

ISSN 1936-5349 (print)
ISSN 1936-5357 (online)

HARVARD

JOHN M. OLIN CENTER FOR LAW, ECONOMICS, AND BUSINESS

PRICE-FIXING POLICY

Louis Kaplow

Forthcoming in *International Journal of Industrial Economics* (2018)

Discussion Paper No. 945

01/2018

Harvard Law School
Cambridge, MA 02138

This paper can be downloaded without charge from:

The Harvard John M. Olin Discussion Paper Series:
http://www.law.harvard.edu/programs/olin_center

The Social Science Research Network Electronic Paper Collection:
<https://ssrn.com/abstract=3101088>

PRICE-FIXING POLICY

Louis Kaplow*

Abstract

The prohibition against price fixing is competition law's most important and least controversial provision. Yet there is far less consensus than meets the eye on what constitutes price fixing, and prevalent understandings cannot be reconciled with principles of oligopoly theory. This article (1) presents a fundamental reconceptualization of our understanding of horizontal agreements, (2) develops a systematic analysis of price-fixing policy that focuses on its deterrence benefits and chilling costs, and (3) compares this direct approach to commentators' favored formulations that typically involve some sort of formalistic communications-based prohibition. By targeting a subset of means rather than the illicit ends, conventional formulations tend to impose liability in cases with lower deterrence benefits and greater chilling costs than those reached under a direct approach and to incur greater administrative costs as well.

JEL Classifications: D43, K21, L13, L41

Keywords: price fixing, competition policy, antitrust, conspiracy, horizontal agreement, collusion, oligopoly

Forthcoming in the International Journal of Industrial Economics, 2018

© Louis Kaplow. All rights reserved.

*Harvard University and National Bureau of Economic Research. I am grateful to Joseph Harrington, the editor, and referees for comments; Harvard's John M. Olin Center for Law, Economics, and Business for financial support; and David Geiger for research assistance. Disclaimer: I occasionally consult on antitrust cases, and my spouse is in the legal department of a financial services firm.

Harvard University, Hauser Hall 322, 1575 Massachusetts Ave., Cambridge, MA 02138, USA. Tel.: +1 617 495 4101.

1. Introduction

Competition laws throughout the world condemn price fixing and penalize it harshly. The prohibition's uncontroversial character masks foundational questions about what constitutes price fixing and how competition regimes should best determine its presence. This article, drawing on Kaplow (2011a, 2011b, 2011c, 2013a), presents a systematic, ground-up analysis of the problem.¹ The central features of this direct approach—defining the social problem, figuring out how best to detect its presence, and determining appropriate sanctions—are as basic as they are different from what is emphasized in most prior policy analysis of the scope of price-fixing provisions.

Section 2 begins by offering a reassessment of conventional attempts to define horizontal agreements so as to distinguish classic cartels (involving, say, secret pricing discussions in hotel rooms) from the broader category of coordinated oligopolistic price elevation. Most approaches entail some type of communications-based proscription that turns out to be difficult to formulate (not just in its details, but at its heart) and is necessarily rooted in formalism rather than any substantive notion grounded in oligopoly theory.² It is not a purpose of this section to argue—as some have implicitly done in the past—that the proper definition of price fixing should somehow dictate policy; rather, the aim is to motivate a fresh analysis of the policy problem that starts from first principles.

Section 3 outlines a direct approach to price-fixing policy. The social problem with successful coordinated price elevation has both static and dynamic elements, the latter underappreciated. Another key point is that attempts to deter this harmful activity risk the imposition of chilling costs, requiring explicit attention to the appropriate tradeoffs. Detection poses the greatest challenge because firms seek to hide illegal activity, and, thus concealed, it can be difficult to distinguish from competitive behavior that we do not wish to distort. Conditions being conducive to successful coordinated price elevation are an important prerequisite to liability because conduciveness is necessary for significant harm to arise whereas unconducive conditions both negate this possibility and further suggest that the chilling costs of mistakes are likely to be larger. Sanctions should be well calibrated to harm not only because strong sanctions are needed to deter significant price elevations and but also because low sanctions tend to be optimal for small overcharges due to chilling concerns.

Section 4 explicitly compares this direct approach to the commonly advanced communications-based prohibition. The approaches differ when liability must be based on circumstantial evidence (with smoking-gun evidence, there is liability regardless). The direct approach tends to assign liability in cases posing the greatest danger and the lowest chilling costs whereas the conventional one ends up targeting settings that involve both lower harm and greater risks of chilling desirable behavior. In addition to being a dominated strategy regarding outcomes, the conventional method is less administrable: it demands an additional, difficult inquiry: determination not only of whether coordinated price elevation occurred but also by what (unobserved) means it was accomplished. On reflection, it should not be surprising that poorer

¹To streamline the presentation here, I offer an abbreviated treatment of many complex subjects, skip over others, and largely omit these writings' voluminous references to relevant literatures on industrial organization economics, competition policy, and law. I also do not address vertical agreements, the subject of Kaplow (2016).

²This article's contrast of formalism to economic analysis is familiar in competition law, such as in the ongoing struggle over the proper meaning and application of TFEU 102. A difference is that, in the present setting, economists have acquiesced in or even embraced formalism rather than criticizing it, perhaps in part because the degree of formalism embedded in conventional understandings is insufficiently appreciated.

targeting and greater cost are the consequences of a legal prohibition that requires subtle inferences about firms' covert means rather than confining attention to a direct assessment of how the evidence bears on the ends they achieved, which need to be examined in any event.

Section 5 offers some concluding remarks, including important reservations regarding policy implications in light of limits to existing empirical knowledge and institutional features of different competition law regimes.

2. Horizontal agreements

This section addresses conventional notions of horizontal agreement that most commentators believe should and do define the reach of competition law's prohibition of price fixing. Even though the analysis in this section points to interdependence as the most cogent notion of horizontal agreement, this convergence does not imply that interdependence should therefore be taken as the beacon that dictates what counts as price fixing. A central theme of this article is that optimal competition policy should not be extracted from definitions and resolutions of interpretive debates but instead should be derived functionally—an enterprise undertaken in section 3 and compared with the conventional approach in section 4.

2.1. Illustrations

In my prior writing, I develop two examples that dramatize the severity of the problem with defining agreement in ways that excuse pure interdependence. In the first, I present a dialogue between two competitors in which they each negotiate prices explicitly and offer assurances of no secret price cuts, resulting in prices that move from a competitive level to a joint maximizing level and stay there indefinitely. This scenario depicts a clear, express, illegal agreement by standard accounts.

I then point out that one can reinterpret this dialogue as price postings by two adjacent gas stations. Consider that the information content of one proprietor stating “I will raise my price to \$3.29 and I promise not to engage in any price cuts without immediately informing you” is entirely conveyed by the owner posting \$3.29 on the station's large price sign. And the other owner's statement that “I will match your price of \$3.29 and I promise not to engage in any price cuts without immediately informing you” is likewise fully communicated by promptly posting \$3.29 on that station's large price sign. (More complex “negotiations” can occur in rounds, such as when the first owner sharply raises its price by posting \$3.29, the second responds with \$3.09, the first counters with \$3.19, and then the second likewise posts \$3.19, with both subsequently sticking to that price.) Most regard this latter depiction as a clear instance in which there is no agreement.

In my second example, I begin with a classic cartel discussion in a hotel room that comes to closure on a price increase, after which the firms implement it. Again, this constitutes a clear, express, illegal agreement. Then, I gradually tweak this story into seriatim press conferences that most would deem an obvious case of no agreement. This metamorphosis is implemented in stages. First, one can imagine that the language of the conversation in the hotel room is carefully “lawyered” by substituting transparent euphemism and innuendo for words like “agree” and “promise.” Next, one can adjust the setting by substituting video conferencing and increasing

the time lag between statements. Finally, one might invite the media to operate the cameras and broadcast the discussions.

In the variations within each example, the practical outcome (prices and quantities) and social harm are identical. Moreover, essentially the same process of interchange takes place: Content-wise, the same communications occur (objectively viewed), and the same understandings are generated (subjectively viewed). Indeed, they are identical by stipulation.³

These illustrations are highly consequential. Within each example, the opposed cases are not close shades of gray by conventional accounts; they are polar opposites that ought to be easily distinguishable. But, upon examination, the differences are trivial, and there are none whatsoever with regard to the content communicated. Perhaps one could define agreement to require magic words, so a little wordsmithing creates immunity. Or stipulate that clearing a hotel room of smoke and inviting the media gives competitors *carte blanche*. But such rules would largely nullify the price-fixing proscription. Sufficient reflection on these variants of familiar settings indicates that conventional definitions of agreement have one of three radical implications: classic cartel arrangements (secret negotiations in hotel rooms) are legal; pure interdependence is illegal after all; or trivial distinctions (use of magic words, changes in who operates video cameras) determine what is illegal.

2.2. Terminology

Ambiguous, inconsistent, and even contradictory usage of key terms has generated great confusion with regard to the notion of horizontal agreement. Following typical but not uniform practice, I use *independent* action to denote competitive behavior wherein each firm takes other firms' prices as given (think one-shot Nash equilibrium). For example, grocery stores all raising their retail prices of oranges in response to an equivalent increase in wholesale prices is understood to be independent for present purposes. *Interdependent* behavior is that which is based on anticipation of other firms' reactions—notably, when firms charging a supracompetitive price do not cut their prices to steal market share from rivals because of anticipated responses (think successful coordination in supergames).⁴

Terms like agreement, conspiracy, and concerted action (language associated, for example, with Sherman Act Section 1 and TFEU Article 101) are often regarded to denote some subset of interdependent behavior, leaving a nontrivial remainder outside the terms' scope. A central difficulty is that standard, pertinent definitions of each of these terms refer to a meeting

³In practice, particularly when the interchanges take a simple form (perhaps the first firm states a single, higher price and the others match), we may be less sure that the public version constitutes coordinated oligopoly price elevation. Moreover, different versions may be more or less effective: for example, public press conferences may be more ambiguous because each firm is aware that others are addressing multiple audiences, or public statements may be more effective because they are more credible. Such possibilities reinforce this article's approach under which communications are viewed as evidence, to be interpreted in context. If what is prohibited is instead the act of communication itself, as many suggest, then the examples (by construction) are indistinguishable. Moreover, differences in effectiveness (in either direction) do not alter whether or not what has transpired constitutes "communication," a point elaborated in subsection 2.3.

