such contrived threats because those modifications would be profitable for the threatener. In contrast, under the contrived-threat test, the threatener has no incentive to make the threat in the first place given that any modification would be unenforceable. The credibility test is thus less desirable, because the proof above shows that such contrived threats can only harm both parties.

d) A contrived threat can be credible because of efforts to improve the threat's credibility. Contrived threats can be made in ways that create formal commitments to carry them out and thus make them credible. Threateners might also take earlier steps that increase the credibility of their contrived threats. For example, if a threat to breach would be contrived because the prospect of damages makes breach unprofitable, the threatener could take on debt that reduces the expected amount she will pay in damages. Alternatively, if a threat to breach would be contrived because performance would be profitable, the threatener could take steps to increase her costs (or to fail to avoid

⁷⁵ For example, a loan shark might find it advantageous to develop a character for severely punishing nonpayment, and that character might thus cause him to carry out a threat to kill someone who declines to pay, even though he hides the killing in a way that means it can have no reputational effects.

⁷⁶ Bar-Gill and Ben-Shahar, 83 Tex L Rev at 726–27, 756–57 (cited in note 5). See also generally Oren Bar-Gill and Omri Ben-Shahar, *Threatening an “Irrational” Breach of Contract*, 11 S Ct Econ Rev 143 (2004).

such cost increases) in order to make performance unprofitable and thus make a threat to breach credible.

Such efforts may ultimately make a threat credible and thus make the modifications induced by it enforceable under a credibility test. If those modifications are sufficiently profitable to the threatener, they can make such efforts worthwhile even though the efforts would be unprofitable if the party were unable to make the threat. Bar-Gill and Ben-Shahar acknowledge that this effect is both possible and undesirable, and they recommend deviating from their credibility test when a party engages in efforts to make her threat more credible.⁷⁷ A contrived-threat test would instead lead directly to the desirable conclusion because, absent an ability to make the threat, such unprofitable efforts to make a threat more credible would not occur. Accordingly, without the ability to communicate such threats, neither those contrived efforts nor the threatened actions would occur. The threat is thus contrived and modifications induced by it should be unenforceable.⁷⁸

e) A contrived threat can be credible when there are low odds of great harm. A contrived threat that is very unlikely to be carried out can still be credible enough to induce agreement given the magnitude of the threatened harm. For example, suppose that someone puts a gun to your head unless you agree to modify a contract. Carrying out the threat harms the threatener because pulling the trigger gets her nothing, incurs legal risks, and costs a bullet. But suppose there are 1 percent odds that the threatener will carry out the threat because of reputational effects or because she may be irrational, deranged, or on drugs. Even if you accurately perceive those low odds, you would likely consent to modification because a 1 percent risk of death exceeds the harm from the modification. The threat can thus be credible enough to induce modification even though it is clearly contrived. A contrived-threat test instead makes the modification

⁷⁷ Bar-Gill and Ben-Shahar, 83 Tex L Rev at 732–34 (cited in note 5); Bar-Gill and Ben-Shahar, 33 J Legal Stud at 415–17 (cited in note 57). See also Mark Seidenfeld and Murat C. Mungan, *Duress as Rent-Seeking*, 99 Minn L Rev 1423, 1427 (2015) (concluding that credible threats should be deemed wrongful when the costs of making the threat credible and carrying it out exceed the benefits of the threatened action).

⁷⁸ Consistent with my conclusion, courts have interpreted the good faith requirement of a “legitimate commercial reason” to mean a reason “outside the control of the party seeking the modification.” *T & S Brass and Bronze Works, Inc v Pic-Air, Inc*, 790 F2d 1098, 1105 (4th Cir 1986). Thus, a party cannot bootstrap itself into a valid reason for breaching by contrived efforts to incur costs, debts, or commitments that make performance unprofitable.

void and thus eliminates any incentive to make the threat in the first place. Enforcing such modifications would instead encourage such threats, which clearly seems undesirable.

2. Why contrived threats should be deterred even if they are not credible.

Bar-Gill and Ben-Shahar's theory also has problematic implications for noncredible threats. First, Bar-Gill and Ben-Shahar conclude that modifications produced by noncredible threats should *also* be enforceable whenever victims accurately perceived the noncredibility of the threat.⁷⁹ They reason that, in such cases, the threat could not have induced the modification. Their approach thus limits the possible scope of duress doctrine to noncredible threats that the victims wrongly perceived to be credible.

Efforts to make contrived threats look credible are certainly harmful. For example, a threatener might claim that her costs or debts are higher than they are, suggesting that she is likely to breach if no modification is obtained, even though breach is in fact unprofitable.⁸⁰ Such claims are directly policed by contract doctrines that deem threats to be in bad faith when they are dishonest in fact.⁸¹ However, as the preceding Section shows, a contrived threat can be credible even if the threatener is completely honest about her situation, so contract law correctly also deems threats to be in bad faith whenever they are unsupported by objective reasons that would produce breach.

Nor is there any reason for the law to get into complicated inquiries about whether a contrived threat was perceived to be credible. We are always better off if such a contrived threat is not made, rather than forcing the other side to guess about its credibility. If the contrived threat induces an adverse modification, it must have been perceived to be credible enough. As noted above, a threat can be credible enough even if the victim knows the odds are very low that it will be carried out and even if fully carrying out the threat is clearly unprofitable but it can be carried out in steps. Courts inquiring about victims' perceptions will often make errors, and may even suffer from hindsight bias,

⁷⁹ Bar-Gill and Ben-Shahar, 33 J Legal Stud at 393–94 (cited in note 57); Bar-Gill and Ben-Shahar, 83 Tex L Rev at 727–28, 730–31 (cited in note 5).

⁸⁰ See Graham and Peirce, 52 L & Contemp Probs at 11, 23 (cited in note 68).

⁸¹ See, for example, *Roth Steel Products v Sharon Steel Corp*, 705 F2d 134, 146 (6th Cir 1983).

because courts will have access to information that the victims do not have about the threats' credibility.

Moreover, even a contrived threat that fails to induce modification is harmful. If the contrived threat fails to induce modification because the victim wrongly perceived that it was not credible, then the threat will be carried out, the prospect of which harms both parties *ex ante*. If the contrived threat fails to induce modification because the victim correctly perceived that it was not credible, then there is still no benefit to allowing the threat; to the contrary, there is some harm because the threat leads to unnecessary conflict and negotiation costs. Because contrived threats never have any benefit and can create great harm, it is better for contract law to always deter them by eliminating any potential profits from making them.

Second, Bar-Gill and Ben-Shahar's approach produces no theory about which of the noncredible threats that are wrongly perceived to be credible should make induced modifications voidable; instead, they conclude that this issue turns on normative considerations beyond their analysis.⁸² My analysis provides a clearer conclusion: all modifications induced by contrived threats should be void, whether or not those contrived threats were credible.

3. The credibility test does not fit contract law.

A final advantage of the contrived-threat test is that it fits well with existing contract-law results, explaining why contrived threats are improper regardless of whether they will be carried out. In contrast, as Bar-Gill and Ben-Shahar acknowledge, their theory conflicts with current contract law.⁸³ Indeed, their insistence that credible threats should never void modifications directly conflicts with the no-reasonable-alternative element, because credible threats are the ones most likely to leave victims with no reasonable alternatives. Their approach thus immunizes the worst kind of contrived threat: a credible one that will be carried out only because the threat was made. Current law also

⁸² Bar-Gill and Ben-Shahar, 83 *Tex L Rev* at 744 (cited in note 5) ("While we argue that whenever a threat is credible the deal should be enforced, we do not argue that whenever a threat is not credible, the deal should not be enforced. . . . A normative theory . . . is necessary to determine which among these noncredible threats are coercive."); *id* at 731–32.

⁸³ *Id* at 719, 721, 737, 753–54; Bar-Gill and Ben-Shahar, 33 *J Legal Stud* at 392–94, 422–23 (cited in note 57).

eschews any separate inquiry into whether a reasonable person would have perceived the threat to be credible, concluding (as I do) that it suffices if the threat actually induced the modification.⁸⁴

F. Comparing Other Proposed Tests

1. Professors Nozick and Schelling on threats versus warnings.

My contrived-threat test is partly related to the distinction that Nozick, Schelling, and others have drawn between impermissible “threats” on the one hand and permissible “offers” and “warnings” on the other.⁸⁵ Nozick’s analysis is particularly important because it has influenced legal scholarship on contractual duress and unconstitutional conditions. However, their distinctions and analyses differ from mine in various important respects.

First, while I use a pure prediction baseline to measure threats and warnings, Nozick and other philosophers combine a prediction baseline with a moral baseline, and thus their test condemns some “threats” that I would deem uncontrived warnings. Although Nozick defines an impermissible “threat” to exist when the threatened conduct would leave the recipient worse off than the expected course of events, he defines “expected” to “straddle *predicted* and *morally required*.”⁸⁶ Thus, his definition of an impermissible threat includes adverse deviations *not only* from what predictably would occur without any threat (which I would call a contrived threat) *but also* from what is morally required, even though it conforms to what would predictably occur without the threat (thus including what I deem uncontrived warnings).⁸⁷

⁸⁴ See Restatement (Second) of Contracts § 175, comment c (1981) (“[T]he question is, did the threat actually induce assent on the part of the person claiming to be the victim of duress. . . . [I]t is not essential that a reasonable person would have believed that the maker of the threat had the ability to execute it.”).

⁸⁵ See Robert Nozick, *Coercion*, in Sidney Morgenbesser, Patrick Suppes, and Morton White, eds., *Philosophy, Science, and Method* 440, 447–58 (St Martin’s 1969); Thomas C. Schelling, *The Strategy of Conflict* 123 n 5 (Harvard 1960).

⁸⁶ Nozick, *Coercion* at 447, 450 (cited in note 85) (emphasis in original). Other scholars likewise define threats by combining prediction and moral baselines. See, for example, Alan Wertheimer, *Coercion* 206–21 (Princeton 1987); Scott Altman, *Divorcing Threats and Offers*, 15 L & Phil 209, 211–15 (1996).

⁸⁷ See, for example, Nozick, *Coercion* at 450 (cited in note 85).

One problem with Nozick's definition is that he never defines what is "morally required" (which he equates with "morally expected"), which adds a vague, conclusory element to his analysis.⁸⁸ Indeed, the typical contract modification is induced by threats to breach, which some might argue always deviate from what is morally expected,⁸⁹ a conclusion that would mean that such threats are always impermissible and the modifications are always voidable.⁹⁰ This would conflict with current law and be undesirable because the proof above shows that enforcing modifications induced by uncontrived "threats" is desirable.

No matter what he means by it, Nozick's moral prong has bite only when it deems threats impermissible even though the threatened action conforms to the predicted course of events—that is, only when it condemns uncontrived "threats."⁹¹ The above proof shows that this bite is undesirable. To be sure, such uncontrived threats by supposition threaten action that deviates from some (unspecified) moral norm, which makes them undesirable relative to compliance with that moral norm. But that moral deviation will by definition occur if the law prevents communication of uncontrived "threats" to take that deviant action. Thus, enforcing modifications induced by such uncontrived "threats" to engage in immoral action will not only reduce such immoral conduct but also leave both parties better off, given the morally unacceptable conduct that would occur if such modifications could not be enforced.

The fact that the threatened immoral conduct would occur may suggest that the underlying problem is that the law does not sufficiently penalize that conduct. But the remedy for that is increasing that penalty. If penalties are not sufficiently large to deter the immoral conduct, banning uncontrived warnings about that conduct will merely assure that the immoral conduct occurs

⁸⁸ Id at 447, 450 ("In some such situations it will be unclear what P is morally expected to do, and hence unclear whether his statement is a threat or an offer.").

⁸⁹ See, for example, Seana Valentine Shiffrin, *The Divergence of Contract and Promise*, 120 Harv L Rev 708, 730–33 (2007).

⁹⁰ If a threat adversely deviates from what is morally expected but not from what is predicted, Nozick suggests that the threat should be condemned whenever the victim would prefer what is morally expected to what is predicted. Nozick, *Coercion* at 451 (cited in note 85). Thus, if performance were morally expected, Nozick would always condemn threats to breach because victims prefer contract performance over breach.

⁹¹ If a threat adversely deviates from what is predicted, Nozick condemns the threat even if it does not deviate from what is morally required. Id at 453. Thus, the only effect that his morally required prong can have is to condemn some threats that the predicted prong would not.

instead of preferable modifications. Moreover, there are usually good reasons why the law does not impose greater penalties on immoral conduct, such as the fear that imprecision in application of the law will also deter desirable conduct. Thus, even optimal legal regulation will leave some immoral conduct undeterred, the harm of which could be reduced by enforcing modifications induced by uncontrived warnings about that conduct.

Second, these scholars often use a different prediction baseline than I do. Although in parts Nozick uses the same no-threat baseline as I do,⁹² in other parts he and other philosophers use a *pre-threat* baseline that turns on what the victim would have expected to occur before the threat was made.⁹³ Given that nearly every person who receives a threat to breach a contract expects contract performance before he hears about the threat, this pre-threat expectation test would condemn all contract modifications induced by a threat of breach. This would not only deviate from actual contract law, but would also condemn desirable uncontrived warnings whenever the warnings were unexpected by the victims. Deterring such unexpected warnings would thus increase unexpected harm to the victims and prevent mutually desirable modifications to avoid it. Schelling uses a different pre-threat baseline to define a “warning,” which under his approach turns on whether, immediately before making the threat, the threatener would have found it in her interest to take the threatened action.⁹⁴ Schelling’s definition of a warning would thus include cases in which parties make contrived commitments that give them independent incentives to take the threatened action, which I note above should be condemned as contrived threats.⁹⁵ Schelling also draws no normative conclusions about what he calls threats because his work is instead about analyzing (quite brilliantly) the credibility and effectiveness of threats.⁹⁶

Third, Nozick concludes that contrived threats should be deemed permissible “offers” when they induce agreements that the recipient feels leave him better off than the predicted course of events.⁹⁷ That is, even if a contrived threat results in

⁹² Id at 453–57.

