Starting Up a Brand New Day—Four Key Consumer Protection Issues for the Five New Commissioners

Randy Shaheen identifies four key areas of consumer protection enforcement worthy of reconsideration as the new FTC Commissioners begin their terms. In particular, he highlights potential changes relating to identifying implied claims in advertising, defining when a “significant minority” of consumers have been misled, determining the appropriate value for consumer redress, and the use of warning letters.

A Primer on Bargaining: How Mergers May Affect Negotiated Prices

In our continuing series of articles explaining complex antitrust economics concepts to lawyers, Dov Rothman and David Toniatti provide a primer on how economists approach bargaining. They describe a basic framework and use that framework to discuss how mergers of suppliers, mergers of buyers, and vertical mergers may affect negotiated prices.

The Role of the Circle Principle in Market Definition

Bryan Keating, Jonathan Orszag, and Robert Willig discuss the importance of continued adherence to the “circle principle” in market definition. The authors explain how abandoning the circle principle can lead to key distortions in evaluating the competitive effects of mergers and respond to recent arguments for diminishing the role of the circle principle in market.

Book Review: Competition Policy and Collusion

Wentong Zheng reviews The Theory of Collusion and Competition Policy, a new book by Wharton School of Business Professor Joseph E. Harrington, Jr. Harrington’s book focuses on the theoretical impact of competition law and enforcement, as well as its optimal design. While largely a survey of economic research on competition policy and collusion, Harrington also identifies gaps in the existing economic literature on collusion and proposes areas for future research.
Starting Up a Brand New Day—Four Key Consumer Protection Issues for the Five New Commissioners

Randal M. Shaheen

For over a year the Federal Trade Commission has had only two Commissioners, Acting Chairman Maureen Ohlhausen, who has since been nominated for a judicial position, and Commissioner Terrell McSweeny, whose term has expired but is continuing to serve pending confirmation of her replacement. After a long nomination/confirmation process it seems that we are finally on the verge of having a new Federal Trade Commission Chairman. And not just a new chairman but also, for the first time since the Commission was formed in 1914, five new FTC Commissioners appointed in the same calendar year. While transitions are always a good time to take stock and evaluate what is working well and what could work better, this is a particularly opportune time to consider changes to how the Commission approaches some important aspects of its consumer protection mission.

Almost uniquely in Washington, the FTC’s Consumer Protection mission has proceeded with a substantial amount of bi-partisanship and has largely avoided the controversies that plague so much of what else goes on in D.C. However, there is always room to improve and so, mindful of the axiom, “If it ain’t broke, don’t fix it,” and conscious of the fact that other practitioners no doubt have their own wish lists of possible changes, this article lays out for consideration four key areas where newly appointed Commissioners might consider changes to consumer protection enforcement: identifying implied claims in advertising, defining when a “significant minority” of consumers have been misled, determining the appropriate value for consumer redress, and the use of warning letters.

Implied Claims

Section 5’s prohibition on deception applies to both express and implied claims. While express claims focus on what the advertisement literally says, implied claims are inferential or indirect messages that consumers take away from the overall context of the ad. It has until recently been almost axiomatic that the FTC has the power—and expertise—to determine the presence or absence of implied claims in consumer advertising. Thus, the FTC is not required to conduct consumer research studies to find implied claims and bring enforcement actions based upon such findings. The Commission has explained the process this way:

1 The pending nominees are Joseph Simons (Chairman), Noah Phillips, Christine Wilson, Rohit Chopra and Rebecca Slaughter. For more information on their backgrounds, see It’s a Full Slate for the FTC, ALL ABOUT ADVERTISING LAW.COM (Apr. 6, 2018), https://www.allaboutadvertisinglaw.com/2018/04/its-a-full-slate-for-the-ftc.html.

2 While this article is focused on the FTC’s traditional consumer protection enforcement, it is worth noting that the FTC’s enforcement activities relating to privacy and data security increasingly raise similar questions—for example, what implied messages do reasonable consumers takeaway from a company’s privacy disclosure and how should the Commission appropriately assess whether to impose consumer redress and, if so, the appropriate amount of redress.
Of course, the Commission must find that a representation, omission, or practice occurred. In cases of express claims, the representation itself establishes the meaning. In cases of implied claims, the Commission will often be able to determine meaning through an examination of the representation itself, including an evaluation of such factors as the entire document, the juxtaposition of various phrases in the document, the nature of the claim, and the nature of the transaction. In other situations, the Commission will require extrinsic evidence that reasonable consumers reach the implied claims. In all instances, the Commission will carefully consider any extrinsic evidence that is introduced. 3

While the FTC's Deception Statement takes something of a middle ground on when the Commission can determine for itself the presence of an implied claim and when it must rely upon extrinsic consumer survey evidence, in several recent high profile matters, then Commissioner Ohlhausen questioned whether the Commission's exercise of that power has gone too far. 4 The new Commission should give these arguments significant consideration—first, because the case law the Commission relies upon (discussed below) should be interpreted in a more limited fashion than is the Commission's common practice and, second, because there are sound policy arguments to limit the Commission's exercise of power in this area.