⁴For ease of exposition, this presentation is simplified, focusing on the core situations that most have in mind. For example, "competitive" behavior need not be perfectly competitive behavior, repeated behavior can take many forms, and coordination need not be limited to pricing. Furthermore, additional terms that have generated some confusion are not defined here. For greater elaboration, see Kaplow (2011c, 2013a). Finally, note that variants of the word "coordinate" are not precisely defined here; the intended meaning is the same as the familiar usage in merger guidelines that distinguishes unilateral from coordinated effects.

of the minds, harmony of opinion or action, or mutual understanding—the very essence of interdependence, which is exactly what commentators mean to distinguish when using these very words. Of course, the law sometimes uses language in special ways, so it is possible to define agreement and the like in some other fashion (although longstanding conspiracy law does not in fact do so). At a minimum, it is confusing to follow the frequent practice—by courts, agencies, and commentators alike—of using such terms without either precise canonical refinement or a series of crisp examples that articulate the intended boundary—if, indeed, some notably narrower subset of interdependent behavior is meant to be specified.⁵ This unmet semantic challenge reflects the conundrum posed by the examples in the preceding subsection.

2.3. *Communications*

Firms' use of communications seems to be at the crux of what many commentators take to be the legal notion of agreement despite the fact that the foregoing definitions make no reference to communications or, more broadly, to the means that firms employ to achieve interdependence. Here and throughout, I criticize deeming the sorts of communications used by firms as *constitutive* of agreement (which is necessary and sufficient for illegality with regard to ordinary price fixing) rather than as a (sometimes very important) source of *evidence* bearing on liability, discussed in subsection 3.2.3 and elsewhere, or as a *facilitating practice* that might separately be condemned, discussed at the end of this subsection.

Moving to the substance, the notion of communication is exceedingly broad, encompassing all manner of conveying information, whether by signs, symbols, gestures, other behavior, or complex language systems. Even the clearest examples that commentators seek to exclude from their definitions of agreement centrally involve communications, such as gas stations' posting of price signs or price announcements at sequential press conferences. Indeed, if communications constitute agreement, vastly more behavior than interdependence is reached, for even perfect competitors' routine price quotations communicate.

Therefore, a central feature of any communications-based definition of what counts as illegal must be a limitation to some particular sorts of communications, demarcated perhaps by content, mode, or context. Because proposed specifications vary widely, are ambiguous, and sometimes are omitted altogether, I find it useful to refer abstractly to the collection of included communications or other acts—the use of which is deemed to constitute agreement—by the set X , and all other communications or acts—deemed insufficient—by the set X' . But which communications are in this set X ? And by what principle is this to be determined?

An appealing answer would be to undertake direct policy analysis: that is, to include in set X those actions for which prohibition would on-balance be desirable with regard to the benefits of deterring coordinated oligopolistic price elevation and the costs of chilling desirable behavior. This formulation, however, is just an awkward manner of stating what I call the direct

⁵Related problems arise with the familiar invocation of “plus factors.” At a foundational level, there is neither a clear identification of the baseline to which these factors must add (perhaps it is merely parallel conduct, which does not distinguish competitive behavior) nor a sharp articulation of the goal line that must be crossed. The most mentioned plus factors are acts against self-interest and a motive to conspire (both of which define interdependence, indicating that it alone is sufficient after all) and another is noncompetitive performance (such as elevated prices, which again can be produced by pure interdependence). The other frequently mentioned plus factor is smoking-gun evidence or close to it, but such is direct evidence and therefore does not address the need to make inferences from circumstantial evidence, which is the challenge that plus factors are supposed to address.

approach to price-fixing policy, examined further in sections 3 and 4. But if this is eschewed, what is the alternative? And why would it be appealing?

To elaborate these questions, consider the problem of functional equivalents. As we divide the sea of possible sorts of communications into sets X and X' , there will in many settings be elements in the prohibited set X that have functional equivalents (or close substitutes) in the permitted set X' . If we freely allow this state of affairs, circumvention is invited. But if we include all functional equivalents in X , we have really abandoned the approach. Why? Because the function that firms mean to serve is the successful achievement of interdependent price elevation, it follows that a classification system that finds illegality whenever that function is served is tantamount to a prohibition of successful interdependence, precisely the rule that these commentators purport to reject. As a consequence, any substantially distinguishable (narrower) prohibited set X must inevitably be determined to a significant extent on formalistic rather than functional grounds. In addition, as we will see in section 4, this method of defining the legal prohibition poses immense problems of proof when the means employed (were some in set X or were all in set X' ?) must be inferred from circumstantial evidence.

Consider further how communications might be divided. Motivated by the example with two gas stations—where discussions are deemed illegal whereas price postings conveying identical information and generating equivalent understandings are permissible—it may seem that the implicit distinction is that the use of language is prohibited but signs and symbols are allowed. But such a distinction misunderstands the concept of language, which is defined as the use of signs and symbols (of any sort) to communicate. This point is driven home by the existence of rich sign languages. Should their use lead to imprisonment (because they are languages) or exoneration (because they are mere signs and symbols)? Presumably the former, but once symbolic expression is deemed illegal, what isn't?

Stepping back, we can see that the direct regulation of what language may be used, which communications-based approaches attempt to do, is a fraught enterprise. To better understand the predicament, my work (Kaplow 2011c, 2013a) examines scholarship in linguistics, philosophy, social and natural sciences, and sign language studies. Language use, like communication, is fundamentally an intersubjective phenomenon, undergirding the difficulty of using the concept to distinguish cases in which the same views are conveyed and the same understandings are generated, and rendering mysterious the conventional orientation toward the topic. Human language is highly flexible, constantly evolving and adapting. It often expresses key messages by drawing on the capacity to read between the lines. It can also be difficult for outsiders to decode (a familiar example being substitute terminology employed by drug dealers who fear wiretapping). These reasons underlie the futility of a prohibition limited to magic words or requiring the explicit presence of specific traits in parties' communications.

An additional reason to be skeptical of targeting language while excusing behavior that achieves interdependence is that such an approach is obtuse. The implicit premise is that words speak louder than actions, but the opposite maxim is better rooted in common sense and, not surprisingly, in the teaching of scholars of strategy, including business strategy with regard to the interaction of firms in an oligopoly. See, for example, Schelling (1960, pp. 107, 117), Porter (1980, p. 103). Is a gas station owner's private assertion of a pricing intention more credible than its actual posting of a higher price?

Before leaving the subject, consider a qualitatively different manner in which certain forms of communication can be relevant to competition policy: as facilitating practices that

might be proscribed as such. To elaborate, consider three qualitatively different ways in which communications in the guise of facilitating practices may be relevant to liability.

First, such communications might be constitutive of agreement and hence define illegality, the subject until now. (In trade associations, information exchanges are often explicit and are certainly legally sufficient with regard to the agreement question, leaving open the question whether the sorts of information exchanged give rise to illegality.) Second, facilitating communications may constitute a type of evidence bearing on illegality, which it would be under section 3's direct approach. (Regular meetings in hotel rooms, even with no proof of what was said, may be taken to support an inference of illegality when combined with other evidence of coordinated oligopolistic price elevation.) Third, certain practices might also or instead be deemed illegal precisely because of their tendency to facilitate coordinated price elevation. (The very act of regularly meeting with competitors in hotel rooms could be deemed illegal.) Of course, most communications would not sensibly be proscribed, but some might be, and without regard to whether price elevation can be demonstrated.

2.4. Oligopoly theory

For the horizontal agreement requirement to be grounded in economic substance, it has to embody an important feature of oligopoly theory. To elaborate, under competition law, something called agreement is ordinarily taken to be a necessary condition for liability for behavior by a group of firms, and, for price fixing, it is a sufficient condition. For this to be good policy, the proffered notion of agreement must correspond to something real and relevant.

Consideration of economic substance poses a powerful *prima facie* objection to defining horizontal agreement narrower than interdependence. All appreciate that the underlying behavior and its impact on social welfare are the same regardless of the means by which oligopolistic firms manage to coordinate their actions to elevate price by a given amount. (Examination of various means can serve as evidence of what transpired, in the manner discussed in section 3, but this is qualitatively different from using all manner of evidence for the purpose of inferring which means were employed, where the choice of means itself determines whether the behavior is prohibited.)

We can also examine oligopoly theory for the light it sheds on coordination itself. The relevant theory pertains to noncooperative repeated games: settings in which externally binding agreements are unavailable (matching the prohibition on legal enforcement of cartel contracts).⁶ Coordinated price elevation must be self-enforcing to succeed, a familiar point that is as applicable to classic cartels as it is to more informal arrangements. See, for example, Kühn and Vives (1995, 43). Repeated (rather than one-shot) games are the focus because they encompass strategic interaction in which firms may anticipate rivals' reactions and respond to others' behavior. An equilibrium in such a game (when it differs from the equilibrium of the one-shot game) is the aspect of this theory that most closely corresponds to the notion of agreement in the price-fixing setting. Such an equilibrium is often characterized as involving a meeting of the minds, which is a standard definition of agreement, conspiracy, and the like—and the essence of interdependence. In contrast, it is hard to identify any fundamental division in oligopoly theory that corresponds to what many imagine to exist when attempting to define agreement more

⁶In this respect, all communications, including explicit cartel agreements, are cheap talk: they may facilitate coordination (equilibrium selection) but not stability.

narrowly. That is, the distinction that many would have the law make central is essentially nonexistent in economic theory.