⁹³ Id at 447–53. See also Wertheimer, *Coercion* at 207 (cited in note 86).

⁹⁴ Schelling, *The Strategy of Conflict* at 123 n 5 (cited in note 85).

⁹⁵ See Part I.E.1.d.

⁹⁶ Schelling, *The Strategy of Conflict* at 121–23 (cited in note 85).

⁹⁷ Nozick, *Coercion* at 448–49 (cited in note 85).

$Cr > Tr$, Nozick concludes that it should count as a permissible offer if $Mr > Cr$.

However, there is no benefit from creating a legal exception for contrived threats that induce modifications that leave the recipient better off. If it were really true that $Mr > Cr$, then the threatener could always induce the modification without any contrived threat. Indeed, the fact that the threatener felt a threat was needed suggests that she really believed $Cr > Mr$. The best way to test whether the recipient really believed $Mr > Cr$ would be to prevent contrived threats that leave the recipient worse off than Cr . This preserves the recipient's autonomy right to choose for himself whether he prefers the modification, rather than having that right turn on whether a court concludes that the recipient really believes the modification made him worse off.

Fourth, and related to all the above points, Nozick believes that the distinction between impermissible threats and permissible offers or warnings is inescapably normative.⁹⁸ Many other scholars share this view.⁹⁹ In my account, however, the distinction is purely factual: it turns solely on what, in a no-threat world, the threatener would have done. This factual distinction has, as I prove above, clear normative implications given the consequences that flow from it. But the distinction is a factual one that does not require independent normative criteria. That is critical because we are dealing with threatened actions that are otherwise lawful and thus by definition are within the bounds in which contract law and constitutional law allow parties and governments to pursue their own normative goals.

2. Changed circumstances.

Other leading scholars conclude that the enforceability of modifications should depend on the existence of various types of changed circumstances. Although such changed circumstances often correctly indicate an uncontrived warning that should lead to enforceability, their tests are under- and overinclusive.

⁹⁸ *Id.* at 449.

⁹⁹ See, for example, Charles Fried, *Contract as Promise: A Theory of Contractual Obligation* 95–99 (Harvard 1981); Sullivan, 102 Harv L Rev at 1446 n 133 (cited in note 6) (collecting sources).

Professor Alan Schwartz concludes that modifications should be enforceable when prompted by any changed circumstances.¹⁰⁰ However, many changed circumstances will not even make performance unprofitable, let alone so unprofitable that threateners would not perform even if modification were impossible. Unless the changed circumstances have that effect, threats not to perform are contrived threats despite changed circumstances, and enforcing such modifications under Schwartz's overinclusive test will thus encourage contrived threats, with undesirable consequences.

Judge Richard Posner concludes that modifications should be enforceable when changed circumstances make performance by a willing party impossible without a modification.¹⁰¹ However, even if performance is possible, it will not actually occur if changed circumstances make performance less profitable than breach. In such cases, the threat to breach is an uncontrived warning and failing to enforce modifications under Posner's underinclusive test will thus lead to breaches that both parties find worse than modification.

Professor Varouj Aivazian, Professor Michael Trebilcock, and Judge Michael Penny conclude that modifications should be enforceable only when changed circumstances affect risks that are remote and for which the threat recipients are superior risk bearers.¹⁰² They reason that other modifications reallocate contractual risks away from superior risk bearers.¹⁰³ However, even when their test is not met, threateners would (if modification were impossible) still breach when doing so would be more profitable than performance. When that is so, threats to breach are uncontrived warnings, and their underinclusive test will prevent desirable modifications. True, modifications induced by such contrived warnings may, as they argue, reallocate risks in a way that is less efficient than the original contracts.¹⁰⁴ But if the modifications are prevented, we will not get performance of the original contracts. We will get breaches that all parties regard as

¹⁰⁰ Alan Schwartz, *Relational Contracts in the Courts: An Analysis of Incomplete Agreements and Judicial Strategies*, 21 J Legal Stud 271, 312 (1992).

¹⁰¹ Richard A. Posner, *Gratuitous Promises in Economics and Law*, 6 J Legal Stud 411, 421–22 (1977).

¹⁰² Varouj A. Aivazian, Michael J. Trebilcock, and Michael Penny, *The Law of Contract Modifications: The Uncertain Quest for a Bench Mark of Enforceability*, 22 Osgoode Hall L J 173, 196–97 (1984).

¹⁰³ See id at 193–94.

¹⁰⁴ Id (arguing that because risks in a contract presumably “have been assigned to the superior . . . risk bearer,” allowing a party to “reallocate [risk] to the other party by contract modification” creates inefficiencies).

worse than the modifications. Aivazian, Trebilcock, and Penny worry that the prospect of risk reallocation through modification will reduce threateners' incentives to avoid risks that make them want to breach.¹⁰⁵ But the contrived-threat test eliminates that concern by deeming threats contrived if nonperformance became advantageous only because of changed circumstances that resulted from contrived decisions that the threatener would have found unprofitable but for the ability to make a threat that induces modification.¹⁰⁶

Professor Jason Johnston concludes that modifications should be enforceable only when sought because changed circumstances made performance of the original contract unprofitable.¹⁰⁷ Although such changes in circumstances often indicate an uncontrived warning, his test is both under- and overinclusive.

First, even if changed circumstances make performance of the original contract unprofitable, sometimes the prospect of paying expectation damages is even less profitable. In such situations, threats not to perform are contrived despite those changed circumstances, and enforcing modifications under Johnston's test will encourage such contrived threats, with undesirable consequences. Thus, courts are right to require further evidence that unprofitability really would have led to nonperformance even if no modification were possible.¹⁰⁸

Second, even if changed circumstances do not make performance of the original contract unprofitable, threats not to perform may be uncontrived warnings. For example, suppose that performing the contract remains profitable, but a third party has offered a higher price that would be even more profitable. Suppose further that the threatener prefers to take the higher price (and risk contract damages) over performing the original contract. If so, threats not to perform are uncontrived warnings, and refusing to enforce modifications will thus undesirably lead to breaches that make both parties worse off. Thus, contract law is right to enforce modifications when a party would otherwise take an even more profitable offer.¹⁰⁹

¹⁰⁵ Id at 194.

¹⁰⁶ See Part I.E.1.d.

¹⁰⁷ Jason Scott Johnston, *Default Rules/Mandatory Principles: A Game Theoretic Analysis of Good Faith and the Contract Modification Problem*, 3 S Cal Interdisc L J 335, 339–41, 366, 375 (1993).

¹⁰⁸ See note 34 and accompanying text.

¹⁰⁹ See notes 30–32 and accompanying text. Posner would go even further and enforce any modification if the threat to terminate were prompted by a third party offering a

Third, even if changed circumstances make performance of the original contract unprofitable, that may not justify modifications induced by contrived threats unrelated to performance. For example, in *Wolf*, the buyers' divorce was a changed circumstance that made performance unprofitable because they consequently had to resell the house.¹¹⁰ However, their threat was not to breach, but rather to complete the house purchase and then resell to an undesirable purchaser, which was even more unprofitable. That threat remained contrived, and thus the court was correct not to enforce the modification.

II. UNCONSTITUTIONAL CONDITIONS

The preceding Part establishes that contract-law standards correctly deem a threat invalidly coercive only if two elements are both met: (1) the threat was contrived (and thus "improper") and (2) carrying out the threat would leave the victim significantly worse off (thus meaning that he had "no reasonable alternative" but to agree). This Part establishes that those same two elements should and do apply to the unconstitutional-conditions doctrine. Part II.A establishes that those elements explain the Spending Clause case law on unconstitutional conditions affecting states. Part II.B provides a general proof of the normative desirability of the contrived-threat test and establishes that the same two elements explain the case law on unconstitutional conditions affecting private persons.

A. Unconstitutional Conditions Affecting States

The Medicaid-defunding-threat decision follows straightforwardly from the above contractual analysis. As the Court has noted, its Spending Clause precedents incorporate contract-law standards on duress, viewing the relationship between the federal government and the states as contractual in nature.¹¹¹ The

higher price. Posner, 6 J Legal Stud at 424 (cited in note 101). His position is overbroad because such threats are uncontrived warnings only if the higher price would have led to nonperformance (despite the risk of contract damages) even if no threat could be communicated.

¹¹⁰ See Part I.B.2.

¹¹¹ See note 4 and accompanying text. A key premise for the contractual characterization was that the federal government could not directly command states to accept the Medicaid expansion. *National Federation of Independent Business v Sebelius*, 132 S Ct 2566, 2602 (2012) (Roberts, joined by Breyer and Kagan) (plurality) ("NFIB"). If the federal government could do so, it would not matter whether the defunding threat was coercive, because coercion would be permissible.

threat to defund Medicaid unless states accepted the Medicaid expansion thus violated the relevant standard if it was the case both that (1) the threat was contrived and (2) carrying out the threat would have left the states significantly worse off.

1. Findings showing a contrived threat.

Much of Chief Justice John Roberts's opinion in *National Federation of Independent Business v Sebelius*¹¹² ("NFIB") focused on the second element because it was important to distinguish some key precedent.¹¹³ But before doing so, he offered an analysis that established that the first element was met by the threat to take away preexisting Medicaid funds, stating:

[Proposition 1] The States claim that this threat serves no purpose other than to force unwilling States to sign up for the dramatic expansion in health care coverage effected by the Act. Given the nature of the threat and the programs at issue here, we must agree. [Proposition 2] We have upheld Congress's authority to condition the receipt of funds on the States' complying with restrictions on the use of those funds, because that is the means by which Congress ensures that the funds are spent according to its view of the "general Welfare." [Proposition 3] Conditions that do not here govern the use of the funds, however, cannot be justified on that basis. When, for example, such conditions take the form of threats to terminate other significant independent grants, the conditions are properly viewed as a means of pressuring the States to accept policy changes.¹¹⁴

Proposition 1 was a finding that the threat had "no purpose" but to force states to accept the modification.¹¹⁵ This finding meant the threat was contrived—that is, absent an ability to make the threat, Congress would not have terminated preexisting Medicaid. Proposition 2 further indicates that threats are acceptable when they reflect Congress's view that the use of its funds advances the general welfare only if certain conditions are met.

¹¹² 132 S Ct 2566 (2012).

¹¹³ Id at 2604–05 (Roberts, joined by Breyer and Kagan) (plurality).

¹¹⁴ Id at 2603–04 (Roberts, joined by Breyer and Kagan) (plurality) (material in brackets added).

¹¹⁵ Id at 2603 (Roberts, joined by Breyer and Kagan) (plurality). For a similar conclusion, see Mitchell N. Berman, *Coercion, Compulsion, and the Medicaid Expansion: A Study in the Doctrine of Unconstitutional Conditions*, 91 Tex L Rev 1283, 1323, 1334–35 (2013).

Such a congressional view would indicate that Congress would find cutting off funding preferable to continuing it without conditions, which would make a threat to cut off funding an uncontrived warning of what Congress would do if attaching the condition were impossible. Proposition 3 concludes that this is not the case when Congress threatens to terminate significant independent grants, which Congress would have no reason to discontinue but for the fact that the threat pressures states to agree to a change, thus indicating a contrived threat.

The conclusion that the threat to eliminate preexisting Medicaid funds was contrived seems entirely correct. There was no evidence that any significant faction in Congress (let alone a congressional majority) preferred eliminating Medicaid over having preexisting Medicaid continue. The threat to defund preexisting Medicaid was thus a contrived threat to engage in action that Congress would not have engaged in if it were unable to make the threat. Indeed, now that the Supreme Court has prevented the Secretary of Health and Human Services from threatening to eliminate Medicaid to induce states to accept the Medicaid expansion,¹¹⁶ we have not seen any congressional movement to eliminate Medicaid, even though Congress has gotten only more conservative since the Congress that enacted the ACA.¹¹⁷ To put it another way, Congress's clear preference ranking was Medicaid Expansion > Preexisting Medicaid > No Medicaid. If unable to impose Medicaid expansion, then Congress would continue to provide preexisting Medicaid. Thus, the threat to instead eliminate preexisting Medicaid was a contrived threat that penalized states' choices not to accept the Medicaid expansion. Indeed, Roberts explicitly confirmed that he thought the threat was contrived when he concluded that, absent the condition, "[p]ractical constraints would plainly inhibit, if not preclude, the Federal Government from repealing the existing program."¹¹⁸

Conversely, suppose that we lived in a different world where Congress really thought that preexisting Medicaid was working so poorly that the nation would be better off eliminating it rather than allowing it to continue, but also thought that the Medicaid expansion would be better than both options. That is, suppose its preference ranking were Medicaid Expansion > No

¹¹⁶ *NFIB*, 132 S Ct at 2608 (Roberts, joined by Breyer and Kagan) (plurality).

¹¹⁷ See Holly Fechner, *Managing Political Polarization in Congress: A Case Study on the Use of the Hastert Rule*, 2014 Utah L Rev 757, 761–62.

¹¹⁸ *NFIB*, 132 S Ct at 2606 n 14 (Roberts, joined by Breyer and Kagan) (plurality).

Medicaid > Preexisting Medicaid. In that imaginary world, the threat to eliminate Medicaid would be an uncontrived warning, which would make it thus constitutional. In this situation, constitutionality would be desirable because if the law prevented Congress from communicating such an uncontrived warning, Congress would have eliminated Medicaid. By definition, that would leave the states worse off because the states preferred agreeing to the Medicaid expansion to having Medicaid eliminated.