Acting Chairman Ohlhausen is not the first Commissioner to argue that the Commission is not well suited to use its own expertise to assess the possible presence of implied claims. This is a bipartisan issue. No less a luminary than Robert Pitofsky, who served as the FTC Chairman under President Clinton and previously as both a Commissioner and the Director of the Bureau of Consumer Protection, questioned the Court's granting the Commission such authority back in 1997.

Why questions of meaning should be submitted to the virtually unreviewable discretion of five commissioners of the FTC has never been articulated. Unlike other instances of deference to regulators as part of the administrative process, there is no reason to believe that commissioners of the FTC have unusual capacity or experience in coping with questions of meaning, nor any indication that successful regulation of advertising requires a balance of related regulatory considerations that commissioners are in a special position to handle. 5

The assumption that the power to determine the presence or absence of implied claims rests in the Commission derives largely from the Seventh Circuit’s decision in Kraft. 6 However, a closer reading of Kraft and other cases supports only the rare exercise of this power.

In the Kraft case the Commission had challenged a series of implied claims in advertising for Kraft Cheese Singles, asserting the existence of implied claims without the benefit of extrinsic consumer survey evidence. On appeal to the Seventh Circuit, Kraft argued that the Commission as a matter of law could not find the presence of an implied claim without extrinsic evidence of consumers’ perceptions of the challenged advertisements. The challenged ads noted that Kraft Cheese Singles were made from 5 ounces of milk so you got "all the calcium you need." The Commission argued that the ad falsely implied that a Kraft cheese single contained the same amount of calcium as 5 ounces of milk when, in fact, some calcium was lost during processing.

---


6 Kraft, Inc. v. FTC, 970 F.2d 311 (7th Cir. 1992).
While the Seventh Circuit held that “the Commission does not have license to go on a fishing expe-
dition to pin liability on advertisers for barely imaginable claims falling at the end of this spectrum
[obvious to barely discernible],” the Court found the Commission could exercise its own judg-
ment—rather than rely upon surveys—in situations where claims are “implied, yet conspicuous.”
Similarly, in Zauderer the Supreme Court held that no survey was required because it was self-
evident that when an attorney advertised that no fees were incurred unless the client won its case,
most consumers would understand “fees” to also include “costs.”

It is difficult to be certain whether the FTC has limited its implied claim divination authority to
situations where such claims are “conspicuous.” In some cases during the investigation phase
the staff may conduct a consumer survey in preparation for possible litigation, yet withhold that
fact from the advertiser. If the case subsequently settles, the existence of the survey may never
be known outside the Commission. However, the fact that Acting Chairman Ohlhausen has
expressed concern on several occasions that the Commission has proceeded to find implied
claims without the use of survey evidence is a clear indication of her knowledgeable belief that
the FTC has not appropriately limited the use of its authority.

What then were some of the findings of implied claims—without the benefit of survey evi-
dence—that then Commissioner Ohlhausen objected to in the POM Wonderful matter? In one
instance the Commission found that POM had made clinically “established” claims regarding
pomegranate juice relating to the prevention of heart disease and prostate cancer. At least some
of the ads in question noted in the body of the ad itself that the studies were “initial,” “preliminary”
and “may one day” prove effective. While the Commission may have been correct that, in
the context of the advertisement as a whole, the qualifying language was inadequate, it seems a
stretch to suggest that the implied claim at issue is “self-evident” or “conspicuous.” Similarly, the
majority found an implied claim that POM Wonderful prevents or reduces the risk of heart disease
(as opposed to supports heart health) based upon an ad that makes the rather general claims that
“scientific research . . . has uncovered encouraging results in prostate and cardiovascular
health.”

Similarly, in a case involving allegedly deceptive claims for an app that promised to help detect
skin cancer, then Commissioner Ohlhausen dissented, arguing that the Commission should have
used consumer survey evidence rather than its own judgment in making a determination that the

7 Id. at 320.
9 The failure to share the outcome of the survey with the advertiser raises another issue, which is the propensity of the Commission in con-
sumer protection matters to not make survey evidence or its experts available to the advertiser unless and until the matter proceeds to adju-
dication. The Commission has argued that it needs to protect its advantage should litigation ensue and that parties do not typically “share”
their evidence prior to litigation. While generally true in the private setting this argument ignores three important facts. First, to the extent
the advertiser has experts or survey evidence it is almost always shared with the Commission in an effort by the advertiser to avoid the vot-
ing out of a complaint. Thus, the “advantage” the Commission seeks is usually one-sided and some would say unfair. Second, the
Commission is both prosecutor and judge, ultimately charged with determining the truth of the matter rather than simply zealously repre-
senting one side. Arguably, opening up its experts and/or survey evidence to critique by the advertiser would lead to better, not worse deci-
sion-making, in that regard. Indeed, on the competition side, it is not at all uncommon for the Commission’s economists to exchange ideas
and arguments with the economists retained by the parties in a merger investigation. Finally, given that the Commission is a public rather
than a private party, the public interest may be better served by more transparency.