3. Direct approach to price-fixing policy

A direct approach to the determination of optimal policy toward price-fixing is, in its general outline, as straightforward as it is different from most prior policy writing on the scope of the prohibition. First is elaboration of the social problem, which involves exploring not only what is objectionable but also why this is so—because the magnitude of harm may differ across settings—and what sorts of risks might arise with regard to the chilling of benign behavior as a consequence of mistaken identification. Second is determination of when the problem is present. This topic requires the most attention due to the difficulty of the task (in all but cases with smoking-gun evidence) of distinguishing successful coordinated price elevation from ordinary competitive behavior, the many potential avenues of inference and their interrelationships, and the just-mentioned concern for errors of over- as well as underinclusion. Third is an assessment of sanctions, which has an important connection with concerns about detection.

3.1. Social problem

3.1.1. Welfare effects

The core, familiar problem with coordinated oligopolistic price elevation is that prices are raised above the competitive level. This price elevation reduces both consumer and total welfare (and output reductions may not be efficiently allocated across firms). An obvious point to keep in mind, which proves to be important below, is that larger price increases are more harmful than smaller ones. This is particularly so with regard to deadweight loss, which rises disproportionately with price.

Dynamic effects are predominantly adverse as well, reinforcing the conclusion that coordinated oligopolistic price elevation is socially detrimental. Elevated prices induce additional entry, which has two types of consequences (Mankiw and Whinston 1986). First, entry tends to be excessive due to the business-stealing effect, which is the only relevant factor in homogeneous goods industries that are the setting in which coordinated price elevation tends to be most feasible and seems most often to have occurred. This dynamic consideration reinforces the need to deter such behavior. Second, when products are differentiated, additional entry tends to enhance product variety, a social benefit that entrants do not fully capture because they typically cannot extract inframarginal consumers' surplus. This countervailing gain is largest when differentiation is greatest, but that is when coordinated price elevation is unlikely to occur, so this potentially countervailing force tends to be unimportant here.⁷

Elevated prices tend to reward not only entry but also any activities that put firms in a position to charge above-cost prices or capture greater market share when doing so. It is well understood that the allocative inefficiency of supracompetitive pricing may have favorable dynamic consequences for investment, such as with regard to intellectual property. Accordingly, competition regimes like that in the United States tolerate monopoly pricing when the dominant

⁷This point accentuates the paradox-of-proof objection to communications-based approaches developed below because conditions that are less conducive in this respect not only favor liability under such approaches but also indicate a smaller net social harm (or, possibly, the presence of a net gain).

position was obtained and is maintained through efficient, nonexclusionary behavior. It is recognized that this permissiveness can be costly because incentives may be excessive and some induced investment is wasteful (rent seeking), but it is granted in part because of expected net dynamic gains.

Coordinated oligopolistic price elevation is qualitatively different and generally worse in this regard. Firms are rewarded not for outperforming their rivals but rather to the extent that they refrain from such competition. Even though social benefits may sometimes ensue, they have an insufficient nexus to justify price fixing. Moreover, price fixing may actually reduce dynamic benefits. Successful coordination dampens the tendency of better firms to serve more consumers. And the anticipation of this effect reduces the ex ante reward to innovation. Overall, dynamic considerations tend to reinforce the case for discouraging coordinated price elevation.⁸

3.1.2. Framework for decision-making

If the detection of successful interdependent price elevation was cheap, easy, and accurate, the only remaining question would be how best to set sanctions. Unfortunately, it is not. Detection is costly and, particularly in cases involving circumstantial evidence, highly imperfect. Therefore, a central feature of optimal policy design can aptly be understood in terms of setting an appropriate burden of proof (Kaplow 2011d), which as a practical matter involves determining how much of what sorts of evidence in various contexts should be deemed sufficient to establish liability.

First, consider cases in which firms have actually engaged in coordinated oligopolistic price elevation. Here, the central benefit of the prospect of liability is deterrence, and, correspondingly, the social cost of false negatives (mistaken exoneration) is the dilution of deterrence. Second, consider cases in which the firms have not in fact elevated prices in a coordinated fashion. There, the prospect of sanctions will tend to chill (discourage) beneficial activity, notably in situations in which firms ex ante anticipate that their actions may generate a significant likelihood of false positives.

The pertinent cost-benefit calculus, therefore, involves trading off deterrence benefits and chilling costs (and also administrative costs). Demanding stronger proof tends to reduce both deterrence and chilling, whereas greater openness to less definitive demonstration enhances deterrence but amplifies chilling. Making this tradeoff is a multifaceted endeavor because, as will be seen in subsections 3.2 and 4.3, different channels of proof vary substantially in their contributions to deterrence and their risks of chilling desirable behavior. Accordingly, approaches toward evidence and inferences need to be functional, attending directly to these benefits and costs of assigning liability.

Likewise, it is dangerous to employ formalistic criteria that arbitrarily limit some modes of proof, privilege others without regard to context, and alter the target of inference in a way that deviates from or disregards the social objective. Under the direct approach developed in this section, what matters about detection and possible errors of either type are their effects on social welfare. Under formalistic legal tests, by contrast, errors are ordinarily defined by reference to the test itself, which at best is a proxy indicator of what behavior should be subject to sanctions. In particular, applying punishment in a case with, say, no actual prohibited communications, is regarded as a false positive. However, when such findings are made in cases in which

⁸This analysis reveals a core deficiency in the argument of some leading commentators that interdependent price elevation should be excused because monopolists' price elevation is legal. (In addition, this argument's logic excuses classic cartels, which these commentators seek to prohibit.)

coordinated oligopolistic price elevation actually occurred (but without proof of the legally requisite communications), the consequence of these so-called false positives is to enhance the deterrence of socially undesirable activity. Under the direct approach, this effect is taken to be a virtue, not a vice.

3.2. Detection

A direct approach focuses on the presence and magnitude of coordinated oligopolistic price elevation. An immediate implication, which is of central relevance for section 4, is that the means by which firms accomplish price elevation (whether particular sorts of communication or otherwise) is not in itself important. To be sure, information about firms' means may have evidentiary value. Of particular note, smoking-gun evidence of secret price-fixing meetings will be sufficient for liability. An implication is that much of the challenge concerns cases with only circumstantial evidence or other indicators that are not definitive. In such instances, proof comes in many varieties and none should be privileged a priori. Instead, the weight of any evidence—alone but more often in combination with other supportive and contrary evidence—should reflect consideration of the pertinent inferences and application of the decision-making framework articulated in subsection 3.1.2, that is, a functional rather than formalistic analysis.

3.2.1. Market-based evidence of successful oligopolistic coordination

Despite overlaps and interactions, it is convenient to categorize indicators of successful coordinated price elevation into four groups: pricing patterns, price elevation, facilitating practices, and the conduciveness of conditions (the last of which is taken up separately, in subsection 3.2.2, for emphasis). Because of the long history of confusion in the legal domain with regard to pricing patterns, it is useful to begin with a basic clarification: the presence of parallel pricing or other common behavior is not usually indicative of coordinated price elevation for the simple reason that it ordinarily characterizes competitive interactions as well. Vibrant competition tends to involve parallel pricing, such as when competitive grocers each respond to higher wholesale prices by raising their retail prices. In order to distinguish successful interdependence from competitive, independent, rivalrous behavior, it is not probative to examine traits shared by both categories; such evidence does not, as a matter of logic, even begin to make the case for coordinated price elevation. Instead, analysis must focus on differences: behavior consistent with oligopolistic coordination and inconsistent with competition, which favors liability; and behavior consistent with competition but inconsistent with interdependence, which disfavors liability.

With this point in mind, consider three phases of successful oligopolistic coordination: raising prices from a competitive to a supracompetitive level, maintaining elevated prices, and price drops, notably, as a consequence of price wars. Initiation or further augmentation of oligopolistic price elevation may be marked by a sharp price increase. That said, it is immediately apparent that one must distinguish sudden price increases in competitive markets, such as in the case of the aforementioned common cost shock for our grocers. Therefore, it is necessary to rule out concurrent cost increases or sudden demand shifts in order to infer coordinated elevation. A further complication is that sophisticated firms may anticipate precisely such inferences, leading them to adjust prices more gradually in order to disguise their behavior (Harrington 2005, Harrington and Chen 2006, Marshall, Marx, and Raiff 2008). But such subterfuge is not always feasible, and the greater complexity may necessitate more frequent

and elaborate communications that themselves leave traces. Note the implication that interdependent price increases *unaccompanied* by detailed, direct interfirm communications may be easier to distinguish from competitive behavior than when more explicit interchanges occur.

Firms seeking to maintain effective coordination may tend to have stickier prices or market shares than is true with competitive interaction because they wish to avoid frequent, potentially contentious fine-tuning (Carlton 1986, Levenstein 1997, Pesendorfer 2000). There are other explanations for pricing inertia, but sometimes they will be insufficient to explain stickiness, particularly when cost or demand changes are substantial.

Sudden, sharp price drops are as suspicious as sharp price increases, again, in the absence of cost or demand shocks. Price wars may be necessary to punish cheaters, or previously successful coordinated price elevation may quickly disintegrate as conditions for continuation become less conducive. When prices fall precipitously without other explanation, the preexisting price may well have involved coordinated elevation.⁹

In addition to searching for suspicious pricing patterns, one can attempt to detect price elevation itself. Although this method can sometimes be useful, it must be employed with particular care because of the risk of misidentification and thus of chilling efficient behavior of competitive firms. The problem is that elevated prices are ones above the competitive level, which is marginal cost, but accurately measuring marginal cost can be quite difficult—a familiar point in other competition policy settings, such as the assessment of market power and identification of predatory pricing.¹⁰ The concern here is that marginal cost may be underestimated due to the exclusion of fixed costs that really are variable, misattribution of supposedly common costs, and a failure to recognize that marginal costs may be rising steeply when near full capacity. Confident measurement of price elevation may sometimes be possible, particularly when there is supracompetitive pricing in some regions or to certain customer groups but not others (enabling a more controlled comparison). And other times, econometric techniques may illuminate whether firms' responses to changes in demand and cost indicate competitive behavior or action more akin to that of a monopolist, although these inferences can be difficult (Nevo 1998, Corts 1999). In any event, the decision-making framework of subsection 3.1.2 counsels in favor of caution, making it sensible in many settings to demand strong evidence of a significant price elevation before assigning liability on that basis.