The actual case also involved an uncontrived warning that the Court did sustain. Namely, the Court held that Congress could threaten to withhold new Medicaid funds from states that did not agree to the Medicaid expansion.¹¹⁹ That threat was an uncontrived warning because the only reason Congress was willing to offer the new funds was to get the Medicaid expansion. Thus, if Congress could not condition those new funds on the states' expansion of Medicaid, then Congress would not provide the new funds at all. In other words, Congress's preference ranking was Medicaid Expansion > Preexisting Medicaid > Additional Funding Without Medicaid Expansion. If Congress could not threaten to withhold additional funding unless the states expanded Medicaid, this ranking means that Congress would just provide states with preexisting Medicaid funding. That makes the threat to withhold additional funding an uncontrived threat, which the Court allowed. A contrary ruling would have removed the option to choose Medicaid expansion with the new funds over preexisting Medicaid, which would have made states that preferred that option (as well as Congress) worse off.

Roberts again confirmed this analysis because he concluded that while the threat to withhold *new* Medicaid funds was "offering" funds, the threat to deny *existing* Medicaid funds would "penalize" states.¹²⁰ This distinction would not make sense if he were determining whether a threat was an offer or penalty relative to a baseline of what Congress legally *could* do, because Congress could legally withhold either source of funds. But it does make sense if we understand Roberts to be using a baseline of what Congress *would* have done absent the condition.

Further confirmation was provided because Roberts's analysis of *Steward Machine Co v Davis*¹²¹ concluded that the threat there was constitutional for reasons that showed it was an

¹¹⁹ Id at 2607 (Roberts, joined by Breyer and Kagan) (plurality).

¹²⁰ Id (Roberts, joined by Breyer and Kagan) (plurality).

¹²¹ 301 US 548 (1937).

uncontrived warning.¹²² In *Steward*, Congress assessed a tax to fund federal unemployment insurance, but it gave a tax credit to employers who paid into state unemployment plans.¹²³ The Supreme Court held that the threat to withhold tax credits from employers in states without unemployment plans did not coerce states to adopt such plans.¹²⁴ As Roberts correctly observed, that holding rested on the premise that “Congress was willing to direct businesses to instead pay the money into state programs only on the condition that the money be used for the same purposes.”¹²⁵ This premise meant that the threat to withhold tax credits was an uncontrived warning because, if Congress had no ability to obtain the condition, Congress would have provided no tax credit rather than give unconditioned tax credits.

To put it another way, in *Steward*, Congress’s preference ranking was Conditioned Tax Credits > No Tax Credits > Unconditioned Tax Credits. Thus, the threat to deny tax credits was just an uncontrived warning of what would happen if tax credits could not be conditioned on states’ provision of unemployment insurance. Prohibiting the conditioned tax credits would thus leave the states who accepted them worse off because their acceptance showed that they preferred the tax credits with conditions to no tax credits at all. In other words, offering the conditioned tax credits to states added an option that they might prefer to the zero tax credits that they would otherwise get, rather than penalizing a choice not to adopt state unemployment insurance by withholding unconditioned tax credits that would otherwise be forthcoming. Accordingly, *Steward* correctly concluded that Congress was using “temptation” (adding options preferable to the but-for world) rather than “coercion” (imposing penalties worse than the but-for world).¹²⁶

The joint dissenting opinion in *NFIB* wrongly suggested in dicta that the sole test of coercion was whether the no-reasonable-alternative test was met.¹²⁷ That test could condemn uncontrived warnings about denials of funding or tax credits, which would leave states worse off because the states would then suffer those denials even when they preferred to agree to

¹²² *NFIB*, 132 S Ct at 2603 (Roberts, joined by Breyer and Kagan) (plurality).

¹²³ *Steward*, 301 US at 574–76.

¹²⁴ *Id* at 585, 591.

¹²⁵ *NFIB*, 132 S Ct at 2603 (Roberts, joined by Breyer and Kagan) (plurality). See also *Steward*, 301 US at 587–91.

¹²⁶ *Steward*, 301 US at 589–90.

¹²⁷ *NFIB*, 132 S Ct at 2661 (Scalia, Kennedy, Thomas, and Alito dissenting).

the conditions necessary to avoid them. It would also deviate from the underlying contract-law principles (which the joint opinion agreed applied¹²⁸) because those principles also require a contrived (that is, improper) threat. However, this suggestion was dicta because other parts of the joint opinion indicate that the dissenting justices believed the actual congressional threat to eliminate preexisting Medicaid was contrived, stating that “Congress would surely” not want to leave our “most vulnerable groups . . . out in the cold” by eliminating Medicaid in states that did not accept the expansion.¹²⁹ Because the actual threat was contrived, eliminating the ability to make it would not (and did not) result in a Medicaid elimination that would harm the states. In any event, because the votes in the Roberts opinion were necessary for the judgment, those votes limit the holding of unconstitutionality to threats that not only leave states with no reasonable choice but also are contrived.

*South Dakota v Dole*¹³⁰ is not to the contrary. To be sure, one might reasonably conclude that the threat in that case—to withhold federal highway funding from states that did not ban underage drinking¹³¹—was a contrived threat. But some Supreme Court language suggests that the Court did not think so.¹³² More important, as stressed above, a threat can be invalidated as coercive only when it meets *two* elements: (1) the threat must be contrived *and* (2) carrying out the threat must leave the states significantly worse off (meaning that they had no reasonable alternative but to agree). Thus, even if the threat in *Dole* were contrived, that threat could not be invalidated unless it also met the second element, and *Dole* focused on the conclusion that this second element was not met, as the next Section details.

2. The lack of a reasonable alternative.

The rest of Roberts’s analysis in *NFIB* established that the second necessary element, the lack of a reasonable alternative,

¹²⁸ Id at 2659–60 (Scalia, Kennedy, Thomas, and Alito dissenting).

¹²⁹ Id at 2665 (Scalia, Kennedy, Thomas, and Alito dissenting).

¹³⁰ 483 US 203 (1987).

¹³¹ Id at 205.

¹³² Roberts distinguished *Dole* because there the condition requiring a twenty-one-year-old drinking age was directly related to Congress’s purpose in making sure that its highway spending advanced safe interstate travel. *NFIB*, 132 S Ct at 2604 (Roberts, joined by Breyer and Kagan) (plurality). Such a purpose suggests that Congress might prefer eliminating highway funding over funding highways on which underage drinkers could drive. If so, then the threat in *Dole* was an uncontrived warning.

was also met, finding that states had “no real option but to acquiesce” because eliminating federal Medicaid funding threatened a “loss of over 10 percent of a State’s overall budget.”¹³³ This finding confirms that (as under contract law) the Spending Clause version of this element merely requires showing that carrying out the threat would have left the victim significantly worse off. Roberts distinguished *Dole* on the ground that there the threat was to eliminate funding that was less than 0.5 percent of the state’s budget.¹³⁴ The opinion thus suggests that losing less than 0.5 percent of a state’s budget does not leave it significantly worse off, but that losing 10 percent of a state’s budget does.

As noted above in the discussion of the underlying contract standards, where to draw the line on what counts as a significant difference turns on administrative grounds related both to litigation costs and to concerns that inevitable inaccuracies in adjudication (such as incorrectly finding a threat to be contrived) might create overdeterrence problems.¹³⁵ It thus makes sense to limit the doctrine to cases with significant underdeterrence concerns; if carrying out a threat would leave states less than 0.5 percent worse off, then modifications induced by that threat cannot harm states any more than that.

3. The critiques.

In *NFIB*, the Government argued that the threat was uncoercive because the Medicaid expansion was so generous that it left the states better off.¹³⁶ This parallels the Nozick argument that I rejected above: that a contrived threat should be deemed a permissible offer if it compels a party to accept an agreement that makes it better off.¹³⁷ As the joint dissent pointed out, if the states really believed that the expansion left them better off, then a threat to take away preexisting Medicaid funds would have been unnecessary.¹³⁸ The fact that (with the threat now removed) many states have not accepted the Medicaid expansion¹³⁹ confirms they must not actually prefer the expansion to

¹³³ Id at 2604–05 (Roberts, joined by Breyer and Kagan) (plurality).

¹³⁴ Id at 2604 (Roberts, joined by Breyer and Kagan) (plurality).

¹³⁵ See Part I.D.1.

¹³⁶ *NFIB*, 132 S Ct at 2665–66 (Scalia, Kennedy, Thomas, and Alito dissenting).

¹³⁷ See text accompanying note 97.

¹³⁸ *NFIB*, 132 S Ct at 2666 (Scalia, Kennedy, Thomas, and Alito dissenting).

¹³⁹ See Elizabeth Weeks Leonard, *Crafting a Narrative for the Red State Option*, 102 Ky L J 381, 381 (2014) (noting that, as of one year after the *NFIB* opinion, “fourteen states have rejected [Medicaid] expansion”).

preexisting Medicaid. To be sure, many argue that this state choice is financially irrational given the huge share of federal funding for that expansion. But the legal premise for coercion analysis is that states have a constitutional autonomy right to choose for themselves whether to accept the Medicaid expansion,¹⁴⁰ and in making autonomous choices, states are free to either assess the financial implications differently or make decisions based on reasons unrelated to finances. Sometimes we admire such decisions as standing on principle. To allow such contrived threats when the federal government and courts believe the end result makes the states better off, but the states do not, would violate the posited constitutional autonomy right.

Justice Ruth Bader Ginsburg argued that the threat could not be coercive because Congress has the legal right to eliminate Medicaid funding.¹⁴¹ But Spending Clause cases incorporate contract-law principles, and her argument conflicts with the contract-law principle that a threat to engage in otherwise-lawful action can be coercive when it is contrived and leaves the victim no reasonable choice but to agree.¹⁴² Her argument also conflicts with the unconstitutional-conditions doctrine, which invalidates many threats to take action (like cutting off funding) that is itself perfectly legal.¹⁴³ These doctrines are correct because, no matter what Congress legally could do, a threat that deviates from what Congress would do without the challenged condition imposes a penalty on the state's constitutional choice.

Ginsburg also argued that the threat was not coercive because "Congress [had] not threatened to withhold funds earmarked for any other program."¹⁴⁴ Roberts disputed her premise, concluding that preexisting Medicaid and the expansion were so different that they were not "all one program."¹⁴⁵ How can we tell whether programs are separate? The answer cannot turn, as Roberts correctly concluded, on how Congress "styled" the programs, because the whole point is to impose some limit on

¹⁴⁰ See note 111. See also *NFIB*, 132 S Ct at 2601–02 (Roberts, joined by Breyer and Kagan) (plurality).

¹⁴¹ *NFIB*, 132 S Ct at 2630 (Ginsburg concurring in part and dissenting in part, joined by Sotomayor).

¹⁴² See Parts I, III.

¹⁴³ See Part II.B.2.

¹⁴⁴ *NFIB*, 132 S Ct at 2634 (Ginsburg concurring in part and dissenting in part, joined by Sotomayor).

¹⁴⁵ Id at 2605 (Roberts, joined by Breyer and Kagan) (plurality).

Congress's power.¹⁴⁶ The answer must instead depend on why it functionally matters whether programs are separate. The reason it matters is that, if the threatened cutoff concerns an independent program, there is generally no reason to think that (absent linkage to the threat) Congress would have wanted to discontinue that independent program, which means that the threat to do so is likely contrived. Thus, whether preexisting Medicaid should count as an independent program from the expansion turns not on formalisms but on whether we think Congress would have preferred to continue preexisting Medicaid in a world where it could not be linked to the expansion. Because Congress clearly would have preferred this, the two should count as separate programs.

Ginsburg further claimed that the Court's holding made no sense because Congress could instead have first repealed Medicaid and then offered a new Medicaid with the expansion.¹⁴⁷ However, her premise again conflicts with contract-law principles, which have long treated rescission and replacement as the same as modification.¹⁴⁸ Nor is her premise accurate, because repealing Medicaid would still be a contrived threat if it would not have occurred absent the prospect of inducing states to agree to the expansion.¹⁴⁹

Finally, Ginsburg argued that the threat was constitutional because Congress and the states had already agreed that Congress could amend Medicaid, and the expansion was a foreseeable use of that power.¹⁵⁰ This argument is quite different—it claims that coercion does not matter because voluntary state consent to the expansion was unnecessary, given that states had already prospectively agreed to any amendment in their earlier Medicaid contracts. This claim raises a different contract-law issue: When a contract gives one side unilateral power to amend a contract, what are the limits on how that power can be exercised without the other side's consent? Roberts's opinion that such a unilateral power can be exercised to make only minor modifications is consistent with how contract law treats such provisions,

¹⁴⁶ Id (Roberts, joined by Breyer and Kagan) (plurality).

¹⁴⁷ Id at 2629, 2636 (Ginsburg concurring in part and dissenting in part, joined by Sotomayor).

¹⁴⁸ See Restatement (Second) of Contracts § 89, comment b (1981).

¹⁴⁹ After all, "a threat need not be verbally expressed; it may be perfectly clear from actions performed what the threat is." Nozick, *Coercion* at 444 (cited in note 85).

¹⁵⁰ *NFIB*, 132 S Ct at 2630, 2638–39 (Ginsburg concurring in part and dissenting in part, joined by Sotomayor).

because without that implied limit the initial contract would contain illusory promises and thus be too indefinite to enforce.¹⁵¹ This explains why past minor Medicaid modifications were permissible even when unilaterally imposed by Congress.