10 Ohlhausen POM Wonderful Statement, supra note 4, at 1–2, Fig. 13.
11 Id. Fig.12.
app promised to diagnose skin cancer as accurately as a dermatologist. The Commissioner argued that consumers could have just as easily taken away a message that the app was rather a helpful self-diagnostic tool to be used in conjunction with regular visits to a dermatologist. Further, as Commissioner Ohlhausen pointed out, this is not merely a semantic debate, as the quantity and quality of the required substantiation depends upon the nature of the claim. A claim that the app is a substitute for visiting a dermatologist presumably would require a much higher level of substantiation. And, as she concluded in her dissent, consumers may also be unintentionally harmed:

If the Commission continues to adopt such conclusions without any evidence of consumers’ actual interpretations, and thus requires a very high level of substantiation for health-related apps, we are likely to chill innovation in such apps, limit the potential benefits of this innovation, and ultimately make consumers worse off.

The FTC’s power to ascertain the existence of implied claims is not so dissimilar to how federal courts have wrestled with this issue in Lanham Act false advertising proceedings. In Lanham Act cases, challenges to express claims need not be supported by consumer survey evidence, while challenges to implied claims must prove the existence of such claims through the use of survey evidence. However, courts have fashioned an intermediate position whereby consumer survey evidence is not required for claims which can be found in the challenged advertising by “necessary implication.” Courts have defined “necessary implication” in several different but related ways, including:

1. “recognized as readily as if it had been explicitly stated;”
2. “the necessarily implied claim is the ‘practical, grammatical [and] syntactical equivalent’ of the literal words used;”
3. The implication is “unambiguous.”

By way of example, in a case involving Clorox, the advertiser promoted the whitening ability of its detergent “without chlorine” and then used the tag line, “Whiter is not possible.” The court found that survey evidence was not required to find an implied claim that the detergent provided whitening ability that was the equivalent of using a detergent plus chlorine. This outcome seems to fit squarely within all three of the above tests for “necessary implication.”

The FTC’s authority to use its own “expertise” to assess the possible presence of implied claims should be defined in a similar, rather limited, manner and the new Commission should take a hard look to see whether, as Acting Chairman Ohlhausen has suggested, it has abused its authority in this regard in the past.

Further, there are compelling arguments as to why the limited use of such power is also good policy. In the Kraft case the advertiser made a number of arguments that mirror many of those made by Chairman Pitofsky. At the outset, Kraft noted that consumers’ perception of advertising...
is shaped by a host of variables including their “social and educational backgrounds, the environment in which they view the ad, and prior experiences with the product advertised.”\(^{19}\) Having a five-member Commission discern how these variables play out among a much larger population is inherently unreliable. Further, Kraft argued that the Commissioners are fairly homogeneous in that they are highly trained attorneys and so cannot reflect the broader, more diverse public to whom products and services are marketed.\(^{20}\) Finally, Kraft attempted to factually rebut the argument that Commissioners have expertise in this area by noting that false advertising cases make up a small part of the agency’s overall work, that most Commissioners come to the job with little or no prior experience in advertising, and that the average tenure of an FTC Commissioner is fairly short.\(^{21}\) While the Seventh Circuit rejected Kraft’s argument that the Commission could never use its own expertise to find the existence of implied claims, it did embrace these concerns in significantly limiting the circumstances under which the Commission could do so.

Several other arguments also support this position. First, there are now numerous instances where the Commission has reversed its views on how consumers perceive particular claims, suggesting that its expertise in this area is overblown and has been exercised too broadly.