Yet another approach examines firms' use of facilitating practices. Such practices might provide evidence of coordinated oligopolistic behavior or might themselves be a basis for liability, as discussed in subsection 2.3. In either case, it is important to examine the existence and importance of other, potentially procompetitive explanations.

3.2.2. Conduciveness of conditions

The conduciveness of conditions to successful coordinated price elevation refers to the degree to which price elevation is mutually profitable, firms' ability to coordinate on a particular price elevation, and their ability to sustain an elevated price by detecting and punishing defection. This subject deserves special attention here for two reasons. First, it is an extremely

⁹Punishment of coordinated oligopoly through detection of price cuts poses interesting and subtle strategic issues regarding firms' incentives to cut prices and cartel stability that are analyzed further in Harrington (2003) and Kaplow (2011a, 2013a). This analysis is also related to leniency policy and other enforcement tactics. See, for example, Aubert, Rey, and Kovacic (2006), Miller (2009), Motto and Polo (2003), and Spagnolo (2008).

¹⁰On distinguishing unilateral price elevation—which complexifies the use of this means of detection because such unilateral behavior involves prices above marginal cost—see Kaplow (2011a, 2013a).

important factor in determining liability under a direct approach in cases in which market-based evidence plays a dominant role. Moreover, it's mishandling under the communications-based approach—giving it negative rather than positive weight—constitutes a central reason that the approach goes awry, as will be explored in section 4.3.¹¹

To begin the analysis under a direct approach, conduciveness is a necessary condition for successful coordinated price elevation. If coordinated oligopoly pricing is impossible in a given case, we can be fairly confident that it is not taking place, and if conditions are merely quite uncondusive, success is unlikely. A key qualification is that the investigator or adjudicator may be mistaken about how uncondusive conditions actually are. Because firms ordinarily understand their industry better than outsiders do, strong evidence of clear attempts to coordinate or of success in doing so tends to show that the observer's contrary analysis of conditions is mistaken. In any event, assessments of the conduciveness of conditions should generally influence inferences about successful coordination.

It should also be emphasized that, although conduciveness is a necessary condition to success, it is not a sufficient one, even when conditions are highly conducive. A central reason for this reservation is that deterrence may be effective. In other respects as well, the relationship between conduciveness and successful coordination is neither tight (quite to the contrary) nor readily and confidently perceived by industry outsiders. Therefore, it ordinarily makes sense in cases largely reliant on marketplace behavior to insist on strong evidence of conduciveness and also to require substantial evidence of success, using the sorts of evidence discussed in the preceding subsection.

Conduciveness bears not only on the probability that coordinated price elevation exists but also on the magnitude of potential deterrence benefits and chilling costs, the central concerns under a direct approach toward price fixing. When conditions are not very conducive, any successful oligopolistic coordination is likely to result in price elevations that are small and short-lived. Moreover, chilling costs are likely to be greater when conditions are less conducive because such industries are likely to be more nearly competitive. Consider that chilling price increases in reaction to increases in cost or demand is especially inefficient when price initially equaled marginal cost. Therefore, condemnation when conduciveness is low has modest benefits and potentially large costs. This point is all the more significant because there are many markets in which conditions are quite uncondusive; if such cases are not screened out, they may give rise to many false positives, generating high chilling costs (and running up the legal system's administrative costs). In contrast, when conditions are highly conducive, undeterred firms that succeed tend to do the greatest damage, generating large and long-lasting price elevations. In addition, chilling costs are less worrisome because highly competitive behavior is less likely. This double distinction—greater deterrence benefits and lower chilling costs—makes it essential for price-fixing policy to accord significant positive weight to the conduciveness of conditions in cases relying primarily on circumstantial evidence.

¹¹Because the pertinent conditions themselves have been the subject of an extensive literature and are not particularly controversial, I forgo elaborating them here. Some of the familiar factors are the number of firms, product heterogeneity, other asymmetries across firms, price transparency, the structure of the buyer side of the market, demand uncertainty, and the coordinating group's market power.

3.2.3. Internal evidence and interfirm communications

Internal evidence may bear positively or negatively on each of the factors examined in the preceding two subsections. Cost or marketing data may indicate firms' beliefs about marginal cost and demand, and other internal evidence (such as strategy and decision memos and meeting notes) may convey firms' thinking and motivation for pricing and other actions. For example, when a firm suddenly raises its price, it will have reasons, usually reflected internally. If there was a corresponding increase in cost, this will be indicated by internal cost data, projections, and the decision-making process. For sudden price drops, similar internal information will have been generated that bears on whether the firm is a secret price cutter, responding to others' perceived cheating, or reacting to changes in cost or demand. A lack of internal information pertaining to changes in demand makes a demand shock an unlikely explanation for a price move. On the other hand, when a new product launch is shot down because studies detail the firm's lack of production capability or consumers' lack of interest, suppression of product heterogeneity (a facilitating practice) is an unlikely explanation.

Reliance on internal evidence is subject to important limitations. First, ascertaining the true explanation for the actions of a large firm can be difficult since its decisions are generated by groups of individuals with complex, overlapping, and sometimes conflicting duties and opinions. Second, lawyers and others may anticipate how internal tracks might give rise to liability and hence attempt to avoid leaving clear footprints. Even so, it seems plausible that sometimes fairly reliable inferences will be possible. Large firms cannot operate effectively without a sense of their costs and market demand, and they ordinarily engage in explicit, often elaborate decision-making that involves many individuals. It is difficult for lawyers to reach broadly and deeply into an organization, controlling how myriad individuals speak, write, email, and otherwise behave. Moreover, hiding coordinated behavior can be particularly difficult. Consider how hard it would be for multiple firms to fabricate a consistent story, which could survive scrutiny, about a supposed cost shock that did not occur. In contrast, with a real cost shock, substantial corroborating material would be created in the ordinary course of business by each firm, and it would be consistent with external indicators.

Interfirm communications constitute an additional and often important source of evidence in detecting successful oligopolistic coordination—that is, (even) under a direct approach that does not make liability turn on the particulars of firms' communications. Not only can smoking-gun evidence of secret meetings itself be determinative of liability, but public interfirm communications, such as interactive advance price announcements, may be revealing. Returning to our cost shock example, when this is the true explanation for firms' price changes, they will not need to engage in elaborate orchestration.

Also, as discussed in subsection 2.3, interfirm communications (both secret meetings and other activity, such as advance price announcements) can be viewed as facilitating practices. Such activity can constitute evidence bearing on whether successful coordinated price elevation took place (the focus here) and be seen as tactics that might themselves be prohibited.

Finally, consider how probative the absence of evidence of explicit, detailed interfirm communications might be. Because firms attempt to keep such communications secret, a failure to find them hardly proves that they did not occur. In addition, successful coordination is sometimes possible without elaborate communications. Both points interact with the analysis of the conduciveness of conditions in subsection 3.2.2. When conditions are most conducive—particularly when the number of firms is small—explicit communications are less essential and more difficult to detect. Therefore, when the danger of oligopolistic price elevation is the

greatest, we should be less bothered by the lack (or limited nature) of such evidence. But when conditions are less conducive—there are many firms and their coordination problem is complex—frequent, explicit discussions among large numbers of individuals may be required, so the absence of evidence of such interchanges would significantly weaken any inference of successful coordination.

3.2.4. Disallowance of cases based on circumstantial evidence?

It is an uncontroversial legal proposition that circumstantial evidence is sufficient to establish an antitrust conspiracy, and many regard this feature to be important because the prohibition of price fixing drives conspiracies underground, making detection difficult. Nevertheless, cases that rely primarily on circumstantial evidence tend to pose the greatest concern due to the possibility of mistaken imposition of liability and concomitant chilling effects. Moreover, permitting such proof may induce the initiation of many weak cases under some institutional arrangements. Accordingly, it is natural to contemplate a reform that would significantly limit the use of circumstantial evidence, perhaps even disallowing cases based primarily on such proof.¹²

Under the direct approach employed here, such a limitation does not in principle require an independent assessment. Throughout, in considering the appropriate weight to accord to different types of evidence, circumstantial or otherwise, one is consistently led to ask whether reliance on a manner of proof generates deterrence benefits in excess of chilling costs. Ultimately, the merits of evidentiary restrictions depend on empirical matters: the ability to achieve deterrence if sanctions are limited to cases with powerful direct evidence (many believe that deterrence is currently inadequate, although perhaps sanctions could be raised further), the related question of how often successful coordinated price elevation will turn up direct evidence, and how much error is involved through judicious use of circumstantial evidence (the precision of such efforts depending on the quality of the competition regime).

If one were to allow liability to be based only on direct evidence, further questions would be presented. First, just what is direct evidence? Most obvious might be secret meetings or written agreements. But other internal evidence can be fairly direct: it might document such meetings or be sufficiently clear regarding why firms raised or maintained prices to yield the requisite confidence. Of course, internal evidence varies greatly in terms of how clearly it indicates someone's views and whether, for example, authors of emails knew what they were talking about. Second, evidence is often stronger when viewed together: ambiguous messages suggesting coordinated price elevation, for example, might be contradicted by powerful evidence of unconducive conditions. If the latter were to be ignored because it is merely circumstantial, a mistaken inference may go uncorrected. The more challenging question is whether to disallow the imposition of liability when there is little or no direct evidence in support, a question contemplated further in this article's conclusion.

3.3. Sanctions

Detection must be combined with sanctions in order to deter coordinated oligopolistic price elevation. To provide an effective deterrent, fines and damages—viewed together for

¹²This alternative regime may be more attractive by comparison to a communications-based prohibition in light of the analysis in subsection 4.3.

present purposes—must reflect the magnitude of harm as well as the likelihood of detection.¹³ If expected sanctions are too low, as some fear them to be in most jurisdictions (particularly in light of what many regard to be a low detection rate), deterrence will be inadequate.