4. *King v Burwell*.¹⁵²

In *King v Burwell*, the challengers argued that the ACA should be interpreted to deny federal tax credits to states that did not create their own health insurance exchanges.¹⁵³ The Court concluded that *Chevron* deference did not apply because of the extraordinary-case exception.¹⁵⁴ Because the IRS interpretation was clearly contrary to the views of the current Congress, this holding fits the pattern of the Court's past cases, which have applied the extraordinary-case exception to *Chevron* deference if and only if the agency interpretation conflicts with *current* congressional views on a politically important topic.¹⁵⁵ As I have shown before, this understanding of the extraordinary-case exception fits other evidence that the true basis of *Chevron* deference is, and should be, a current-preferences default rule that applies only when the enacting legislature's meaning is unclear and the agency action is a good proxy for current enactable preferences.¹⁵⁶ The Court then concluded that, when the relevant provisions were read in context with other statutory language and structure, the ACA did not threaten to deny federal tax credits to states that refused to create insurance exchanges.¹⁵⁷

But what if the statutory language had instead been ambiguous or (as the dissent argued) clearly did make such a threat? Such a threat would clearly have been contrived because (if making this threat were impossible) the enacting Congress surely would have preferred giving tax credits without this condition over denying tax credits. After all, as every justice recognized in

¹⁵¹ Id at 2605–06 (Roberts, joined by Breyer and Kagan) (plurality). See also Restatement (Second) of Contracts § 2, comment e (1981); Samuel Williston, 1 *A Treatise on the Law of Contracts* § 4:24 at 756 (Thomson/West 4th ed 2007) (Richard A. Lord, ed); Joseph M. Perillo and Helen Hadjiyannakis Bender, 2 *Corbin on Contracts* § 5.28 at 142 (West rev ed 1995).

¹⁵² 135 S Ct 2480 (2015).

¹⁵³ Id at 2487–88.

¹⁵⁴ Id at 2488–89.

¹⁵⁵ See Einer Elhauge, *Statutory Default Rules: How to Interpret Unclear Legislation* 101–07 (Harvard 2008) (showing this pattern in the case law).

¹⁵⁶ See id.

¹⁵⁷ *King*, 135 S Ct at 2489–96.

NFIB, the overriding goal of the Congress that enacted the ACA was universal insurance coverage.¹⁵⁸ The ACA imposes on all states community-rating and guaranteed-issue requirements that would drive healthier persons out of individual insurance markets unless coupled with both tax credits and mandates that keep the individuals in.¹⁵⁹ Because the ACA conditions the mandates to buy insurance on the federal tax credits received, withholding tax credits would also have lifted the mandates for nearly 87 percent of individuals in states that refused to create exchanges.¹⁶⁰ Withholding federal tax credits would thus have caused a collapse of state insurance markets that would have been precisely opposite to the congressional goal of universal coverage.¹⁶¹ Indeed, it would have triggered a collapse of individual insurance markets below the levels that prevailed before the ACA, which would have been contrary to the aims of even congressional opponents to the ACA.

The statutory threat that would have been created by the challengers' interpretation would also have satisfied the no-reasonable-alternative element because carrying out the posited threat would have left states significantly worse off. The direct financial harm alone would have been 6 percent of the average state budget,¹⁶² twelve times the 0.5 percent in *Dole* and closer to the 10 percent that *NFIB* held was way past the constitutional line.¹⁶³ Moreover, this financial harm would have been multiplied because the threatened action would also have caused individual insurance markets to collapse in the states that refused to establish their own exchanges. Not only would this have been harmful to their citizens and state economies (and thus to state tax revenue), but in order to deal with all the newly uninsured, these states "would almost certainly find it necessary to increase

¹⁵⁸ *NFIB*, 132 S Ct at 2606 (Roberts, joined by Breyer and Kagan) (plurality); id at 2613 (Ginsburg concurring in part and dissenting in part, joined by Sotomayor, Breyer, and Kagan); id at 2664–65 (Scalia, Kennedy, Thomas, and Alito dissenting).

¹⁵⁹ *King*, 135 S Ct at 2485–87, 2493.

¹⁶⁰ Id at 2493.

¹⁶¹ Id at 2493–94, 2496.

¹⁶² Brief of *Amici Curiae* Jewish Alliance for Law & Social Action (JALSA), Jewish Social Policy Action Network (JSPAN), Jewish Council on Urban Affairs (JCUA), Boston Alliance for Community Health, Lawyers' Committee for Civil Rights and Economic Justice, and Professors of Health Law and Constitutional Law in Support of Respondents, *King v Burwell*, Docket No 14-114, *9 (US filed Jan 26, 2015) (available on Westlaw at 2015 WL 350366).

¹⁶³ *NFIB*, 132 S Ct at 2605 (Roberts, joined by Breyer and Kagan) (plurality); id at 2663 (Scalia, Kennedy, Thomas, and Alito dissenting).

[their] own health-care expenditures substantially,” and indeed the states would have been legally obligated to fund the newly uninsureds’ emergency care in state hospitals.¹⁶⁴ The total harm to states would surely have exceeded the 10 percent that *NFIB* held was obviously unconstitutional.

Accordingly, the same principles that explain why the Court was correct to hold the Medicaid-defunding threat unconstitutional show that interpreting the ACA to threaten to withhold tax credits would have also made that tax credit threat unconstitutional. One might try to argue that the two cases differ because in *King* the posited threat would have pressured states indirectly by withholding new tax credits from its citizens. But that was precisely the form of threat at issue in *Steward*, which held that the applicable legal standard to judge such threats was the same legal standard that *NFIB* applied to the Medicaid-defunding threat.¹⁶⁵ A holding that threats to deny new benefits cannot be coercive would also conflict with the unconstitutional-conditions doctrine, which often condemns withholding new benefits unless individuals waive their constitutional rights, and with underlying contractual principles, which often condemn threats to withhold new benefits to induce new contracts.¹⁶⁶ In any event, the posited threat in *King* would not have merely withheld new benefits: given the associated regulation, it would have threatened a collapse of individual insurance markets below levels that prevailed before the ACA.

Interpreting the ACA as threatening to withhold tax credits would thus have conflicted with the canon requiring that statutes be interpreted to avoid constitutional invalidity. Indeed, as Justice Anthony Kennedy correctly stated at oral argument, the modern canon of constitutional avoidance requires avoiding any interpretation that raises significant constitutional doubt or complexities, which would certainly have been raised by interpreting the ACA to make the posited threat.¹⁶⁷ Further, even if

¹⁶⁴ Id at 2657, 2664 (Scalia, Kennedy, Thomas, and Alito dissenting).

¹⁶⁵ Id at 2603–05 (Roberts, joined by Breyer and Kagan) (plurality) (applying the *Steward* “undue influence” standard); id at 2659 (Scalia, Kennedy, Thomas, and Alito dissenting).

¹⁶⁶ See Parts II.B.2, III. The correct baseline is a predictive baseline, not a historical baseline, although history is sometimes relevant to prediction. See text accompanying notes 227–28.

¹⁶⁷ Transcript of Oral Argument, *King v Burwell*, Docket No 14-114, *16–18, 49 (US Mar 4, 2015) (available on Westlaw at 2015 WL 2399405) (“*King* Transcript”). See also *National Labor Relations Board v Catholic Bishop of Chicago*, 440 US 490, 508–09 (1979).

one (in my view, incorrectly) concluded that the statute threatened to withhold tax credits so clearly that the canon of avoidance was inapplicable, that threat would have to be invalidated as unconstitutional. Some have argued that the remedy for such unconstitutionality is unclear, but the routine remedy for unconstitutional conditions has always been voiding the condition, not denying the government benefit to everyone.¹⁶⁸ The contrived-threat test explains why: such conditions are unconstitutional only when the threat was contrived, which by definition means that (if unable to impose the condition) the enacting legislature would have preferred granting the benefit over withholding it. This same legislative preference ranking means that statutory default rules on severability, which require severing the statute in a way that is most likely to advance the enacting legislature's preferences, would have required voiding the threat to withhold benefits, rather than voiding the grant of benefits for everyone.¹⁶⁹

¹⁶⁸ That was true both in *NFIB* and in all the cases considered in Part II.B.2. When, for example, the Court held it unconstitutional to withhold a tax exemption from veterans who refused to give a loyalty oath, the remedy was to give the tax exemption to all veterans, not to deny it to all veterans. See *Speiser v Randall*, 357 US 513, 527–29, 531 (1958).

¹⁶⁹ See Elhauge, *Statutory Default Rules* at 133–34 (cited in note 155). The dissenting justices simply ignored the fact that their interpretation would have resulted in a tax credit threat that was unconstitutional under their own interpretation of the unconstitutional-coercion doctrine in *NFIB*. But one of the dissenting justices, Justice Antonin Scalia, suggested at oral argument that his view was that no prior case had ever held that the Court could “rewrite” a statute when a clear provision was held unconstitutional. *King* Transcript at *17 (cited in note 167). However, no such rewrite would have been required: the condition “by the State” could simply have been voided in the provisions that calculate tax credits based on premiums and coverage months in an “[e]xchange established by the State.” 26 USC § 36B(b)(2)(A), (c)(2)(A)(i). This would have involved, if anything, less “rewriting” than voiding the language that created the unconstitutional Medicaid-defunding threat. In any event, Scalia was mistaken in his claim that no prior case has ever held that a rewrite was the appropriate approach. In *United States v Booker*, 543 US 220 (2005), which held that mandatory sentencing guidelines without jury fact-finding were unconstitutional, the Court stressed that severability analysis does not require simply grammatically separating the invalid terms, but rather requires choosing the option that Congress would have most preferred, and it noted that this sometimes requires “modification of a statutory provision,” which the Court did by making the guidelines advisory. *Id.* at 246–49. Scalia himself acknowledged that this decision required “rewriting” the statute. *Id.* at 284–85 (Stevens dissenting in part, joined by Souter and Scalia). Moreover, Scalia favored a different interpretation that would have “rewritten” the statute as well: namely, adding jury fact-finding to the sentencing process on the ground that Congress would have preferred that instead. *Id.* at 246 (pointing out that the dissent, joined in relevant part by Scalia, required the “addition” of provisions requiring jury fact-finding); *id.* at 303–04 (Scalia dissenting in part).

B. Unconstitutional Conditions Affecting Private Persons

1. A general proof.

Unconstitutional-conditions cases involve situations in which the government has the power to deny a benefit and persons have a constitutional autonomy right not to engage in certain conduct, and the government conditions the benefit on the person not exercising that right. Sometimes courts declare such conditions unconstitutional, but sometimes they do not. The essential critique is that there is no coherent theory for explaining when courts should declare such conditions unconstitutional.¹⁷⁰ But the following proves that the contrived-threat test does provide a coherent theory of which conditions violate the constitutional right not to engage in that conduct.

All persons necessarily prefer unconditioned benefits over both conditioned benefits and no benefits. Persons who take conditioned benefits must also prefer them to no benefits, and thus for them Unconditioned Benefits > Conditioned Benefits > No Benefits. Persons who reject conditioned benefits must find them worse than no benefits, and thus for them Unconditioned Benefits > No Benefits > Conditioned Benefits.

The government threatens no benefits unless persons accept the conditioned benefits. There are two possibilities depending on whether, if conditioned benefits could not be offered, the government would prefer to provide the benefits or not.

a) Uncontrived warning. If the government prefers Conditioned Benefits > No Benefits > Unconditioned Benefits, then threats to provide no benefits are uncontrived warnings because if constitutional law did not allow the conditioned benefits, then the government would provide no benefits. Uncontrived warnings that benefits will be denied to those who reject the condition thus impose no penalty relative to the but-for world; they instead add a conditioned-benefits option that some persons prefer to the but-for option of no benefits. To be sure, some persons would be even happier with unconditioned benefits, but that is not what they would get if the conditioned benefits were unconstitutional. Prohibiting the conditioned benefits would be bad for both the government and persons who would accept the condition, because they both must prefer conditioned benefits to no benefits. Prohibiting the conditioned benefits would not affect

¹⁷⁰ See notes 3–7 and accompanying text.

persons who would reject the condition, because they would get no benefits either way. Allowing the conditioned benefits would thus make the government and some persons better off without making any persons worse off.

b) Contrived threat. If the government instead prefers Conditioned Benefits > Unconditioned Benefits > No Benefits, then threats to deny benefits to persons who do not accept conditioned benefits are contrived because if constitutional law does not allow the conditioned benefits, then the government would provide unconditioned benefits. Such contrived threats penalize decisions to reject the condition by denying unconditioned benefits that would have been given in the but-for world. Making a contrived threat thus has precisely the same effect as imposing a fine on those who exercise the constitutional right contrary to how the government wishes.

For persons who reject the condition, the contrived threat means they will get no benefits, rather than the unconditioned benefits they would have received without the threat. This effect makes these persons and the government worse off because they both prefer unconditioned benefits to no benefits.

For persons who accept the condition, the contrived threat means they will get conditioned benefits rather than unconditioned benefits. This makes these persons worse off because they prefer unconditioned benefits. But it makes the government better off because it prefers conditioned benefits to the unconditioned benefits it otherwise would have given. The benefits are the same either way. The only difference is that, with the threat, the constitutional right would be exercised how the government wishes; whereas, without the threat, the constitutional right would instead be exercised how the person wishes. But the existence of the constitutional autonomy right by definition means that the Constitution embodies the normative principle that the right should be exercised how the person prefers, unless the person receives benefits that he is willing to exchange for those rights. Adding a contrived threat converts the situation so that, without any difference in benefits received, the right is now exercised how the government prefers instead. The government has basically taken over the autonomy right without providing any offsetting benefits. The effect is identical to the effect that imposing a fine has on those persons who comply and thus do not pay the fine. Given the premise of the constitutional autonomy

right, the harm to these persons by definition exceeds the gain to the government.

In other words, contrived threats force persons who accept the condition to give up their constitutional rights in exchange for nothing, because they get the same benefits they would have gotten in the but-for world where conditioned benefits were not possible. That necessarily violates the posited constitutional right. Indeed, the contrived threat harms all persons whether or not they accept the condition. Because the government conditioned the benefits, it must find the policy gain from individuals who accept the condition greater than the policy harm from denying benefits to individuals who reject the condition, but that policy gain is unconstitutional by premise and thus illegitimate.