For example, the Commission’s 1980 Endorsement Guides offered advertisers using testimonials with “non-typical” results the option of either disclosing the typical results or clearly disclosing the “limited applicability of the endorser’s experience to what consumers may generally expect to achieve” (i.e., “results not typical” or “your results may vary.”).\(^{22}\) However, in conjunction with a later review of its Endorsement Guides, the Commission staff conducted two consumer surveys causing the Commission to conclude that such disclaimers did not serve to dispel an understanding by reasonable consumers that the represented experiences were typical. As a result, the Commission modified its guidance so that advertisers are now required to disclose what results are typical in connection with any “atypical” consumer testimonial.\(^{23}\)

Similarly, the FTC had opined that “up to” claims were not misleading so long as the “up to” result was what the average consumer can reasonably expect to achieve.\(^{24}\) In 2012, however, the Commission revisited the issue and, in connection with consumer research it had commissioned, concluded that window manufacturers could advertise “up to” energy savings claims only if “all or almost all” consumers are likely to achieve the maximum savings claimed.\(^{25}\) Here too it seems that the Commission’s initial attempts to determine on its own what the reasonable consumer takeaway might be were subsequently amended when actual consumer research was done.\(^{26}\)

\(^{19}\) Kraft, 970 F.2d at 318.

\(^{20}\) Id.

\(^{21}\) Id. at 319 (citations omitted).

\(^{22}\) Fed. Trade Comm’n, Guides Concerning the Use of Endorsements and Testimonials in Advertising § 255.2 (1980).

\(^{23}\) Id. 73 Fed. Reg. 72,374, 72,387 (Nov. 28, 2008) (to be codified at 16 C.F.R. pt. 255) (“Based upon the staff’s empirical research and its law enforcement experience, the Commission believes that disclaimers regarding the limited applicability of an endorser’s experience to what consumers may generally expect to achieve are unlikely to be effective. . . .”).


\(^{26}\) Various members of the FTC later expressed their view that the “all or almost all” standard was intended to apply in situations where consumers could not readily verify the claims for themselves and not in situations like savings claims where a retailers advertises “savings of up to __ \%.” However, the prior Commission matter involving “up to” claims involved fuel savings claims which also seem difficult for consumers to independently verify.
In each of these instances the Commission’s change of heart came about because it conducted consumer research, suggesting that the Commission’s prior effort to use its alleged expertise regarding consumer understanding of these claims—without the benefit of actual consumer research—was less than perfect. While these are only two examples, there are many other instances where the Commission has put itself in the shoes of ordinary consumers that have never been “tested” by a subsequent consumer survey, leaving open the question of how many other claims the Commission may have mistakenly interpreted.  

Misleading a “Significant Minority” of Consumers

My second issue is related to some extent to the first. When the Commission challenges an express claim, e.g. “relieves headache pain in less than one hour,” there is little reason to be concerned with what percentage of consumers take away the express claim. Simply by virtue of the claim being express, and therefore open and obvious, one can assume that most, if not all, consumers took away the misleading, challenged message. This question becomes important, however, when the allegedly deceptive claim is implied. In that case, some consumers may not take away the implied claim so that the question becomes: how many consumers can an advertiser deceive before the advertisement is deemed to violate Section 5? The Commission’s Statement on Deception provides at least a partial answer, noting that advertisements must be interpreted from the perspective of the “reasonable” consumer:

An advertiser cannot be charged with liability with respect to every conceivable misconception, however outlandish, to which his representations might be subject among the foolish or feeble-minded. Some people, because of ignorance or incomprehension, may be misled by even a scrupulously honest claim. Perhaps a few misguided souls believe, for example, that all “Danish pastry” is made in Denmark. Is it therefore an actionable deception to advertise “Danish pastry” when it is made in this country? Of course not. A representation does not become “false and deceptive” merely because it will be unreasonably misunderstood by an insignificant and unrepresentative segment of the class of persons to whom the representation is addressed.

The Commission then goes on to state that “[a]n interpretation may be reasonable even though it is not shared by a majority of consumers in the relevant class, or by particularly sophisticated consumers. A material practice that misleads a significant minority of reasonable consumers is deceptive.”

Of course, this begs the question as to what constitutes a “significant minority” or when the percentage of misled consumers is “insignificant.” Unfortunately, the FTC has avoided clear guidance on this question. In one instance, a company commissioned a survey that showed consumer confusion levels of 15 percent and argued that such a finding was “insignificant,” but the Adminis-

---

27 While these two examples involved the elapsing of a substantial period of time between the Commission’s first interpretation of the claim and the later modification, leaving open at least the possibility that consumer understanding of the claims may have shifted over time, the Commission’s ability to stand in the shoes of reasonable consumers should apply equally to determining when consumer understanding of a particular claim may have changed. Further, although each of the cited examples involved instances where the consumer surveys evidenced a more narrow or restrictive interpretation of the claim than the Commission had previously adopted, there is no reason to believe that the converse could not be true in some instances.


29 Deception Policy Statement, supra note 2, at 3 (citing Heinz W. Kirchner, 63 F.T.C. 1282, 1290 (1963)).

30 Id. at n.20 (citing Heinz W. Kirchner, 63 F.T.C. 1282 (1963)).
In another case, an Administrative Law Judge found that a consumer confusion level of 11 percent was sufficient to justify a finding of deception. However, the FTC to date has not articulated specific guidance as to what percentage is needed to satisfy its significant minority standard.