A further challenge in setting sections is to avoid excess because of the central concern with chilling costs. At a basic level, raising sanctions poses a tradeoff between deterrence and chilling. A more subtle point is that this tradeoff makes it important to gear sanctions to the magnitude of harm. Regarding deterrence, it is obvious that greater overcharges require higher sanctions whereas small overcharges can be deterred with lower sanctions. Often overlooked, however, is that chilling concerns affirmatively favor low sanctions in the latter setting. Chilling costs are greatest when industries are actually competitive, and, in those situations, apparent (although mistakenly identified) price elevations will usually be small. Hence, insistence on low sanctions in cases with low estimated price increments provides a useful, if incomplete, safety valve that diminishes chilling costs associated with the mistaken imposition of liability.

This analysis raises concerns about the practices of leading jurisdictions—such as the United States and European Union—that tend to impose fines equal to a given fraction of firms’ revenues, with little regard for the magnitude of the overcharge (a practice motivated by its simplicity). Overcharges vary widely: some are sufficiently high that price fixing may be profitable even after paying such fines, and others are so small that these fines may greatly exceed what is necessary for deterrence while also risking significant chilling costs.¹⁴ In addition, recent literature suggests that fines based on revenues suffer from deficiencies with regard to cartel prices and stability. See Bageri, Katsoulacos, and Spagnolo (2013) and Katsoulacos, Motchenkova, and Ulph (2015).

Fines and damage awards are increasingly being supplemented by imprisonment.¹⁵ A standard rationale is to help offset underdeterrence when monetary sanctions are inadequate. In particular, imprisonment (and fines on individuals) can usefully address two sorts of agency problems in firms. First, when a firm does wish to price fix, it may find it difficult to induce the cooperation of key employees if they individually face large sanctions. Second, a firm that wishes to avoid price fixing may nevertheless have lower-level employees who can profit thereby (perhaps by securing large bonuses or promotions). If the most a firm can do when caught (and itself having to pay huge fines) is fire the responsible employees, they may be insufficiently deterred, in which event government sanctions provide a valuable supplement.

¹³A recurring confusion in the literature reflects a failure to appreciate the basic implications of deterrence logic, originally explained in this setting by Posner (1969). Specifically, some wonder how oligopolistic firms can be expected to desist from coordinated price elevation in light of its rationality, whereas firms that anticipate sufficient penalties will no longer find such coordination to be rational.

¹⁴Regimes that limit severe sanctions, including the imprisonment of individuals, to cases with direct, smoking-gun evidence need not be much concerned about such chilling costs. However, the fact that sanctions are severe even when little or no overcharge seems to have occurred can undermine enforcement to the extent that the requirement that punishment, if any, be extreme may induce enforcers to avoid such cases altogether. This article focuses on cases in which liability is predicated on circumstantial evidence (whether under a direct or communications-based prohibition), wherein the concern for false positives is significant.

¹⁵My prior writing also examines injunctions. Although they do not seem central to price-fixing enforcement as a matter of theory (injunctions do not deter and ultimately must be enforced via the threat of other sanctions in any event) or actual practice, some commentary nevertheless takes injunctions to be central and then argues that, due to difficulties of implementation, liability for price fixing must accordingly be circumscribed.

4. Comparison of approaches

Section 3 presents a ground-up analysis of price-fixing policy, and the approach that emerges bears no resemblance to the communications-based sort of prohibition that most commentators advance. The latter method focuses on a subset of means to an undesirable end rather than on the end itself, and it does not focus on the tradeoff of deterrence benefits and chilling costs that is central under a direct approach to the problem. Accordingly, the conventional method does not seem to be a plausible candidate for an optimal policy. Nevertheless, a communications-based prohibition has monopolized attention for decades, so there is value in undertaking a systematic, side-by-side comparison with a direct approach to price fixing. As will be emphasized throughout this section, the emphasis is on cases based on circumstantial evidence because, when there is direct, smoking-gun evidence of price fixing, both regimes would apply sanctions (recall the discussion in subsection 3.2.3).

4.1. Preliminaries

To begin, let us restate the conventionally favored approach in operational terms and reflect further on how it relates to the social objective. Section 2 revealed significant difficulties in defining this prohibition—in distinguishing basic cases of what are taken to be clear violations from obvious instances of permissive activity, in articulating the agreement concept with language that captures a notion notably narrower than interdependence, and in dividing the phenomenon of communication and, relatedly, human language into usefully distinct subsets. These difficulties are implicitly revealed by commentators' general failure either to state their proposed rules in canonical form or to present a series of sharp examples that illustrate the intended division. The earlier analysis also shows that the contemplated sort of categorization must inevitably entail a substantial degree of formalism that is ungrounded in oligopoly theory.

For these reasons, this section will employ the abstraction introduced in subsection 2.3 of using the set X to refer to all types of communications (or other activity) deemed to constitute a prohibited agreement, with the complementary set X' referring to all those that do not. This formulation allows the analysis to be fairly general rather than to depend on the particulars of one or another possible restrictive definition of agreement.

Section 2 already alluded to the challenge that such a communications-based definition poses for detection—keeping in mind that our focus is on cases involving circumstantial evidence. (Recall again that those with clear, smoking-gun evidence generate condemnation in any event.) Not only will some acts in X and X' inevitably be virtually indistinguishable when viewed directly, clearly, and closely, but let us now also contemplate how much harder they will be to differentiate when viewed indirectly, obscurely, and from afar.

In reflecting on the envisioned enterprise, a number of problems are immediate. First, it is familiar that the harms associated with coordinated oligopolistic price elevation are essentially unrelated to the means by which it is achieved: use of some acts in X or only those in X' . To justify this otherwise arbitrary distinction requires that the communications-based method have a significant advantage over a direct approach with regard to detection, but a priori the opposite seems likely. The relevant question is whether exclusive reliance on circumstantial evidence of the use of practices in X will be better in distinguishing true coordinated price elevation from actual competition than will the alternative of considering the evidence most relevant to that

concern and examining directly how it illuminates the distinction rather than a qualitatively different (and welfare-irrelevant) one. This question largely answers itself.

Further suggesting skepticism about the communications-based approach, consider differences across industry settings. Deterrence benefits are smallest and chilling concerns greatest when conditions are more competitive, which is when they are less conducive to coordination. But these are the settings, as we will now see, in which the use of some acts in X and thus a finding of liability is more probable under the communications-based approach (following the typical but possibly mistaken assumptions of many courts and commentators).

The central policy justification for circumscribed liability—a hallmark of the communications-based prohibition—would seem to be the problem of false positives that may chill beneficial conduct. Nevertheless, it generally makes more sense to address this concern directly. That is, one should use all probative evidence, combining and weighing it in a manner that reflects deterrence benefits and chilling costs, and set a burden of proof that optimally trades off these two considerations. What makes no sense is to privilege a subset of the evidence and misuse much other evidence in the course of answering a socially irrelevant question (the use of acts in X), which, moreover, sometimes entails giving important evidence weight that is opposite from what it deserves.¹⁶

4.2. Paradox of proof: overview

The difference between the core implications of a direct approach and those of a communications-based prohibition can be appreciated by examining what I refer to as the paradox of proof. This paradox grows out of the interplay of two starting points:

- demanding more than the demonstration of successful interdependence by also requiring the use of certain sorts of communications in some set X, and
- needing to infer whether the latter is so from circumstantial evidence.

When communications by assumption are not observed (because we are concerned here with cases that lack smoking-gun evidence—wherein liability would be imposed under both approaches), one must nevertheless infer whether or not certain means of communication were used. In circumstantial evidence cases, this by definition involves an assessment of market conditions, notably, how conducive they are to successful oligopolistic coordination and whether such successful coordination appears to have occurred. Because the outcome, interdependent oligopoly pricing, might have come about in a number of ways, the process of making inferences about whether the unobserved communications employed by the defendants included at least some acts in set X or only ones in set X' is challenging, to say the least.

Under a communications-based approach, the analysis of whether we can infer the use of some acts in set X from circumstantial evidence begins by considering (as under a direct approach) how likely successful oligopolistic coordination is to have taken place. If that is sufficiently unlikely, then it does not make sense to infer from other circumstantial evidence that such coordination is being accomplished through improper means. Since settings involving no

¹⁶Kaplow (2013b) elaborates the general proposition that, if the concern is with excessive chilling costs relative to deterrence benefits, it is optimal to raise the standard of proof (as applied to the correct question) rather than to require the demonstration of some other or additional element (even if that element is significantly correlated to the consequences of concern). See also Kaplow (2011d), identifying the determinants of the optimal burden of proof in settings in which the concern is with providing ex ante incentives.

apparent (and actual) coordination are ubiquitous, the probability of the use of acts in X in such instances is presumably too low to justify a positive inference.

Suppose now that this first question about the presence of successful coordinated price elevation is answered affirmatively with some requisite likelihood. Under the direct approach of section 3, our analysis would be complete (although the requisite likelihood may well differ under a direct approach). Under a communications-based prohibition, however, we must then ask our second question: whether, under the circumstances of the market in question, such success is sufficiently unlikely to have been achieved in the absence of the use of some acts in X that we can accordingly infer that such were employed. This inference process is more involved because it necessitates an additional inquiry. Moreover, it is one that is much more difficult than the first.

Let us examine the logic by which this second question can be addressed. The answer is generally understood to depend on the degree to which market conditions are conducive to successful coordinated price elevation. When conditions are highly unconducive, it is supposed that the likelihood that some act in X was used is low because firms would not risk liability in circumstances in which there is little expected payoff. As successful coordinated price elevation becomes easier, the likelihood of the use of acts in X rises. Eventually, however, as conditions become sufficiently conducive, this likelihood is imagined to fall because firms would be expected to forgo legally risky methods when successful coordinated price elevation is sufficiently easy without them.¹⁷

For concreteness, consider Figure 1, which provides a simple illustration of this phenomenon (with a horizontal dotted line representing a 50 percent likelihood of the use of acts in X, which one might associate with the proof threshold under the preponderance of the evidence rule), and Figure 2, which highlights the implications for liability.