The above argument depends on the existence of a constitutional autonomy right, like the state right not to administer federal programs that was at issue in *NFIB*¹⁷¹ or the individual constitutional rights at issue in unconstitutional-conditions case law. The same analysis would not apply if there were no constitutional autonomy right, but instead only a claim that the federal government was exceeding its Commerce Clause power. For example, suppose that Congress attached to federal spending or tax exemptions a contrived threat to deny those benefits to individuals that do X, that individuals have no constitutional autonomy right not to be compelled to do X, but that compelling X lies beyond Congress's Commerce Clause power. In that case, the contrived nature of the threat means it is coercive—that is, regulatory—but there is no constitutional autonomy right to be free of such regulation. In these cases, the question is instead whether regulatory uses of the federal spending and tax powers should be limited to the scope of the Commerce Clause power. In a system that gives Congress multiple overlapping enumerated powers, it is not at all clear why they should be. Indeed, *NFIB* itself held the contrary, sustaining a regulatory use of the taxing power to penalize the nonpurchase of health insurance even though the Court struck down a mandate to buy health insurance as beyond the Commerce Clause.¹⁷² Consistent with this analysis, courts have not struck down contrived threats to deny benefits when the affected conduct is not protected by any constitutional autonomy

¹⁷¹ *NFIB*, 132 S Ct at 2601–02 (Roberts, joined by Breyer and Kagan) (plurality).

¹⁷² *Id* at 2594–2600 (Roberts).

right, but instead simply lies beyond the scope of the Commerce Clause power.

c) Conclusion. Accordingly, conditioning benefits on the waiver of constitutional autonomy rights violates those rights whenever the threat to deny benefits is contrived. Such conditions do not violate those rights whenever the threat to deny benefits is uncontrived. Conditioning benefits reflects a contrived threat only when, if unable to make such a threat, the government would still provide the benefits.

To put it another way, the contrived-threat test asks: In a but-for world where the condition could not have been adopted, what would the government have done? If the government would have offered the benefit in the but-for world, then denying the benefit imposes a penalty relative to that but-for world, and thus the condition unconstitutionally penalizes the exercise of a constitutional right. If the government would not have offered the benefit in the but-for world, then denying the benefit imposes no such penalty and thus does not violate the recipient's constitutional right.

This analysis shows the error in the general claim that the greater power includes the lesser. When threats are contrived, the greater power to withhold benefits entirely would not be exercised if the supposedly lesser power to condition benefits did not exist. In such cases, the power to condition benefits is not lesser at all because it enables the government to achieve effects that it would not achieve with the supposedly greater power.

Nothing in the above proof requires relying on a contract analogy or on whether benefits were provided in the past. Indeed, as we shall see, the unconstitutional-conditions doctrine fully applies to contrived threats to deny new benefits. The relevant baseline entitlements are set by the Constitution, not by the past provision of benefits. However, one could properly deem those constitutional rights to be part of the initial contract between the government and its states and citizens. Further, regardless of whether one characterizes the initial set of constitutional entitlements as contractual, courts rely on contract principles to explain why that initial set of constitutional entitlements can sometimes be waived (that is, modified) in exchange for government benefits. It thus makes sense to conclude that the same contract principles that allow some such modifications also limit others, so that the limits on contrived threats that induce agreements to modify initial contractual rights also

apply to contrived threats that induce agreements to modify initial constitutional rights. Accordingly, while not required for the proof, the contract analogy provides legal grounds for the same conclusions as that proof.

To illustrate the above points, consider the classic case of *Frost & Frost Trucking Co v Railroad Commission of California*.¹⁷³ California could legally ban truckers from its highways, but truckers had (at the time of the case) a substantive due process right not to be ordered to be common carriers.¹⁷⁴ The Supreme Court held that it was unconstitutional for a California statute to condition highway access on truckers' agreeing to be common carriers.¹⁷⁵ This holding has been critiqued because the truckers preferred being common carriers to lacking highway access.¹⁷⁶ But this critique wrongly assumes that, if unable to condition the benefit, California would have denied highway access to truckers. That seems implausible because trucking was highly valuable to the state economy. California's preference ranking was surely instead Common Carrier Condition > Unconditioned Trucker Access > No Trucker Access. Therefore, if California could not impose the common carrier condition, California would instead give truckers highway access without that condition. The threat to deny truckers access was thus a contrived threat that harmed truckers because they preferred unconditioned access to conditioned access. The Court thus correctly held the statute unconstitutional.¹⁷⁷

The threat in *Frost* was also entirely credible because the statute committed to carry out the threat and enough truckers complied that legislators had no incentive to repeal the statute. This is generally the case with unconstitutional conditions, and it illustrates both why contrived threats can be highly credible and why immunizing credible threats would produce the wrong result.

¹⁷³ 271 US 583 (1926).

¹⁷⁴ *Id.* at 592.

¹⁷⁵ *Id.* at 599.

¹⁷⁶ See Adam B. Cox and Adam M. Samaha, *Unconstitutional Conditions Questions Everywhere: The Implications of Exit and Sorting for Constitutional Law and Theory*, 5 J Legal Analysis 61, 65 (2013).

¹⁷⁷ *Frost*, 271 US at 599. Likewise, it is unconstitutional to condition the ability of corporations to do business in a state on waiving their access to federal court. *Terral v Burke Construction Co*, 257 US 529, 532–33 (1922). This holding is correct because, if unable to adopt that condition, a state would surely prefer having corporations operate in the state without the condition over having no corporations operate in the state. Thus, the threat to withhold the right to operate was contrived.

d) Avoiding analytical confusion. An important analytical benefit of the above analysis is that it carefully separates the issue of whether withholding benefits coercively violates a constitutional autonomy right not to engage in certain conduct from claims that (1) the threatened withholding is illegal, (2) the induced agreement is independently illegal even if uncoerced, or (3) the induced agreement is legal even if coerced because there is no relevant constitutional right to be free of government coercion. Past scholarship has found unconstitutional conditions intractable in part because the scholarship has often mixed up these analytically separate issues.

For example, suppose that the government provided a \$100 prize to the person who wrote the best essay arguing that the next president should be Hillary Clinton. The threat to withhold \$100 seems uncontrived because the government would not provide the \$100 if it could not attach the condition. Such a prize certainly would (and should) be unconstitutional. Does that mean the unconstitutional-conditions doctrine does properly condemn uncontrived threats? No, because the reason that this prize is unconstitutional is *not* that it coerces the prize recipient into giving up her constitutional right of free speech. Instead, it is because, even without coercion, such agreements violate equal protection because they fund speech favoring one political party over another, and they thus harm third parties even if they benefit the agreeing parties. To test this, suppose that the government instead equally funded political speech for both parties, as it has done with presidential campaigns. Then there is no constitutional violation, because the funding is equal. For other examples, whether an uncoerced agreement is independently unconstitutional may be more difficult, but analytical clarity comes from seeing that such conclusions must rest on grounds other than coercion. Just as some private contracts are illegal as a matter of public policy even if they are not produced by duress, likewise some government-person agreements are independently unconstitutional even if not produced by coercion. Analytical clarity requires separating the issues.

e) Difficulties of proof. One might be concerned that it is hard to determine whether a threat is contrived. But often, as in *Frost* and the ACA cases, it is not that hard. Nor has this proved an insuperable problem for contract law, which has to resolve the same issue. Even when it is hard, a difficult factual question beats a vacuous standard. Difficulties of proof are also a common

issue in law. The possibility of adjudicative error creates over-deterrence and underdeterrence, and we have our usual legal tools for dealing with this problem. Here, because determining whether a threat is contrived will produce litigation costs and adjudicative error that deter uncontrived warnings to some extent, it makes sense to limit this doctrine (like the contract doctrine) to cases in which underdeterrence concerns are significant by also requiring evidence that carrying out the threat would have left the victim significantly worse off.

No matter how difficult it may be for courts to figure out the true facts, we cannot avoid the reality that some threats are actually contrived while other threats are not, and that the above proof shows the effects are always harmful for the contrived threats and beneficial for the uncontrived threats. This is important because it means that any legal test that one uses must be judged by the extent to which it successfully deters contrived threats while avoiding deterring uncontrived warnings. As we shall see next, the legal tests used by the Supreme Court do fairly well on this score.

2. The fit with case law.

Supreme Court standards on unconstitutional conditions correlate, albeit imperfectly, with a contrived-threat test. The application of those standards can be improved by clarifying that, in ambiguous cases, the standards should be interpreted to implement a contrived-threat test.

a) Threats to deny funds used for government-desired objectives to those who exercise constitutional rights without using government funds. A series of cases draws the following distinction: Congress can deny funds to persons who use federal funds for activities that Congress does not want to subsidize. But Congress cannot deny funds to persons who use federal funds only for “activities Congress wants to subsidize” based on how those persons exercise constitutional rights that are not federally funded.¹⁷⁸ Although the Court itself admits this “line is hardly clear” or “self-evident,”¹⁷⁹ a contrived-threat test helps define the line.

¹⁷⁸ *Agency for International Development v Alliance for Open Society International, Inc.*, 133 S Ct 2321, 2328–32 (2013) (“AID”) (collecting cases).

¹⁷⁹ *Id.* at 2328, 2330.

Denying funds used for activities that Congress does not want to subsidize is an uncontrived warning because, if unable to impose that condition, Congress would not provide the funds, given that by definition it does not want to subsidize the activities. Allowing this condition will thus make the persons that want to accept it better off without harming anyone.

Denying funds used for activities that Congress does want to subsidize is likely a contrived threat because, if unable to impose the condition, Congress would likely provide the funds because the funds are by definition being used for activities Congress wants to subsidize. The Court thus correctly prohibits such threats, rejecting the contrary view that such threats are constitutional whenever the condition advances congressional objectives.¹⁸⁰ The problem with this contrary view is that all it means is that Congress prefers Conditioned Funds > Funding Activities Congress Wants > No Funding. Thus, if unable to use the challenged condition, Congress would still fund the activities it wants. Accordingly, its threat to instead cut off all funding is a contrived threat that should be prohibited.

Consider *Harris v McRae*,¹⁸¹ which strongly suggested that Congress could not threaten to withhold nonabortion Medicaid funding from women who had abortions using nonfederal funds.¹⁸² Critics have argued that this conclusion offers no sound theory of whether the benefit of receiving nonabortion Medicaid should be included in the transaction.¹⁸³ But this critique assumes that nonabortion Medicaid would have been withheld in the but-for world without the condition. In fact, the Court was right that without the condition, Congress would prefer funding nonabortion Medicaid to eliminating Medicaid, which is demonstrably what Congress has actually chosen ever since *Harris* was decided. A threat to deny all Medicaid funding to those who had privately funded abortions would thus be contrived. Accordingly, no “benefit” of nonabortion Medicaid would be caused by the posited condition. Rather, such a condition would coercively threaten to take a benefit that would exist without any condition.

This precedent further holds that, if the activities Congress wants to subsidize can be separated from the exercise of constitutional rights without those federal funds, then Congress must

¹⁸⁰ Id at 2328.

¹⁸¹ 448 US 297 (1980).

¹⁸² Id at 317 n 19.

¹⁸³ See, for example, Levinson, 111 Yale L J at 1345–47 (cited in note 6).

allow such separation.¹⁸⁴ Thus, Congress may not, by refusing to allow a separation of funds, use an uncontrived warning to provide funds only for activities it wants as an excuse for a contrived threat to withhold those funds because persons engaged in protected unfunded activities. That is, when separated funding is possible and legislatively preferable to denying all funding, a refusal to provide separated funding would be a contrived threat.

For example, *Rust v Sullivan*¹⁸⁵ involved federal funding for family planning.¹⁸⁶ The Court held that Congress could constitutionally deny those funds to organizations that used the funds to counsel abortions, but only because Congress permitted those organizations to separate their activities so that organizations that provided abortion counseling could still get federal funding for other family planning activities.¹⁸⁷ The refusal to fund abortion counseling was an uncontrived warning because Congress preferred not to fund such activities. But had Congress instead refused to provide family planning funds to organizations that used separate nonfederal funds to counsel abortions, then that would have been a contrived threat because Congress did want to fund family planning services other than abortions.

In contrast, in *Agency for International Development v Alliance for Open Society International, Inc.*,¹⁸⁸ the Court held that federal funding to combat HIV/AIDS could not constitutionally be denied to organizations that refused to oppose prostitution, even though it could constitutionally be denied to organizations that used those federal funds to advocate the legalization of prostitution.¹⁸⁹ The condition requiring that federal funds not be used to advocate the legalization of prostitution was an uncontrived warning because Congress did not want to fund such advocacy. But the condition denying federal funds to organizations that refused to oppose prostitution, even if they used those funds to combat HIV/AIDS in ways other than advocating the legalization of prostitution, was a contrived threat because Congress did want to fund efforts to combat HIV/AIDS in other ways and was preventing the separation of such funding.¹⁹⁰

¹⁸⁴ *AID*, 133 S Ct at 2328–32.

¹⁸⁵ 500 US 173 (1991).

¹⁸⁶ *Id* at 178–79.

¹⁸⁷ *Id* at 196–97.

¹⁸⁸ 133 S Ct 2321 (2013).

¹⁸⁹ *Id* at 2330–32.

¹⁹⁰ *See id.*

b) *Penalties versus subsidies*. The Supreme Court has held that conditioning government benefits on persons giving up their constitutional rights to travel, speak, or observe their Sabbath is an unconstitutional penalty,¹⁹¹ but that conditioning benefits on persons giving up their rights to strike or avoid mandatory home visits is a constitutional subsidy.¹⁹² Critics have argued that these two lines of cases are inconsistent.¹⁹³ But whether they are consistent depends on factual predictions of what the government would likely do if unable to adopt these conditions.

In the former line of cases, it is clear that, if left with a choice between providing the benefits without the challenged condition and not providing the benefits at all, the government would choose to provide the benefits. Indeed, that is what has happened. The threats to deny those benefits were thus contrived threats to impose penalties relative to the but-for world.