Lanham Act litigation is of little help in this regard. In one case, a court rejected a deception level of 7.5 percent as too low. In another case, however, a different court concluded that a deception level of 7.5 percent was high enough. Yet another court rejected deception levels in the low to mid-20s as insufficient. The only definitive guidance has been provided by the National Advertising Division of the Advertising Self-Regulatory Council (NAD), which has stated that 20% of survey respondents reporting a “miscommunication” is a “conventionally recognized significant level.”

In the face of all of this uncertainty, what is an advertiser trying to do the right thing supposed to do? Suppose a risk averse advertiser commissions a consumer survey of proposed advertising (a fairly expensive proposition that can cost as much as the high five or six figures.) The survey finds, after the use of an appropriate control that 14 percent of consumers take away a message that the advertiser does not intend and cannot support. The unlucky advertiser has just spent a lot of money on a consumer survey only to find itself in a gray area. Can it safely rely upon cases that suggest a higher requirement for a finding of deception? Of course, the risk averse advertiser can always default to the lowest possible finding for a “significant minority.” However, this course of action doesn’t necessarily benefit consumers. Should 90 percent of consumers not benefit from hearing a product or service message simply because 10 percent of other consumers might be misled? Perhaps the answer should depend in part upon the nature of the claim. Nevertheless, there seems to be no reason why the Commission cannot provide more guidance or even establish a threshold or perhaps a safe harbor rather than have well-intentioned companies play a guessing game and perhaps hold back from communicating certain beneficial messages out of an abundance of caution?

Determining the Appropriate Value for Consumer Redress

The fact that deception of a “significant minority” of consumers can constitute a violation of Section 5 is also one of several issues associated with trying to seek an appropriate amount for...
Consumer redress. In recent years, the Commission has more aggressively sought consumer redress to resolve consumer protection matters. Such redress is typically linked to the amount of consumer harm rather than the advertiser’s profits. As such, redress amounts can be the equivalent of total sales revenue and have ranged as high as nine figures.40 Here again Acting Chairman Ohlhausen has been a consistent voice for more carefully linking the redress sought to actual consumer harm.41 The ABA Section of Antitrust Law made similar recommendations in its 2016 Presidential Transition Report, where it called for “seeking monetary relief that corresponds more closely to the nature of the violation, the extent of consumer injury, and the culpability of the respondent/defendant” as well as providing more transparent guidance on how it determines the amount of any monetary sanction.42

The Commission also should consider providing guidance on what types of arguments it might reasonably consider in evaluating whether to reduce consumer redress. For example, in a case where the claims are express or a substantial majority of consumers were misled by a deceptive claim, no one would likely suggest that the FTC should reduce its redress demand. But what about a situation where most consumers—indeed, perhaps as many as 80 or 85 percent of consumers—were not misled? Is there any justification for requiring the advertiser to pay full redress when the vast majority of its customers got exactly what they expected and bargained for? While the Commission may argue that it exercises its discretion in this area and either opts not to seek redress or settles for a significantly reduced amount, the Commission has never repudiated its ability to seek full redress in such a situation, which makes for a fairly unbalanced negotiating position. Similarly, suppose a product deceptively claims that it will reduce the risk of diabetes but also offers other important vitamins and nutrients. Should the advertiser be able to argue that redress should be reduced based on the fact that the product delivered other valuable benefits and that some consumers might have purchased the product irrespective of the allegedly deceptive claim? No doubt there are other economic arguments that could be made as well, and guidance in this area would be helpful, particularly because most Commission orders simply state the redress number without further explanation.

Use of Warning Letters

A new Commission might also consider whether to continue and perhaps accelerate the recent trend of more frequent use of warning letters. As any FDA practitioner knows, the FDA is a much more prolific sender of warning letters than the FTC. The FDA’s explanation of its use of warning letters states:

> When it is consistent with the public protection responsibilities of the agency and depending on the nature of the violation, it is the Food and Drug Administration’s (FDA’s) practice to give individuals and firms an opportunity to take voluntary and prompt corrective action before it initiates an enforcement action. Warning Letters are issued to achieve voluntary compliance and to establish prior notice. . . . The use of Warning Letters and prior notice are based on the expectation that most individuals and firms will voluntarily comply with the law.

The agency position is that Warning Letters are issued only for violations of regulatory significance.