¹⁷In my fuller treatments of the problem (Kaplow 2011b, 2013a), numerous variations and qualifications are considered, a few of which are sketched in subsection 4.4. In some of them, there is no paradox region, which arises when the relevant curve (see Figure 1) continues to slope upward throughout or, if it does turn down at some point, it does not fall sufficiently to cross the dotted line from above. In that event, the two approaches are qualitatively the same in that both would assign liability in all cases above some requisite level of conduciveness. The difference is that the communications-based approach would set the threshold based on some requisite probability of the use of acts in X (which would in turn require identifying the relevant curve), whereas the direct approach would instead choose the threshold that optimally trades off deterrence benefits and chilling costs.

Figure 1: Ease of Success, Likelihood of Acts in X, and Liability

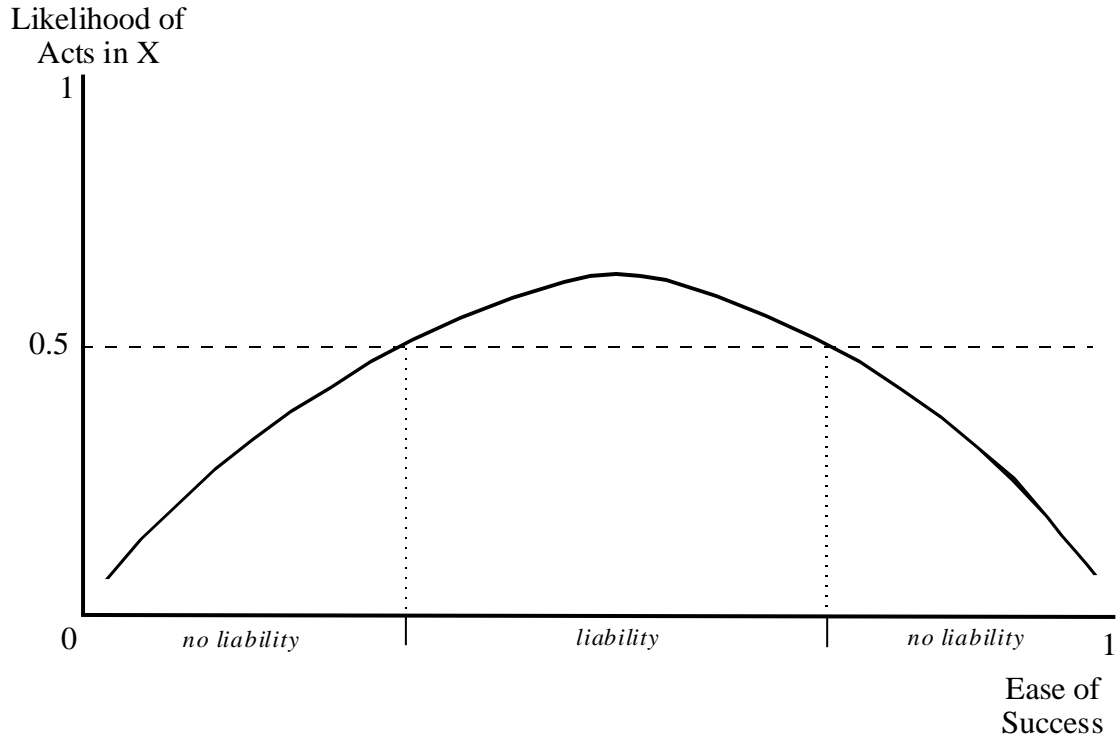
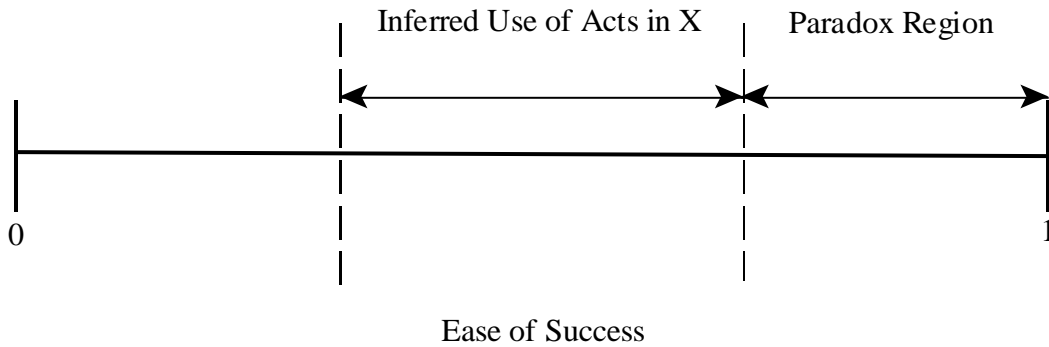


Figure 2: Ease of Success and Paradox of Proof



4.3. Comparison of outcomes

We can see immediately that the implications of a communications-based prohibition are indeed paradoxical as a matter of policy because it assigns liability in cases of moderate danger while exonerating defendants in cases posing the highest threat, that is, where the expected likelihood, magnitude, and longevity of price elevation are the greatest. To develop this point, let us consider some explicit comparisons with a direct approach.

First, suppose that the direct approach would find liability when the ease of successful coordinated price elevation is at least at the *left* boundary of the liability region in these figures. (Recall from subsection 3.2.2 that, under a direct approach, higher conduciveness, all else equal, favors liability, so a direct approach will entail liability starting at some point on the horizontal axis of these figures and including all points to the right.) Regarding deterrence, the communications-based approach gives up on the right, paradox region where, as already stated, harm from oligopolistic price elevation is the greatest.¹⁸ Additionally, chilling costs are worse under a communications-based prohibition (that is, worse per unit of deterrence benefit): in the middle region, where conduciveness is moderate, errors in identifying successful coordinated price elevation will be greater because it is less plausible that it is taking place. Moreover, as explained in subsection 3.1.2, chilling costs from errors tend to be the largest when conditions are more nearly competitive. In contrast, as one moves toward the right, expected error costs fall, so they are accordingly lower in the paradox region, where the communications-based approach fails to assign liability. In sum, the communications-based method concentrates liability in the central region, where deterrence benefits are smaller than in the paradox region and chilling costs are more serious.

To further dramatize this difference in outcomes, consider another comparison: a direct approach that assigns liability starting at the *right* boundary of the liability region—that is, at the division between the liability region and the paradox region. (To ease the exposition, suppose further that there are the same number of potential cases in the middle and right regions.) Observe that, relative to the communications-based approach depicted thus far, this direct approach just reverses the outcomes in the middle and right regions. This comparison is depicted in Figures 3A and 3B.

¹⁸This characterization need not hold at every point because (under either the direct or communications-based approach) it is possible that, as conduciveness rises, firms will, say, hold fewer meetings because they are both less necessary and raise the probability of detection. This offsetting factor tends to reduce the overall chance of success. Of course, if it does so sufficiently, firms would indeed take greater risks. It seems implausible that, overall, the social danger would often be much lower in the right than in the middle region, and discovered cartels achieving significant success (the largest, most prolonged price elevations) have been in industries in which conditions were indeed highly conducive.

Figure 3A: Communications-Based Approach

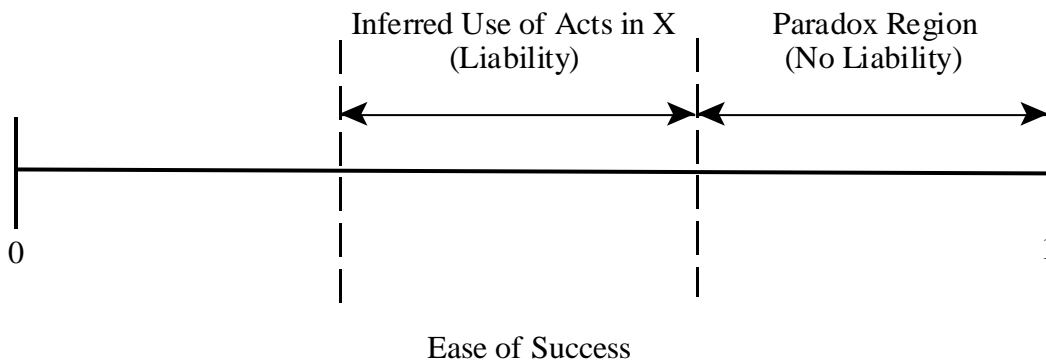
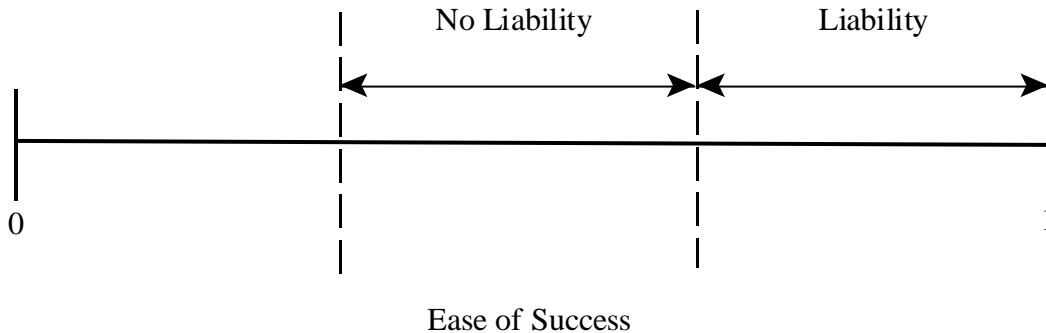


Figure 3B: Direct Approach



Can it possibly make sense to impose liability in cases that pose moderate danger and exonerate defendants when the danger is the greatest (Figure 3A), rather than doing the opposite (Figure 3B)? On an expected basis, situations in the rightmost region are more dangerous than those in the middle region: deterrence benefits are greater and chilling costs are lower. In terms of outcomes, the direct approach strongly dominates the communications-based prohibition.

Note that the analysis in this subsection can be seen as a restatement of the key lesson of subsection 3.2.2: greater conduciveness favors liability because deterrence benefits are higher and chilling costs are lower. A direct approach builds in this feature, whereas the communications-based prohibition perversely does the opposite in the circumstances depicted here.