The validity of the latter line of cases turns on whether one thinks that, if unable to adopt the challenged conditions, the government would have preferred to withhold the benefits rather than provide them without those conditions. The Court found that Congress adopted the no-strike condition because it wanted to make sure food stamps were not used to fund strikes,¹⁹⁴ and it concluded that New York adopted the home visit condition because it wanted to make sure welfare benefits were being used to advance the best interests of children.¹⁹⁵ If the desire to avoid such unwanted uses of federal funds were strong enough, then (if the conditions were unavailable) Congress might well prefer withholding these benefits to providing benefits without conditions that would prevent unwanted uses. If so, then the threats to deny those benefits were uncontrived warnings, and the Court was right that they did not impose penalties. The Court's implicit factual predictions are not self-evidently wrong, and if they are correct, the two lines of cases are normatively consistent. But courts should clarify the penalty/subsidy distinction by explicitly

¹⁹¹ See *Shapiro v Thompson*, 394 US 618, 641–42 (1969) (addressing a condition affecting the freedom to travel); *Speiser*, 357 US at 528–29 (addressing a condition affecting the freedom of speech); *Hobbie v Unemployment Appeals Commission of Florida*, 480 US 136, 146 (1987) (addressing a condition affecting the right to observe Sabbath).

¹⁹² See *Lyng v International Union, United Automobile, Aerospace, & Agricultural Implement Workers of America, UAW*, 485 US 360, 373–74 (1988) (addressing a condition affecting the right to strike); *Wyman v James*, 400 US 309, 326 (1971) (addressing a condition affecting the right to avoid mandatory home visits).

¹⁹³ See, for example, Sullivan, 102 Harv L Rev at 1433–39 (cited in note 6).

¹⁹⁴ *International Union*, 485 US at 370–71.

¹⁹⁵ *James*, 400 US at 315–16, 318–19.

interpreting those terms to be measured relative to what the government would have done without the challenged condition.

c) *Nexus or germaneness*. Another line of cases concerns situations in which the government can deny a development permit outright, but cannot directly take some of the owner's property rights without just compensation. The government threatens to deny the permit unless the owner agrees to a condition that exacts some of its property rights, such as requiring the owner to give an easement to the public. The Supreme Court has held that such threats are constitutional only if the condition's purpose has an "essential nexus" to whatever purpose would be served by withholding the permit.¹⁹⁶ The Court has stressed that this doctrine is just a "special application" of the unconstitutional-conditions doctrine.¹⁹⁷ More generally, the pattern of the Court's results on unconstitutional conditions indicates that the Court is more willing to uphold a condition that is "germane" to any government reasons for withholding the benefit.¹⁹⁸

If a nexus or germaneness exists between the condition and the withholding of the benefit, then a government that cannot use the condition would be more likely to advance the same purpose by withholding the benefit. Offering an unconditioned benefit would be less likely in the no-condition world because that would not serve the same purpose. Accordingly, a nexus or germaneness indicates that threats to withhold benefits are likely uncontrived warnings that should make the condition constitutional.

If no nexus or germaneness exists between the condition and the withholding of the benefit, then an inability to use the condition would be less likely to cause the government to withhold the benefit, because the two serve different purposes. Offering an unconditioned benefit would be more likely in the no-condition world because giving the benefit would serve the same purpose as the benefit that would be given in a conditioned benefit. (For example, providing a conditioned permit and providing an unconditioned permit both supply whatever policy gains flow from real estate development.) Thus, a lack of nexus or germaneness

¹⁹⁶ *Dolan v City of Tigard*, 512 US 374, 386 (1994); *Nollan v California Coastal Commission*, 483 US 825, 837 (1987).

¹⁹⁷ *Koontz v St. Johns River Water Management District*, 133 S Ct 2586, 2594 (2013).

¹⁹⁸ Sullivan, 102 Harv L Rev at 1458–68 (cited in note 6) (summarizing cases).

indicates that threats to withhold benefits are likely contrived threats that should make the condition unconstitutional.¹⁹⁹

However, even if the condition and the denial of a permit further very different purposes, the government may value those different purposes equally, in which case different purposes need not suggest a contrived threat. For example, *Nollan v California Coastal Commission*²⁰⁰ found no nexus between the purpose of the condition (providing public access to the beach) and the purpose of denying the permit (protecting visual access to the ocean).²⁰¹ Although those purposes are different, the government might value them the same and thus value denying the permit similarly to providing a conditioned permit. If so, then holding such conditioned permits unconstitutional will prevent uncontrived warnings and result in the permits being denied outright.²⁰² This is by definition worse for the property owner and the government, because both prefer the conditioned permit to no permit. We can thus improve the doctrine by interpreting nexus and germaneness to mean a similar government valuation of the condition and denial that is likely to make the threat uncontrived.

Contrived-threat theory can also illuminate other features of the doctrine used to determine when conditioned permits are unconstitutional takings. In addition to imposing a nexus requirement, the case law has stressed that there must be some “rough proportionality” between the condition and the social costs imposed by granting the permit.²⁰³ The explanation for this proportionality element under contrived-threat theory parallels the explanation for the contract doctrine that makes the fairness of a modification relevant to judging whether a threat used to induce that modification is enforceable.²⁰⁴ The existence of a

¹⁹⁹ Some argue that any ban on conditioned permits without the needed nexus could easily be evaded by using eminent domain to take the easement and then giving the permit as just compensation. See, for example, Levinson, 111 Yale L J at 1349 (cited in note 6). But if the threat to deny the permit were contrived, then the permit would have been given in the but-for world, and thus giving the permit would provide no compensation at all. Giving the permit would provide compensation only if the threat were uncontrived, in which case the conditioned permit would be constitutional anyway.

²⁰⁰ 483 US 825 (1987).

²⁰¹ Id at 836–37.

²⁰² See William A. Fischel, *Regulatory Takings: Law, Economics, and Politics* 346 (Harvard 1995) (suggesting that the nexus doctrine may often result in permit denials that are worse than the conditioned permits that would otherwise be negotiated).

²⁰³ *Koontz*, 133 S Ct at 2595; *Dolan*, 512 US at 391.

²⁰⁴ See Part I.D.3.

nexus may show that a threat to deny a permit is uncontrived in the sense that the government would truly prefer no development to an unconditioned development. However, a threat to deny a permit unless an unreasonably onerous condition were accepted could still be contrived, because a more reasonable condition would have sufficed to make granting the permit preferable to denying it. This provides clearer guidance on what proportionality means: it means the condition should not exceed the minimum necessary for the government to be willing to grant the permit.

Other issues raised in the recent *Koontz v. St. Johns River Water Management District*²⁰⁵ decision can also be readily explained in terms of contrived-threat theory. First, *Koontz* made clear that threats that violated the nexus-proportionality test were unconstitutional even if the threatened party refused to accept the condition and thus the threatened permit denial was carried out.²⁰⁶ This fits contrived-threat theory because contrived threats are undesirable whether or not they are carried out, and the parallel contract doctrine prohibits them even when they do not induce modifications.²⁰⁷ Second, although the *Koontz* majority did not resolve the issue of how concrete the demand for a condition must be for it to be unconstitutional,²⁰⁸ the dissent by Justice Elena Kagan, joined by Justices Ruth Bader Ginsburg, Stephen Breyer, and Sonia Sotomayor, discussed that issue at length.²⁰⁹ And although framed as an inquiry into the definiteness of the demand, the dissent's reasoning indicated that the ultimate basis for the dissent's conclusion was not the Government's wording, but rather was what the dissent saw as a lack of evidence that the Government would have granted the permit if it could not attach the condition.²¹⁰ Again, this fits contrived-

²⁰⁵ 133 S Ct 2586 (2013).

²⁰⁶ Id at 2595–97.

²⁰⁷ See Part I.C (explaining why contrived threats are undesirable whether or not they are carried out); Part I.B (discussing the *Wolf* case, which condemned an improper threat even though it did not induce a modification).

²⁰⁸ *Koontz*, 133 S Ct at 2598.

²⁰⁹ Id at 2609–11 (Kagan dissenting, joined by Ginsburg, Breyer, and Sotomayor).

²¹⁰ The dissent summarized its analysis of this issue by saying that “[t]he District’s only hard-and-fast requirement was that Koontz do something—anything—to satisfy the relevant permitting criteria” and that if the District were not permitted to make the statements that it made, “the District would opt to simply deny permits outright.” Id at 2611 (Kagan dissenting, joined by Ginsburg, Breyer, and Sotomayor). This reasoning indicates that the dissent concluded that the Government was simply providing an uncontrived warning.

threat theory. If a threat is contrived, the fact that it might be phrased in equivocal or pleasant ways does not alter the undesirability of the threat. Third, *Koontz* held that monetary conditions could violate the doctrine, distinguishing them from taxes on the ground that the monetary conditions were imposed on particular pieces of property.²¹¹ Given that taxes are undeniably coercive, the difference is not that taxes are less coercive than such monetary exactions. The difference is that while there is no constitutional autonomy right to be free of general taxes, there is, in the Court's view, a constitutional autonomy right against selecting a particular person's property for uncompensated takings.²¹² Thus, the coercive nature of the threatened monetary exaction affects the constitutional result only when that exaction is selectively imposed on a particular person and thus affects a constitutional autonomy right.

d) *Plea bargaining*. In a plea bargaining agreement, a prosecutor threatens to bring criminal charges unless the defendant agrees to a certain sentence. The prosecutor thus offers a benefit (nonprosecution) in exchange for the defendant's waiver of constitutional rights. Some argue that plea bargains are coerced by the threat of imprisonment.²¹³ But plea bargains are usually enforced. The reason is that, although the threat of prosecution may leave the defendant with no reasonable alternative to agreeing to the plea bargain, duress exists only if the threat is also improper, which the threat of prosecution ordinarily is not.

How can we determine which threats of prosecution are improper? By determining whether the threats are contrived. If the prosecutor is threatening to bring the same charges she would have brought if no plea bargain were possible, then she is simply offering an uncontrived warning. This does not penalize the criminal for exercising his constitutional rights, because he would have faced the same charges even in the no-threat world.

²¹¹ Id at 2598–2602.

²¹² See *Koontz*, 133 S Ct at 2599 (distinguishing *Eastern Enterprises v Apfel*, 524 US 498 (1998), because the monetary obligation there was imposed generally rather than on a particular piece of property). See also *Koontz*, 133 S Ct at 2608 (Kagan dissenting, joined by Ginsburg, Breyer, and Sotomayor) (acknowledging that *Dolan v City of Tigard*, 512 US 374 (1994), and various state doctrines suggest a distinction between general legislative fees and fees that are selectively imposed on individual properties).

²¹³ See, for example, Albert W. Alschuler, *The Changing Plea Bargaining Debate*, 69 Cal L Rev 652, 695–703 (1981); John H. Langbein, *Torture and Plea Bargaining*, 46 U Chi L Rev 3, 12–19 (1978).

In this case, offering a plea bargain can only make the defendant better off, by giving him the option of agreeing to a sentence he finds less harmful than the expected harm from prosecution in the no-threat world.

In contrast, suppose the prosecutor is overcharging—that is, threatening to bring charges she would not have brought in the no-threat world—in order to coerce a plea bargain. Then the prosecutor is making a contrived threat that does penalize the defendant for exercising his constitutional rights, because he would not have faced the same charges in the no-threat world. Accordingly, plea bargains produced by such contrived threats should be unenforceable.

Because there is usually no observable non-plea bargain baseline, it is generally difficult to determine whether prosecutors have overcharged in order to induce plea bargains. However, sometimes there is relevant evidence on that baseline. For example, suppose that a prosecutor initially indicated an intent to bring only certain charges but then, after the defendant refused to plead guilty, added more charges. The Supreme Court has indicated that such a clearly contrived threat likely violates due process.²¹⁴ But when prosecutors indicate their intent to bring certain charges before the plea negotiations begin, the Court has found that there is insufficient evidence to show a contrived threat that violates due process.²¹⁵ Still, if concrete evidence did exist that the prosecutor's initial charges were deliberately excessive in order to coerce plea bargains, that should suffice to show a due process violation.²¹⁶

²¹⁴ See *Bordenkircher v. Hayes*, 434 US 357, 360–61 (1978):

While the prosecutor did not actually obtain the recidivist indictment until after the plea conferences had ended, his intention to do so was clearly expressed at the outset of the plea negotiations. . . . This is not a situation, therefore, where the prosecutor without notice brought an additional and more serious charge after plea negotiations relating only to the original indictment had ended with the defendant's insistence on pleading not guilty. As a practical matter, in short, this case would be no different if the grand jury had indicted Hayes as a recidivist from the outset, and the prosecutor had offered to drop that charge as part of the plea bargain.

(citation omitted).

²¹⁵ See *id.*

²¹⁶ Although Professor Albert W. Alschuler believes that whether a threat is coercive should be based on a normative rather than predictive baseline, see Albert W. Alschuler, *Constraint and Confession*, 74 *Denver U L Rev* 957, 967–69 (1997), he has recently argued that the sentences imposed at trial are systematically greater than those that would be imposed in a system without plea bargaining. See Albert W. Alschuler, *Lafler and Frye: Two Small Band-Aids for a Festering Wound*, 51 *Duquesne L Rev* 673, 698–99,

3. Comparing other proposed tests.

Illustrious scholars have advocated many other tests for dealing with unconstitutional conditions. However, their tests all either underinclusively allow some contrived threats or overinclusively condemn some uncontrived warnings. Further, when their tests produce results that overlap with the conclusions of a contrived-threat test, my analysis contributes a consequentialist proof in favor of those conclusions that has heretofore been lacking.

Kathleen Sullivan argues that courts should focus on the distributive effects of conditioned benefits.²¹⁷ Such distributive effects may well show that the selective denial or grant of benefits violates equal protection or the constitutional rights of third parties. But if so, then either the threatened denial of benefits or the agreements granting benefits are themselves unconstitutional regardless of any coercion. Professor Cass Sunstein makes a similar argument, but he recognizes that it amounts to abandoning the unconstitutional-conditions doctrine.²¹⁸ Their analyses thus cannot explain the puzzle of why conditions are sometimes held unconstitutionally coercive even when the threatened denial of benefits and the agreement to grant benefits are both lawful. Further, when both are lawful, their tests would immunize all conditioned benefits, which would incorrectly allow contrived threats that penalize the exercise of constitutional rights.