---

40 See, e.g., FTC v. AT&T Mobility LLC, Case No. 1:14-cv-3227-HLM (N.D. Ga. 2014).
41 See, e.g., Dissenting Statement of Acting Chairman Ohlhausen, FTC v. NetSpend Corp., Case No. 1:16-cv-04203-AT ((N.D. Ga. 2017) (arguing that some consumers did not abandon funds and were not entitled to redress).
Significant violations are those violations that may lead to enforcement action if not promptly and adequately corrected. A Warning Letter is the agency’s principal means of achieving prompt voluntary compliance with the Federal Food, Drug, and Cosmetic Act (the Act).43

Of course, there are pros and cons to warning letters. With responsible advertisers (or those who may not be so upstanding but have a well-developed sense of risk avoidance), a warning letter likely stops the alleged harm to consumers more quickly and reduces overall harm. To a certain degree, the FTC’s ability to collect consumer redress helps address any additional consumer harm after the fact, but consumer redress is not perfect. In some cases consumers deserving of redress cannot be identified or fail to come forward. In other cases, a company may plead inability to pay and the redress requirement is suspended in whole or in part.

Warning letters can also be an efficient use of Commission resources. For example, in “Operation Full Disclosure” the FTC sent warning letters to more than 60 national advertisers that it believed were not making adequate disclosures.44 In another more recent example, the FTC sent warning letters to 21 social media influencers, who were allegedly not sufficiently disclosing material connections to advertisers. These letters were follow-ups to more than 90 “educational” letters the Commission had previously sent.45 In each of these instances, the FTC’s letter-writing campaign garnered significant press attention and, at least anecdotally, appears to have resulted in an effort by many of the letter recipients to come into compliance. Time and money would have never permitted the FTC simultaneously to open an equivalent number of law enforcement investigations.

Warning letters also have downsides from an enforcement perspective. Unscrupulous advertisers can ignore them, and companies who modify their behavior but subsequently engage in similar deceptive behavior would not be subject to an order and, thus, not liable for civil penalties based on an order violation. However, on balance, there may be a case for a broader use of warning letters than the FTC has made to date. As is clear from the examples above, the FTC has in recent years made more frequent use of warning or “education” letters. It is perhaps time to examine these efforts and weigh whether they have helped or hindered the Commission’s consumer protection efforts.

As a group, the incoming slate of Commissioners do not bring an extensive amount of consumer protection experience to the table, but perhaps it is fitting that five fresh pairs of eyes take a look at whether to appropriately modify some longstanding Commission practices in the consumer protection area.

---

A Primer on Bargaining: How Mergers May Affect Negotiated Prices

Dov Rothman and David Toniatti

Prices are determined in different ways in different markets. In some markets, a market price is determined by the quantity suppliers choose to supply. In other markets, sellers set prices and buyers choose whether and how much to purchase at those prices. In other markets, buyers organize auctions, sellers offer bids, and prices are determined based on those bids and buyers’ purchasing decisions. In still other markets, buyers and sellers negotiate prices through bilateral bargaining.

In this article, we provide a primer on how economists approach bargaining. We describe a basic framework and use the framework to discuss how mergers of suppliers, mergers of buyers, and vertical mergers may affect negotiated prices. We illustrate the practical application of this framework to each type of merger in the context of (1) shoe suppliers and retailers, (2) hospitals and insurance providers, and (3) video content providers and distributors.

Bargaining Framework

The Setting. We consider bargaining between an intermediary and an input supplier. While bargaining may occur between any buyer and seller, prices in business-to-business transactions are often negotiated.

Gains-from-Trade. An intermediary and an input supplier will transact if there are “gains-from-trade.” There are gains-from-trade if the “gross benefit” the intermediary gets from purchasing from the supplier is greater than the cost the supplier incurs in selling to the intermediary.1

Consider the following hypothetical example of a retailer, Johnnie’s, negotiating with a shoe manufacturer, Nike, over the price of a shipment of shoes. We assume Johnnie’s has already reached agreements to carry Adidas and New Balance. We also assume Nike has already reached agreements to supply Jerry’s and Sportsman’s, two competitors of Johnnie’s.

If Johnnie’s carries Nikes, Johnnie’s will make a certain amount of revenue on its sales of Nikes. Alternatively, if Johnnie’s does not carry Nikes, the consumers who would have bought Nikes from Johnnie’s will buy something else. Some will buy Adidas or New Balance from Johnnie’s, while others will buy Nikes from Jerry’s or Sportsman’s. Johnnie’s will earn a certain amount of profit on the additional sales of Adidas and New Balance, while Nike will earn a certain amount of profit on the additional sales to Jerry’s and Sportsman’s.