4.4. Comparison of administrability

As subsection 4.2 explains, an approach to liability that requires the detection of prohibited communications (acts in X) through circumstantial evidence necessitates answering both the central question under the direct approach—whether successful coordination has taken place—and the additional one focused on here—whether any success required the use of at least some acts in X or only acts in X'. By its very nature, a communications-based prohibition is

more difficult to apply because it subsumes what must be done under a direct approach and also requires an additional, complex inquiry. This further determination actually is even more difficult than may already be apparent. Indeed, the foregoing discussion of the paradox of proof and the further elaboration here have a whimsical feel, indicating a glaring failure to analyze the implications of the broadly favored communications-based approach.

The magnitude of this practical, administrative problem can be seen by reflecting on Figure 1, which depicts a curve relating the ease of success to the likelihood of acts in X. Purely for purposes of explicating the logic of the paradox of proof, I drew a simple, symmetric curve of intermediate height. But in an actual case, where would information on the shape and height of this curve come from? Keep in mind that, unless one knows what this curve looks like, one cannot determine the location of the two boundaries of the middle, liability region, and hence know whether to assign liability—indeed, even if both parties were to stipulate to the likelihood of coordinated price elevation and the ease of successful coordinated price elevation. However difficult it is for economic experts to illuminate the question of whether success has occurred (and it often is quite hard), at least it has been the subject of decades of theoretical and empirical research. By contrast, precious little is known about the supplemental question relating to this all-important curve that must be pinned down if one is to apply the communications-based approach. And the answer, of course, will vary with the facts of each case.¹⁹

Although many variations are explored in Kaplow (2011b, 2013a), let us consider here one respect in which the relevant curve may differ from that in Figure 1, focusing on the breadth of the paradox region. A narrower region is depicted in Figure 4 and a broader one in Figure 5.

¹⁹Actually, the problem is much harder than the text's depiction suggests, among other reasons because the conduciveness of conditions is a multidimensional problem, and one that seems to exhibit limited empirical regularity. A further, substantial challenge is that the curve depends on how the prohibited set X is defined. For example, if the set X were smaller, the curve would be lower, but how much so at various points along the horizontal axis is hard to say. The challenge is magnified by the arguments in section 2 regarding the substantial ambiguity about what existing law requires on this dimension and what various commentators' preferred legal rules actually entail.

Figure 4: Narrow Paradox Region

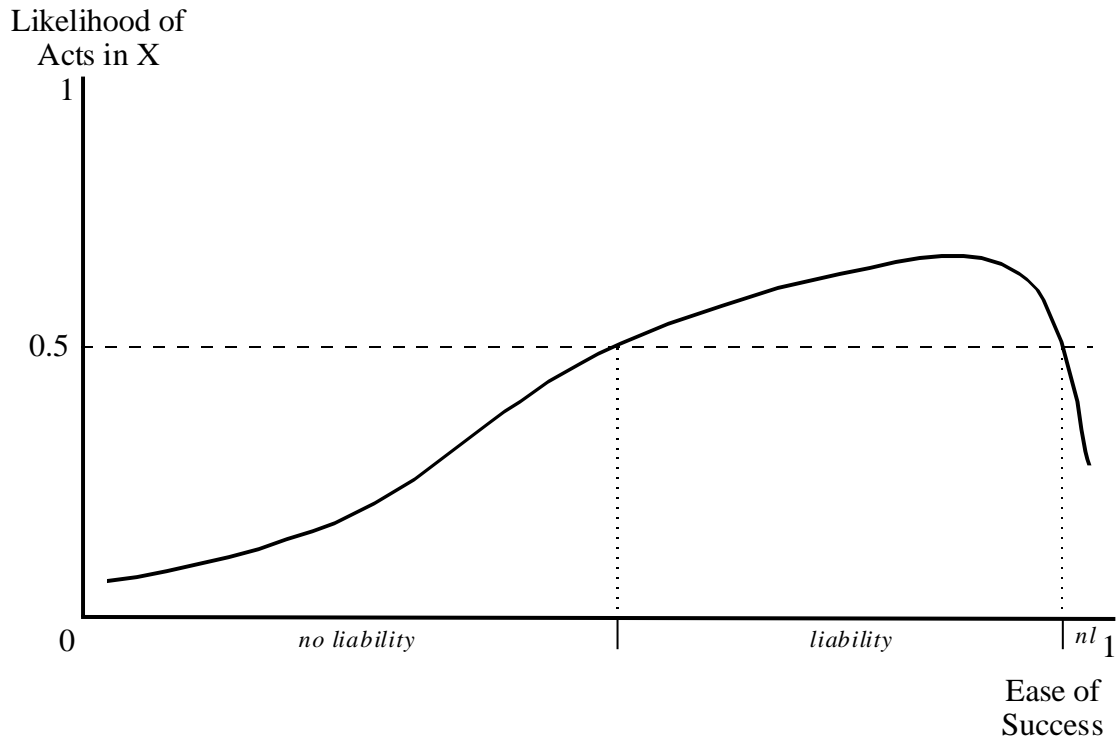
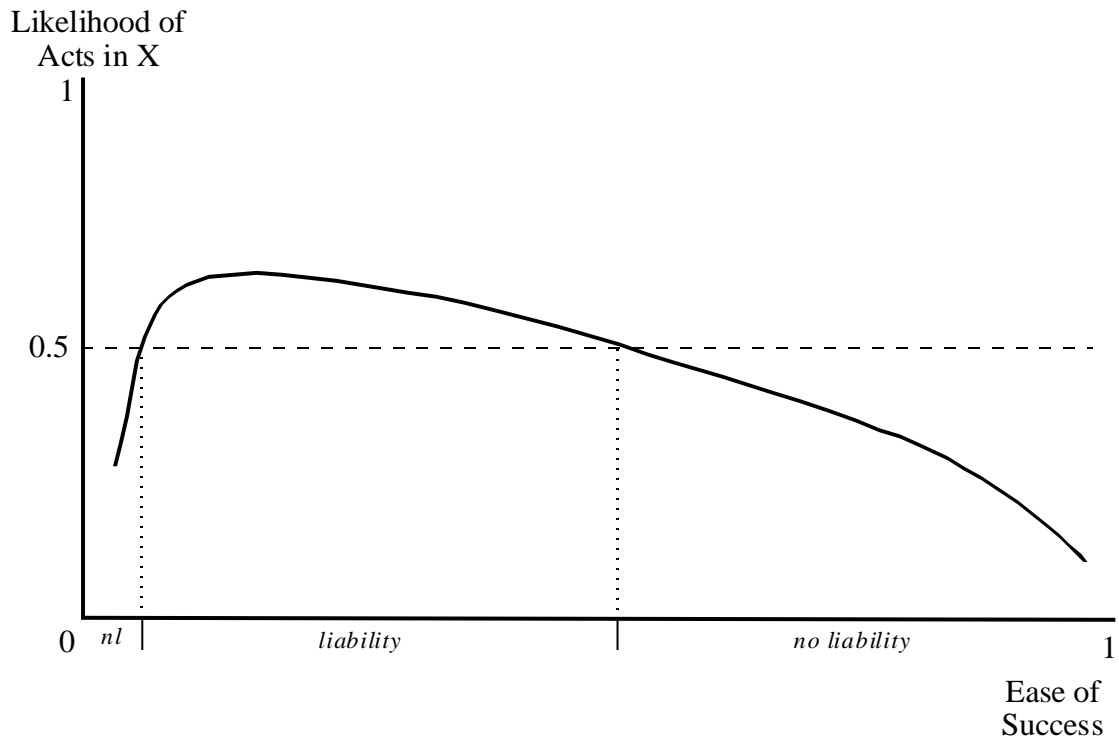


Figure 5: Broad Paradox Region



This juxtaposition suggests a severe problem in applying the communications-based prohibition: as drawn, the liability outcome is reversed between these two situations in virtually every case, specifically, all but those at the extremes of conduciveness. This observation drives home the point that, even if everything else in a case were uncontested, we would literally have no idea whether to assign liability without knowing a good deal about the pertinent curve.

These illustrations also cast further light on the comparison of outcomes in the preceding subsection. The scenario depicted in Figure 4, with a narrow paradox region, implies outcomes fairly similar to those under a direct approach (with a proof burden set so as to impose liability in the right half of the figure). The only difference is the rightmost slice, where the direct approach uniquely assigns liability. In this instance, although the number of cases with different outcomes may be small, the direct approach may be substantially superior. First, those cases that do differ are those posing the most extreme danger and involving the lowest expected chilling costs—that is, they are the strongest cases for liability. Second, under a communications-based approach, the existence of this right segment would encourage firms in situations that are in fact somewhat further to the left to dispute both the ease of success (where on the horizontal axis the case falls) and the precise shape and height of the curve (which determines the boundary). This introduces the possibility of additional error and drives up administrative costs.

A further irony is revealed by reflecting on commentators' views regarding the ability of firms to successfully implement coordinated price elevation without resort to particular sorts of communications. Given the vehemence with which most oppose the direct approach, one might have expected them to believe that the difference between the two methods in terms of outcomes involves a substantial number of cases, such as the situation depicted in Figure 1 or perhaps even Figure 5. Yet many have remarked (albeit usually without much supporting evidence) that successful coordinated price elevation is difficult if not rare in the absence of what they regard to be prohibited acts.²⁰ These unexpected views indicate the extent to which the implications of the paradox of proof are not appreciated.