Professor Philip Hamburger argues that all conditioned benefits should be unconstitutional whenever the condition waives a constitutional right.²¹⁹ But his test cannot explain the

702 (2013). However, he acknowledges that some scholars and justices have reached a contrary empirical conclusion on the latter point. *Id.* at 687–91, 702. My Article does not purport to resolve this empirical debate, but if the Supreme Court were persuaded that Alschuler's position is empirically correct, then my Article would indicate that plea bargaining should, under the Supreme Court's own precedent, be banned generally, rather than policed with the case-by-case standard that is now used.

²¹⁷ Sullivan, 102 Harv L Rev at 1421 (cited in note 6).

²¹⁸ See Cass R. Sunstein, *Why the Unconstitutional Conditions Doctrine Is an Anachronism (with Particular Reference to Religion, Speech, and Abortion)*, 70 BU L Rev 593, 594–95, 608 (1990). Sunstein recommends abandoning the doctrine for two reasons. First, he rejects what he assumes are the only possible baselines to determine when conditions are a penalty or subsidy: (a) what the government could do, (b) the common law, or (c) a historical status quo baseline. *Id.* at 597–604. He is right to reject those tests, but he does not consider a predictive standard that asks what the government would do if it could not condition the benefit. Second, he notes that penalizing constitutional rights is not always unconstitutional. *Id.* at 603. But when that is so, direct coercion is permissible, and thus a key premise for the doctrine is not met.

²¹⁹ Philip Hamburger, *Unconstitutional Conditions: The Irrelevance of Consent*, 98 Va L Rev 479, 480–81 (2012).

puzzle of why courts often hold precisely the opposite. His test would also incorrectly condemn all uncontrived warnings, which would harm both the government and the persons with the constitutional rights in question.

Professor Richard Epstein argues that current constitutional law deviates from ideal constitutional law by weakly protecting contract and property rights and allowing government redistribution of wealth.²²⁰ To move us closer to his ideal, he argues that conditioning benefits should be (a) unconstitutional when his ideal constitution would not allow denial of the benefit (like preventing businesses from operating), but (b) constitutional when his ideal constitution would not allow providing the benefit (like providing food stamps).²²¹ One obvious problem is that one might disagree with his ideal constitution. But the problem is deeper: his approach would actually move us further from his own ideal. When the threat is to deny one of his ideally required benefits, Epstein and I would both condemn contrived threats but he would also condemn uncontrived warnings, thus resulting in precisely the denials that he fears, as well as preventing government-person bargains that would make them both better off. When the threat is to deny one of his ideally prohibited benefits, he and I would both allow uncontrived warnings but he would also allow contrived threats, thus allowing the government to use its ideally prohibited power to penalize the exercise of constitutional rights that Epstein himself deems ideal.

Professor Mitchell Berman argues that threatening to deny benefits should be presumptively unconstitutional if the purpose of carrying out the threat would be to make it more costly to exercise a constitutional choice.²²² Unfortunately, purpose is a murky test, especially because different legislators have different purposes, and Berman stresses that his purpose test deviates from a pure predictive test like mine.²²³ Nor does he define the grounds for rebutting his presumption. Whenever his test produces conclusions that deviate from a contrived-threat test, his test will either undesirably allow contrived threats or undesirably prohibit uncontrived warnings.

²²⁰ See generally Richard A. Epstein, *Unconstitutional Conditions, State Power, and the Limits of Consent*, 102 Harv L Rev 4 (1988).

²²¹ Id at 27–104.

²²² Mitchell N. Berman, *Coercion without Baselines: Unconstitutional Conditions in Three Dimensions*, 90 Georgetown L J 1, 7, 46 (2001).

²²³ Id at 46.

Professor Kenneth Simons would deviate from a predictive baseline when the government plausibly believes the condition is relevant to the legitimate purposes of the government program.²²⁴ But such deviations would undesirably allow a contrived threat that penalizes the exercise of constitutional rights whenever the condition has some relevance but that relevance is not important enough to mean the benefit would be denied without that condition. As Simons acknowledges, his conclusion conflicts with case law prohibiting the government from conditioning benefits on working during the Sabbath.²²⁵ He argues that, without his related-purpose exception, the government could not limit political campaigning by public employees.²²⁶ But while such limits do seem permissible, the reason is not that they are uncoercive, but rather that the constitutional interest of undistorted politics justifies compelling public employees not to campaign, making coercion irrelevant.

Professor Seth Kreimer proposes a mix of historical, equality, and predictive baselines to judge whether threats to withhold benefits are unconstitutionally coercive.²²⁷ However, whenever his historical or equality baselines allow conditions that a predictive baseline would condemn, his test wrongly allows contrived threats that penalize the exercise of constitutional rights. Whenever his historical or normative baselines condemn conditions that a predictive baseline would allow, his test wrongly condemns uncontrived warnings that would benefit both the government and persons with the constitutional right.

True, Kreimer's historical baseline is relevant to the correct test because the past government conferral of unconditioned benefits helps suggest that the government would have likely continued to confer the benefit if it could not add the condition. However, government policy does change over time, and the historical baseline should be rejected whenever there is evidence that it does not accurately predict what the government would have done without the condition. Kreimer argues otherwise because persons rely on prior benefits.²²⁸ But when the historical test deviates from the predictive test, it will be when the historical test

²²⁴ Kenneth W. Simons, *Offers, Threats, and Unconstitutional Conditions*, 26 San Diego L Rev 289, 291–92, 318–20, 323 (1989).

²²⁵ *Id.* at 323.

²²⁶ See *id.* at 319–20.

²²⁷ Seth F. Kreimer, *Allocational Sanctions: The Problem of Negative Rights in a Positive State*, 132 U Pa L Rev 1293, 1359–78 (1984).

²²⁸ *Id.* at 1362.

either (1) allows undesirable contrived threats to withhold a new benefit, for which reliance is irrelevant, or (2) condemns desirable uncontrived warnings about the future denial of old benefits, in which case his test would produce a denial of old benefits that affirmatively harms reliance interests.

Likewise, Kreimer's equality baseline is relevant to the correct test because government conferral of a benefit on all the persons who accept the condition may suggest a purpose for doing so that would likely extend to everyone if the government could not adopt the condition. However, to accurately judge whether the condition unlawfully coerces the threat recipient, this equality baseline should be rejected when it deviates from the predictive baseline. After all, when there is such a deviation, the equality baseline will either (1) allow contrived threats to withhold benefits that others do not get, which penalizes the constitutional autonomy right, or (2) condemn uncontrived warnings about the coming denial of benefits that others do get, in which case his test would produce a denial of benefits that harms everyone.

III. THREATS THAT INDUCE NEW PRIVATE CONTRACTS

A contract is obviously unenforceable if the terms are independently illegal or if the contract was induced by threats of illegal conduct.²²⁹ But threats of otherwise-lawful action to induce an otherwise-lawful new contract can also constitute duress when the threat is improper and the victim has no reasonable choice.²³⁰ This Part shows that such threats are and should be deemed "improper" only when they are contrived.

A. Threats to Exercise Unrelated Rights

Even if a party has every legal right to terminate a contract, a threat to do so "for some purpose unrelated to the contract, such as to induce the recipient to make an entirely separate contract, is ordinarily improper."²³¹ Such facts indicate that the threatener had no reason not to continue performing the first contract and thus would have performed it if unable to communicate a threat designed to induce the separate contract. Such a threat is thus contrived.

²²⁹ See Restatement (Second) of Contracts §§ 176–78 (1981).

²³⁰ See Restatement (Second) of Contracts §§ 175–76 (1981).

²³¹ Restatement (Second) of Contracts § 176, comment e & illustrations 9, 11 (1981).

The same can be true when a party threatens not to enter one contract in order to induce another. In one case, a lessor said it would rent to the purchaser of a leaving tenant's business, but the lessor then threatened not to do so unless the leaving tenant forked over 45 percent of his business-sale proceeds.²³² This threat was held improper, which thus invalidated the new contract to pay 45 percent of the business-sale proceeds.²³³ This fits the contrived-threat test because the lessor had clearly indicated it was happy to rent to the new tenant and threatened otherwise only to extort part of the business-sale proceeds.

In short, if a party threatens to exercise unrelated rights to induce a new contract, such threats are improper when the threat is contrived—that is, when the party would not exercise the unrelated rights in that way but for the ability to communicate the threat. Such contrived threats impose penalties that the victim would not have incurred in the no-threat world. Because the threat of such penalties can induce the victim to agree to terms on the new contract that are not in his interest, the threat eliminates the normal assurance that the new contract allocates resources to the highest-value user. Thus, such contracts are properly held unenforceable.

If, instead, the threat were to exercise unrelated rights in a way that would be in the interests of the threatener even if no new contract could be induced, then the threat would be an uncontrived warning. Because the unrelated rights would be exercised in an adverse way in the no-threat world, the new contract must make the threat recipient better off because he prefers the new contract terms to having the unrelated rights exercised adversely. Thus, in such cases, the new contracts are properly held enforceable.

B. Threats Not to Enter into the Challenged New Contract

1. Threats to withhold desired goods unless the market price is paid.

The typical threat that induces a new contract is a threat to withhold goods unless the market price is paid. Some argue that all such threats are coercive because individuals have no

²³² See *Wolf*, 154 A2d at 630, citing *Hochman v Zigler's Inc.*, 50 A2d 97, 100 (NJ Chanc 1946).

²³³ *Wolf*, 154 A2d at 630.

reasonable choice but to pay to get the goods that they need to live.²³⁴ However, if the seller could not sell to the individual at the market price, then the seller would withhold the good because it would prefer to sell to someone else for the market price. The threat to withhold the good is thus an uncontrived warning. If duress law made the contract unenforceable, this individual would thus not get the good that he needed. (Of course, if price regulations generally required a lower price, then some individuals would get the good, but recall our premise is that the terms of the contract are not independently illegal.) This conclusion is true even if the seller is a monopolist, because then the market price is the monopoly price, and the seller would prefer to sell to someone else at the monopoly price rather than sell to any particular individual at a submonopoly price. Thus, refusals to sell for less than the market price are enforceable no matter how much the buyer needs the good.²³⁵

*Batsakis v Demotsis*²³⁶ shows that this point holds even in extreme cases. In war-torn Greece in 1942, Demotsis borrowed 500,000 drachmae (worth about \$25) in exchange for a promise to later pay \$2,000 after the war.²³⁷ The court held that this agreement was enforceable.²³⁸ Because the war-torn circumstances doubtless made Demotsis desperate for the money, critics have argued that her promise was extracted under duress.²³⁹ But those same war-torn circumstances also likely made having cash highly valuable and lending money highly risky, and no evidence was offered that the high rate charged differed from the going market price. Thus, if Batsakis did not believe the contract would be enforceable, he likely would not have lent the money to Demotsis, which would have made her even worse off given her apparent desperate need for money. Because the threat to withhold the money unless promised \$2,000 was thus an uncontrived warning, the contract was properly held enforceable. Moreover, assuming that the high market rate reflected the desperate need of persons in war-torn Greece for cash, we want such high rates to be enforceable in order to encourage

²³⁴ See, for example, Sullivan, 102 Harv L Rev at 1446–47 (cited in note 6).

²³⁵ See Restatement (Second) of Contracts § 176, comment f, illustration 14 (1981); *Standard Box Co v Mutual Biscuit Co*, 103 P 938, 943–44 (Cal App 1909).

²³⁶ 226 SW2d 673 (Tex Civ App 1949).

²³⁷ Id at 673–74.

²³⁸ Id at 675.

²³⁹ Fried, *Contract as Promise* at 109–11 (cited in note 99).

others to increase the supply of cash to war-torn Greece to meet that need.

2. Threats to withhold desired goods when parties have been cut off from the market.

Sometimes parties might be cut off from the market, so that the market does not provide an effective constraint. If the buyer has been cut off, a seller might threaten not to sell a product unless the seller receives an above-market price. If the seller has been cut off, the buyer might threaten not to buy unless the buyer receives a below-market price. When should such threats be deemed contrived?

First, consider cases in which a seller has taken steps to cut the buyer off from the market in order to try to charge the buyer an above-market price. Then a threat not to sell at less than an above-market price is clearly contrived because, if that threat could not be made, then the seller would be willing to sell to the buyer at the market price and would not have taken steps to cut off the buyer to begin with. For example, if a seller misleads a buyer into thinking that the seller will supply a good at the market price, and thus causes the buyer to forgo other options until it is too late, then a threat not to sell unless an above-market price is paid is contrived and a contract at that above-market price is unenforceable.²⁴⁰

Now suppose that a seller, through no productive effort, stumbles upon a buyer who has been cut off from the market for adventitious reasons. For example, suppose that Bill Gates shows up at the emergency room needing immediate lifesaving treatment, and the doctor says she will not treat him unless he promises to pay her \$1 billion. In this case, the doctor's threat is contrived because if that threat were not possible, the doctor would be willing to treat Bill Gates at the doctor's usual market rate. Thus, Bill Gates's agreement to pay \$1 billion is unenforceable.²⁴¹

If, in the adventitious case, there is no clear market rate, an alternative price baseline might still show that the seller made a contrived threat. This point is illustrated by *Post v Jones*.²⁴² In

²⁴⁰ See Restatement (Second) of Contracts § 176(2)(b) & comment f, illustration 13 (1981).

²⁴¹ See Steven Shavell, *Contractual Holdup and Legal Intervention*, 36 J Legal Stud 325, 338 (2007) (noting that courts hold adventitious holdup contracts to be unenforceable when they are substantially above market prices).