Johnnie’s gross benefit from carrying Nikes is the difference between the revenue Johnnie’s would make on sales of Nikes and the profit Johnnie’s would make on these additional sales of

---

1 It is common to make the simplifying assumption that the seller knows the buyer’s gross benefit and the buyer knows the seller’s cost.
Adidas and New Balance. We assume Johnnie's would earn $10,000 in revenue from the sale of Nikes. We also assume that if Johnnie's does not carry Nikes, Johnnie's would earn $2,000 in additional profit from the sale of Adidas and New Balance. Thus, Johnnie's gross benefit is $8,000 ($10,000 – $2,000). We also assume Nike incurs $3,000 to make the shoes it sells Johnnie's, plus an opportunity cost of $1,000 from the profit Nike would earn on its additional sales to Jerry's and Sportsman's if it does not sell to Johnnie's. Nike's cost is therefore $4,000 ($3,000 + $1,000). Because Johnnie's gross benefit is $8,000 and Nike's cost is $4,000, there are gains-from-trade. Any price for the shipment of shoes between $8,000 and $4,000 will leave Johnnie's and Nike better off relative to not contracting.

**Bargaining Solution.** The price an intermediary and a supplier negotiate divides the gains-from-trade between them. If the intermediary is a more skilled or effective negotiator, it will be able to negotiate a lower price and capture a larger share of the gains-from-trade. If the supplier is a more skilled or effective negotiator, it will be able to negotiate a higher price and capture a larger share of the gains-from-trade. In our hypothetical example, if Nike is a more skilled negotiator, it will be able to negotiate a price closer to $8,000—the highest price Johnnie's is willing to pay. If Johnnie's is a more skilled negotiator, it will be able to negotiate a price closer to $4,000—the lowest price Nike is willing to accept. If Johnnie's and Nike are equally skilled negotiators, they will divide the gains-from-trade equally and negotiate a price that is equal to $6,000.

The negotiated price also will tend to be higher if the intermediary's gross benefit from purchasing from the supplier is larger. This may be the case if the intermediary charges a high price to its own customers. The intermediary's gross benefit will also be larger if it is more dependent on the supplier—e.g., if the intermediary does not have a good alternative to the supplier and so would earn very little profit if it does not purchase from it.

Returning to our hypothetical example, suppose Johnnie's can earn revenue of $12,000 rather than $10,000 on sales of Nikes. Johnnie's gross benefit would now be $10,000 ($12,000 – $2,000) rather than $8,000 ($10,000 – $2,000). If Johnnie's and Nike divide the gains-from-trade equally, the negotiated price will now be $7,000 rather than $6,000. Increasing Johnnie's gross benefit by $2,000 increases the negotiated price by $1,000 if Johnnie's and Nike divide the gains-from-trade equally. Alternatively, suppose again that Johnnie's can earn revenue of $10,000 on sales of Nikes, but can only earn an additional profit of $1,000 rather than $2,000 on sales of Adidas and New Balance if it does not carry Nikes. Johnnie's gross benefit from carrying Nikes would be $9,000 ($10,000 – $1,000) rather than $8,000 ($10,000 – $2,000). If Johnnie's and Nike divide the gains-from-trade equally, the negotiated price will now be $6,500 rather than $6,000.
The negotiated price also will tend to be higher if the supplier’s cost of selling to the intermediary is higher. This may be the case if the supplier’s incremental cost and/or opportunity cost is higher. The supplier’s opportunity cost may be higher if the supplier is less dependent on an intermediary—e.g., if the supplier has a good alternative to the intermediary and so could do well if it does not sell to it. For example, if Nike incurs an opportunity cost of $2,000 rather than $1,000 by selling to Johnnie’s because more consumers would buy Nikes from Jerry’s or Sportsman’s if Nike does not sell to Johnnie’s, then Nike’s cost will be $5,000 ($3,000 + $2,000) rather than $4,000 ($3,000 + $1,000). If Johnnie’s gross benefit from carrying Nikes is $8,000 and Nike’s cost is $5,000, the negotiated price will be $6,500 rather than $6,000. Increasing Nike’s cost by $1,000 increases the negotiated price by $500 if Johnnie’s and Nike divide the gains-from-trade equally.

A merger of suppliers may affect a negotiated price by changing the bargaining stakes. Prior to a merger, if the intermediary and one of the merging suppliers do not reach a supply agreement, the intermediary can purchase from the other merging supplier. After the merger, the intermediary may lose this backup option, because the merged suppliers can negotiate jointly after their merger. Depending on the extent to which the intermediary considers the merging suppliers to be substitutes for each other, the loss of this backup option may increase the importance to the intermediary of reaching an agreement with the merged supplier and may decrease the importance to the merged supplier of reaching an agreement with the intermediary. These changes in bargaining stakes may enable the merged supplier to negotiate a higher price with the intermediary.