5. Conclusion

This article draws on my prior work to articulate a framework for addressing price-fixing policy, one that departs substantially from most literature on the subject both in the questions it poses and, as a consequence, in many of the topics it explores. Section 2 begins with terminological and legal questions relating to the requirement of an agreement, conspiracy, or concerted action with regard to coordinated oligopolistic price elevation. The nearly universal insistence on defining these concepts to be narrower than interdependence proves to be deeply problematic in unappreciated ways: polar opposite cases are not readily distinguished, key terms are difficult to define in the intended manner, standard approaches are at odds with basic understandings about the nature of communication, and the core distinction is ungrounded in oligopoly theory, placing form over substance. This analysis should be seen as removing

²⁰See, for example, Areeda and Hovenkamp (2010, p. 230), Posner (1969, pp. 1574–75), and Turner (1962, pp. 665, 672–73). Observe that if this oft-expressed view were taken as fact, then in U.S. civil litigation plaintiffs' allegations (at the motion-to-dismiss stage) and evidence (at the summary judgment stage or after trial) that successful oligopolistic price elevation occurred would indicate not merely the plausibility that prohibited acts were used by the defendants but a very high likelihood that this was so.

roadblocks, thereby clearing the way to undertake explicit policy analysis of the problem, not as dictating conclusions. Sharper definitions reduce confusion in discourse but do not tell us how different regimes influence social welfare. To understand that, we need a direct approach to the problem, which is sketched in section 3.

This direct approach has three elements: identifying the social problem associated with coordinated oligopolistic price elevation, detecting the presence of such supracompetitive pricing, and determining appropriate sanctions. The analysis of detection receives the most attention because it poses the greatest challenge. Although the path is fairly easy when there is unmistakable smoking-gun evidence, it is also important to navigate the vast sea of potential cases in which most evidence is circumstantial, often relating to marketplace behavior but sometimes valuably supplemented by evidence internal to firms. Throughout the discussion, significant emphasis is accorded to the tradeoff of deterrence, which is enhanced by more aggressive enforcement, and chilling costs, which arise due to the prospect of mistaken classification. Because this latter concern is great, some attention is devoted to the possibility of eschewing liability in cases with only circumstantial evidence, or at least demanding very strong proof, perhaps corroborated by fairly clear internal evidence. In any event, the central framework, under which consequences for social welfare are addressed explicitly and traded off optimally, provides the correct approach to the problem, wherever it may lead.

Section 4 compares the direct approach to what most regard to be the better prescription (and also argue to be an apt description of existing law): a communications-based prohibition that demands (again, focusing on cases involving circumstantial evidence) proof not only of successful coordinated price elevation but also of the use of particular (unobserved) means in achieving it. Not surprisingly, requiring an additional inquiry that is incredibly difficult to undertake poses even greater administrative obstacles than those presented by the direct approach, which themselves are daunting. Worse, by focusing on the wrong question as a matter of policy, the communications-based prohibition produces perverse outcomes. Under conventional understandings of the inference process—elaborated here under the rubric of the paradox of proof—the communications-based prohibition focuses liability on cases that involve both less danger and greater chilling costs than those targeted by the direct approach. The latter instead aims explicitly at cases exonerated under the conventional rule: those with the greatest harm and the least chilling risk. This sharp advantage can hardly be a surprise: after all, a direct approach is explicitly concerned with the maximization of social welfare.

Focusing on the policy analysis of sections 3 and 4, one might be tempted to move straightaway to recommendations for the optimal rule in this domain. As my work emphasizes, however, there are three important sets of reservations. First, even though the analysis is both intensive and extensive, delving deeply into a broad range of issues, much of it traverses new territory and, in the process, reveals previously ignored pitfalls. A central purpose of this project is to launch a new dialogue. Despite significant efforts to fill out the map, I believe that readers should have much more confidence in the methodology—undertaking a direct approach to the social problem at hand, focusing on the three fundamental elements I analyze—than in any particular preliminary conclusions that might be drawn at this stage.

Second, many of the points illuminated by the analysis raise empirical questions about which existing knowledge is limited. (As mentioned, this gulf is even wider with regard to the communications-based prohibition.) To construct an optimal regime, we need to know more about the extent of coordinated oligopolistic price elevation in the economy, what modes of detection are most reliable in different contexts, and the nature and magnitude of chilling costs

that are generated by false positives. Previous commentary, in addition to asking the wrong questions, has been too quick to offer authoritative policy pronouncements despite this substantial void of empirical knowledge. Another motivation for this investigation, therefore, is that a sharper understanding of the policy-relevant questions can guide empirical research.

Third, other portions of my work emphasize institutional issues concerning the practical implementation of any price-fixing regime. Every step from the selection and screening of cases to final adjudication and the determination of sanctions may involve challenging context-dependent factual inquiries that impose heavy demands on even high-quality legal systems. Moreover, the nature of actual institutions varies greatly across jurisdictions and sometimes within them, notably, in the United States. Given the complexity of the task and the potential for significant error and administrative costs, it is possible that the best rule would depend significantly on the institutional setting. As but one illustration, the aforementioned possibility of disallowing liability in cases supported only by circumstantial evidence may be more appealing when mechanisms for screening are weak and decision-making quality is low. Indeed, detection challenges and the possibility that a prohibition may be misapplied or overzealously implemented counsels caution even in the most sophisticated regimes.

Before closing, let us contemplate some practical lessons. Suppose that some cases that are based largely on circumstantial evidence are to be pursued. Then perverse welfare consequences will result if violations are found when conduciveness to successful coordinated price elevation seems questionable, whereas firms are exonerated when success appears likely. Second, in allocating resources to cartel investigations and when deciding whether additional firms should be granted leniency in exchange for supplemental evidence, this article's analysis bears heavily on the question: evidence of what? Does the remaining uncertainty concern whether firms were actually coordinating on higher prices or instead merely the identification of the particular sorts of communications that led to such a result.²¹

Stepping back, the analytical approach to coordinated oligopolistic price elevation advocated herein is aimed primarily at the research community and only secondarily at competition agencies. My intention is to rekindle and advance a long-dormant policy debate on one of the most important aspects of competition law. Clearer thinking should refocus our endeavors and ultimately help to shape policy. To obtain the best understanding of how different regimes affect social welfare, it is best to address this question directly.

References

- Areeda, P.E., Hovenkamp, H. 2010. *Antitrust Law*, vol. 6, 3rd ed. Aspen, New York, NY.
- Aubert, C., Rey, P., Kovacic, W.E. 2006. The impact of leniency and whistle-blowing programs on cartels. *Intl. J. Indus. Org.* 24, 1241–1266.
- Bageri, V., Katsoulacos, Y., Spagnolo, G. 2013. The distortive effects of antitrust fines based on revenue. *Econ. J.* 123, F545–F557.
- Carlton, D.W. 1986. The rigidity of prices. *Am. Econ. Rev.* 76, 637–658.
- Corts, K.S. 1999. Conduct parameters and the measurement of market power. *J. Econometrics* 88, 227–250.

²¹This distinction is also relevant at such time as cases emerge involving algorithmic collusion.

- Harrington, J.E., Jr. 2003. Some implications of antitrust laws for cartel pricing. *Econ. Letters* 79, 377–383.
- Harrington, J.E., Jr. 2005. Optimal cartel pricing in the presence of an antitrust authority. *Intl. Econ. Rev.* 46, 145–169.
- Harrington, J.E., Jr., Chen, J. 2006. Cartel pricing dynamics with cost variability and endogenous buyer detection. *Intl. J. Indus. Org.* 24, 1185–1212.
- Kaplow, L. 2011a. An economic approach to price fixing. *Antitrust Law J.* 77, 343–449.
- Kaplow, L. 2011b. Direct versus communications-based prohibitions on price fixing. *J. Leg. Anal.* 3, 449–538.
- Kaplow, L. 2011c. On the meaning of horizontal agreements in competition law. *Calif. Law Rev.* 99, 683–818.
- Kaplow, L. 2011d. On the optimal burden of proof. *J. Polit. Econ.* 119, 1104–1140.
- Kaplow, L. 2013a. *Competition Policy and Price Fixing*. Princeton University Press, Princeton, NJ.
- Kaplow, L. 2013b. Multistage adjudication. *Harvard Law Rev.* 126, 1179–1298.
- Kaplow, L. 2016. The meaning of vertical agreement and the structure of competition law. *Antitrust Law J.* 80, 563–630.
- Katsoulacos, Y., Motchenkova, E., Ulph, D. 2015. Penalizing cartels: the case for basing penalties on price overcharges. *Intl. J. Indus. Org.* 42, 70–80.
- Kühn, K.-U., Vives, X. 1995. *Information Exchanges Among Firms and Their Impact on Competition*. Office for Official Publications of the European Communities, Luxembourg.
- Levenstein, M.C. 1997. Price wars and the stability of collusion: a study of the pre-World War I bromine industry. *J. Indus. Econ.* 45, 117–137.
- Mankiw, N.G., Whinston, M.D. 1986. Free entry and social inefficiency. *Rand J. Econ.* 17, 48–58.
- Marshall, R.C., Marx, L., Raiff, M.E. 2008. Cartel price announcements: the vitamins industry. *Intl. J. Indus. Org.* 26, 762–802.
- Miller, N.H. 2009. Strategic leniency and cartel enforcement. *Am. Econ. Rev.* 99, 750–768.
- Motta, M., Polo, M. 2003. Leniency programs and cartel prosecution. *Intl. J. Indus. Org.* 21, 347–379.
- Nevo, A. 1998. Identification of the oligopoly solution concept in a differentiated-products industry. *Econ. Letters* 59, 391–395.
- Pesendorfer, M. 2000. A study of collusion in first-price auctions. *Rev. Econ. Stud.* 67, 381–411.
- Porter, M.E. 1980. *Competitive Strategy: Techniques for Analyzing Industries and Competitors*. Free Press, New York, NY.
- Posner, R.A. 1969. Oligopoly and the antitrust laws: a suggested approach. *Stanford Law Rev.* 21, 1562–1606.
- Schelling, T.C. 1960. *The Strategy of Conflict*. Harvard University Press, Cambridge, MA.
- Spagnolo, G. 2008. Leniency and whistleblowers in antitrust, in *Handbook of Antitrust Economics*, P. Buccirossi, ed., 259–303. MIT Press, Cambridge, MA.
- Turner, D.F. 1962. The definition of agreement under the Sherman Act: conscious parallelism and refusals to deal. *Harvard Law Rev.* 75, 655–706.