²⁴² 60 US (19 How) 150 (1856).

1849, three whaling ships happened to find another whaling ship stranded on rocks in the Arctic right before winter and agreed to buy the stranded ship's whale oil for a very low price.²⁴³ Although the whale oil was nominally "auctioned," the stranded ship had more oil than the three ships could take, so they had no need to compete with each other on price.²⁴⁴ The Supreme Court recognized that even when salvage is an "absolute necessity," parties can set salvage prices by contract when the stranded ship has access to "a market and competition."²⁴⁵ However, the Court held that this contract was unenforceable for two reasons. First, the stranded ship had no reasonable choice but to take whatever the three ships offered, because "there was no market . . . [and] no competition."²⁴⁶ Second, taking the whale oil at the default price (which is set by salvage law when no binding contract exists) was still highly profitable because salvaging the oil required little risk or effort and the opportunity costs of filling the cargo space were low given that the whaling season was ending.²⁴⁷

The second finding in *Post* meant that the three ships' threat not to take the oil at the default rate was a contrived threat because, if they could not make that threat, they would have happily salvaged the oil at the still highly profitable default rate. No salvaging was the wrong but-for baseline because salvaging would have occurred in the no-threat world. Indeed, the Court explicitly rejected the three ships' claim that the stranded ship was not harmed because it was better off than if no salvaging had been provided.²⁴⁸ The effects of allowing salvagers to price below the default salvage price in such accidental-monopoly cases is harmful because if ships know that rescues will cost them almost all the value of their cargo, then they will take excessive precautions to avoid the need for rescue.²⁴⁹ In contrast, because the default price was high enough to make salvage profitable, and because finding the stranded ship was adventitious, using the default price as the baseline for judging whether the threat was contrived will not reduce any salvage effort. But this result

²⁴³ Id at 158–59.

²⁴⁴ Id at 159.

²⁴⁵ Id at 157–58. See also *The Elfrida*, 172 US 186, 196–97, 204 (1898) (stressing that the law does not generally regulate the reasonableness of salvage prices and allowing a high contract price when the stranded ship had competitive salvage bids).

²⁴⁶ *Post*, 60 US (19 How) at 159.

²⁴⁷ Id at 160–61.

²⁴⁸ Id at 160.

²⁴⁹ See Steven Shavell, *Foundations of Economic Analysis of Law* 336 (Belknap 2004).

depends critically on having some known, desirable price baseline to determine whether a contrived threat had been made.

Now suppose a seller earns being in a position to sell to a buyer who is cut off from the market. For example, suppose that the rescue ships were actually in the business of searching the seas for stranded ships. If the default price provided by salvage law suffices to compensate those search costs, then a threat to withhold salvage unless paid more than that default price would be contrived because, absent an ability to make that threat, the rescue ships would have incurred the search costs and provided salvage. If the default price provided by salvage law does not suffice to compensate for those search costs, then a threat to withhold salvage unless paid more than that default price would be an uncontrived warning because, absent an ability to make that threat, the rescue ships would not have incurred the search costs and thus would not have been there to provide salvage. This result could also be achieved by adjusting the default price to compensate for efficient search costs.²⁵⁰

3. Bilateral negotiation to split extramarket surplus.

Suppose that neither party has been cut off from the market but the seller's goods are worth more to the buyer than to any other buyer. This is a frequent situation, often because of relationship-specific investments. An employee who works at a firm may develop firm-specific skills that make the employee more valuable to the firm than to any other firm. Or a business that rents a location may find at the time of renewal that the location is more valuable to it than any other location is. In these cases, we have a bilateral negotiation over how to split an extramarket surplus because, although the buyer cannot find any other seller whose goods it values as much, the seller also cannot find any other buyer who values its goods as much. There will be a bargaining range, with the minimum price the seller would take being the market price other buyers would pay, and the maximum price the buyer would pay being the higher value it puts on the goods. The results of such bargaining are indeterminate.

In these cases of bilateral negotiations about the split of extramarket surplus, the contrived-threat test is not helpful, because we have no baseline price to assess whether a contrived threat has been made. One might think that a contract at any

²⁵⁰ See *id.* at 335–36.

price above the seller's reservation price must reflect a contrived threat because the seller would take the reservation price if it were unable to insist on a higher price. However, it is also the case that any price below the buyer's reservation price must reflect a contrived threat by the buyer because the buyer would take its reservation price if it were unable to insist on a lower price. Such negotiations thus cannot help but reflect contrived threats by *both* sides, and a contrived-threat test thus cannot give either side a unique right to void the contract. Moreover, both sides would be worse off if contracts between them were unenforceable, because any contract between them leaves them better off than a contract with others. Thus, contracts produced by such bargaining are enforceable.²⁵¹

C. Blackmail

In the case of blackmail, the threat is to divulge information to third parties unless the victim agrees to pay money. Blackmail law had long been regarded as "one of the most elusive intellectual puzzles in all of law."²⁵² Contrived-threat analysis helps solve that puzzle by showing that all blackmail contracts are necessarily produced by either contrived threats or uncontrived warnings and that each has distinctive harmful effects. Contrived blackmail threats are coercive and cause inefficiency between the contracting parties, without affecting the information revealed to third parties. Uncontrived blackmail threats are uncoercive and benefit the contracting parties, but they prevent information disclosures to third parties whose interests often trump the interests of the contracting parties. To explain blackmail law, we need to understand the combination of these effects.

²⁵¹ Restatement (Second) of Contracts § 176, comment a (1981). One might think that any price above the seller's reservation price is undesirable because it discourages the buyer (for example, a tenant) from making relationship-specific investments that make the seller's goods more valuable to the buyer. However, the seller might also make relationship-specific investments that make its goods more valuable to the buyer. Moreover, if buyer investments of this sort are efficient, the seller can help pay for them. For example, in commercial leases, sellers often give tenants improvement allowances to induce tenants to begin leases.

²⁵² James Lindgren, *Blackmail: An Afterword*, 141 U Pa L Rev 1975, 1975 (1993).

1. Contrived blackmail threats are always coercive and invalid.

First, consider cases in which divulging the relevant information “would not significantly benefit the party making the threat.”²⁵³ Given that the act of revealing this information would not significantly benefit the threatener, it is likely a contrived threat. True, if divulging information is very cheap, one might wonder whether the threatener would divulge the information in the no-threat world for even insignificant benefits. But effort is necessary to collect information and (even when acquired fortuitously) to document the information in a way that can be divulged such that it is sufficiently persuasive to be effective and to overcome the risk of suits for libel or slander. A blackmailer would not likely incur the costs of such effort for insignificant benefits without the prospect of making blackmail threats.

Thus, any threat to divulge information is likely to be a contrived threat if divulging the information confers no significant benefit to the threatener. Allowing such contrived blackmail threats would harm the victim because in the no-threat world he would not pay money and the information would still not be divulged. Because such contrived threats do not affect the revelation of information, preventing such blackmail would not affect third parties. Finally, preventing such contrived blackmail threats would reduce blackmailer efforts that would inefficiently incur costs to induce nothing but an unproductive transfer of wealth. Because these effects are always negative, contract law correctly condemns all blackmail contracts that are induced by such contrived threats, regardless of what sort of information is at issue.²⁵⁴

2. Uncontrived blackmail threats benefit contracting parties but harm third-party interests.

Now consider cases in which divulging information would significantly benefit the threatener. For example, suppose the information were collected by an investigative reporter who would, in the no-threat world, publish the information in an article. In

²⁵³ Restatement (Second) of Contracts § 176(2)(a) (1981). See also Restatement (Second) of Contracts § 176, comment f (1981) (“Clause (a) is concerned with cases in which a party threatens to do an act that would not significantly benefit him but would harm the other party. . . . A typical example is a threat to make public embarrassing information concerning the recipient unless he makes a proposed contract.”).

²⁵⁴ See Restatement (Second) of Contracts § 176(2)(a) (1981).

that case, the threat to divulge the information is clearly an uncontrived warning.²⁵⁵ Accordingly, an agreement to suppress the information is in the interests of both contracting parties, because the victim prefers paying the money to having the information divulged and the blackmailer prefers getting the money to revealing the information. But this blackmail contract does prevent a disclosure of information to third parties that otherwise would have occurred. Whether to invalidate blackmail contracts induced by uncontrived warnings thus turns on whether third-party interests in the information trump the interests of the contracting parties, which varies with the information at issue.

In particular, when the information concerns crimes, the public obviously has a powerful interest in knowing the information because that helps deter crimes that harm the public. Because the public interest in discovering crimes exceeds the interests of the contracting parties in concealing crimes, such blackmail contracts are invalid even when uncoercive. Accordingly, contract law correctly condemns all blackmail about crimes, whether or not divulging the information would confer a significant benefit that would make the threat uncontrived.²⁵⁶ The reason is that all such blackmail about crimes necessarily involves either (1) a contrived threat that creates inefficiency for the contracting parties without affecting information disclosure or (2) an uncontrived warning that suppresses the disclosure of criminal information to third parties whose interests trump the interests of the contracting parties.

When the information involves embarrassing noncriminal information, third-party interests are weaker, and it is less clear whether they outweigh the interests of the contracting parties. In the investigative-reporter example, one might worry about the harm to the reporter's newspaper or its readership from suppressing the information. On the other hand, one might regard third parties' interests in salacious information as weak. Such judgments vary by jurisdiction and court, as well as with the specific information at issue. General contract law thus does not condemn agreements that are induced by threats to divulge embarrassing noncriminal information when divulging would

²⁵⁵ Some believe that the costs of collecting and revealing information are sometimes so trivial that information would be revealed even without any significant benefit. To the extent they are right, such threats are also uncontrived warnings and have the effects described in this Section.

²⁵⁶ See Restatement (Second) of Contracts § 176(1)(b), (2)(a) & comments c, f (1981).

significantly benefit the threatener (in other words, when it is an uncontrived warning),²⁵⁷ leaving it to other bodies of law to decide when such uncoercive contracts should be void on public policy grounds.

Why, however, do some criminal blackmail statutes seem to overinclusively condemn even uncontrived warnings about information with little legitimate third-party interest? Perhaps partly because adjudicative errors in distinguishing between contrived and uncontrived blackmail threats would create strong underdeterrence concerns if such an exception were recognized. Nor are blackmail statutes literally applied in certain cases. Consider the fact that litigation usually ends in a settlement in which someone receives money and agrees not to disclose information about the other party. Because such confidentiality has value, it must have been paid for in the settlement, so this exchange literally fits broad definitions of blackmail. But no one treats it that way. The analysis here explains why. Litigation already provides copious incentives to collect and document the information, so disclosure would likely be nearly costless. Further, litigants generally dislike each other, so disclosure would likely produce serious psychic benefits. Thus, threats to divulge the information collected in litigation are particularly likely to be uncontrived. Adjudicative errors and underdeterrence are thus less of a concern, and an implicit exception to blackmail law is recognized. The analysis here thus helps explain not only why blackmail prohibitions are so broad, but also why nonliteral exceptions are sometimes recognized.

3. Prior literature.

Prior works on blackmail have tended to find it perplexing because they focus either on the legality of the threatened act rather than on the practical effects of the threats²⁵⁸ or on some of those practical effects to the exclusion of others. Some articles focus on protecting the interests of third parties in information,²⁵⁹ but they cannot explain the universal ban on contrived blackmail threats, which do not affect third-party information. Other works focus on the harm to victims and the inefficiency of

²⁵⁷ See Restatement (Second) of Contracts § 176(1)(b) & comment c (1981).

²⁵⁸ See, for example, Jeffrie G. Murphy, *Blackmail: A Preliminary Inquiry*, 63 *Monist* 156, 156–58 (1980); Glanville L. Williams, *Blackmail*, 1954 *Crim L Rev* 79, 162–63.

²⁵⁹ See, for example, James Lindgren, *Unraveling the Paradox of Blackmail*, 84 *Colum L Rev* 670, 672 (1984); Richard A. Epstein, *Blackmail, Inc.*, 50 *U Chi L Rev* 553, 558 (1983).

efforts to collect information with no productive gain,²⁶⁰ but they cannot explain the ban on many uncontrived blackmail warnings, through which the victim benefits, the information would be collected anyway, and the blackmail contract does alter the information revealed to third parties. Professor Scott Altman has offered what he calls a “patchwork” theory that combines many of these effects.²⁶¹ What contrived-threat analysis clarifies is the deeper structure: that all blackmail must have one or the other of these sets of effects and that only the combination can explain the full scope and limits of blackmail law.

CONCLUSION

Contractual duress, unconstitutional conditions, and blackmail have been long-standing puzzles because the focus has been on the wrong baseline: what the threatener should do or legally could do. To resolve these puzzles, one must instead focus on what the threatener *would* do absent the threat. Only contrived threats to take action that differs from this but-for action can harm the threat recipient and be coercive. Uncontrived warnings can only benefit the agreeing parties and thus should be unlawful only when even an uncoerced agreement would be unlawful, usually because the agreement harms third parties. A contrived-threat test not only explains the actual pattern of case law in a clearer fashion, but also provides a sound normative justification for those results.

²⁶⁰ See, for example, Douglas H. Ginsburg and Paul Shechtman, *Blackmail: An Economic Analysis of the Law*, 141 U Pa L Rev 1849, 1857–60 (1993); Leo Katz, *Blackmail and Other Forms of Arm-Twisting*, 141 U Pa L Rev 1567, 1598–99, 1615 (1993); Wendy J. Gordon, *Truth and Consequences: The Force of Blackmail's Central Case*, 141 U Pa L Rev 1741, 1748–50 (1993); Ronald H. Coase, *Blackmail*, 74 Va L Rev 655, 673–74 (1988); Robert Nozick, *Anarchy, State, and Utopia* 84–87 (Basic Books 1974).

²⁶¹ See generally Scott Altman, *A Patchwork Theory of Blackmail*, 141 U Pa L Rev 1639 (1993).