Returning to our hypothetical example, consider a merger of Nike and Adidas. Prior to the merger, if Johnnie’s does not reach an agreement with Nike, some consumers who would have purchased Nikes from Johnnie’s will instead purchase Adidas or New Balance from Johnnie’s, while other consumers will purchase Nikes from Jerry’s or Sportsman’s. The more consumers who would purchase Adidas or New Balance from Johnnie’s, the stronger Johnnie’s bargaining position vis-à-vis Nike and the weaker Nike’s bargaining position vis-à-vis Johnnie’s. The more consumers who would purchase Nikes from Jerry’s or Sportsman’s, the weaker Johnnie’s bargaining position and the stronger Nike’s bargaining position.

8 The supplier’s opportunity cost will be lower if the supplier is more dependent on an intermediary—e.g., if the supplier does not have a good alternative to the intermediary and so could earn very little profit if it does not sell to the intermediary.

9 If the price is $6,500, Johnnie’s net benefit is $1,500 ($8,000 – $6,500) and Nike’s economic profit is $1,500 ($6,500 – $5,000).

10 If one of the merging suppliers does not reach an agreement with an intermediary prior to the merger, many consumers may continue to purchase from the intermediary if the intermediary’s product is just as good or almost as good when the intermediary purchases from the other merging supplier. This means that the intermediary may not lose many sales and the merging supplier may lose a lot of sales if the two parties do not reach a supply agreement. But if the intermediary’s product is much worse when it does not purchase from either merging supplier, many consumers may switch to other intermediaries who have agreements with the merged supplier. This means that the intermediary may lose many sales and the merged supplier may lose fewer sales if the intermediary and merged supplier do not reach an agreement.

11 The magnitude of any such price increase will depend on the extent to which the intermediary considers the merging suppliers to be substitutes for each other. If the intermediary does not consider them to be close substitutes, the merger may have little or no impact on the bargaining stakes. An intermediary may not consider merging suppliers to be close substitutes in markets in which there are multiple suppliers that are not highly differentiated.
Now consider what happens if Nike and Adidas merge and negotiate jointly with Johnnie’s. If Johnnie’s does not reach an agreement with Nike/Adidas post-merger, consumers who, prior to the merger, would have purchased Adidas instead of Nikes from Johnnie’s if Johnnie’s had not reached an agreement with Nike, no longer have the option to purchase Adidas because if Johnnie’s does not reach an agreement with Nike then it also does not reach an agreement with Adidas. Now these consumers will either purchase New Balance from Johnnie’s or purchase Nikes from Jerry’s or Sportsman’s. This latter group of consumers that purchases Nikes from Jerry’s or Sportsman’s represents sales Nike would have lost to Adidas prior to the merger, but now makes through Jerry’s or Sportsman’s after the merger. This group also represents sales Johnnie’s would have made prior to the merger, but now loses to Jerry’s or Sportsman’s after the merger. These diverted sales change the bargaining stakes—they decrease the importance to Nike of reaching an agreement with Johnnie’s and increase the importance to Johnnie’s of reaching an agreement with Nike. These changes in bargaining stakes may enable Nike/Adidas to negotiate a higher price with Johnnie’s.

A merger of suppliers also may generate efficiencies that reduce the merged supplier’s incremental costs. Given lower costs, the merged supplier may be willing to accept a lower price. As in posted price markets, at least some of any merger-specific reductions in incremental costs—efficiencies—may be passed through by the merged supplier to its customers. For example, if the merger of Nike and Adidas reduces Nike/Adidas’ cost of supplying shoes, the additional sales Nike/Adidas makes through Johnnie’s become more valuable to Nike/Adidas. This changes the bargaining stakes by increasing the importance to Nike/Adidas of reaching an agreement with Johnnie’s, and this may enable Johnnie’s to negotiate a lower price with Nike/Adidas.

**Hospital Merger.** Hospitals negotiate reimbursement rates with commercial health insurers. These reimbursement rates depend on the extent to which the hospital and insurer each benefit from the hospital participating in the insurer’s provider network. The hospital may benefit because participation in the insurer’s provider network may increase its patient volume. The insurer may benefit because the hospital’s participation in its provider network may increase the attractiveness of its provider network to purchasers of commercial health insurance—employers and individuals.

A merger of hospitals may increase the reimbursement rates negotiated with an insurer by changing the bargaining stakes. Prior to the merger, if an insurer and one of the merging hospitals do not reach an agreement, the insurer can still contract with the other merging hospital. After the merger, if an insurer and one of the merged hospitals do not reach an agreement, the insurer may not be able to contract with the other merged hospital. Depending on the extent to which consumers consider the merging hospitals to be substitutes for each other, the prospect of offering a provider network that does not include either merging hospital could be much worse than the prospect of offering a provider network that includes one but not the other merging hospital. This change in the bargaining stakes may enable the merged hospital to negotiate higher reimbursement rates with the insurer.12

A merger of hospitals also may generate efficiencies that reduce the merged hospital’s costs of providing medical care to an insurer’s enrollees. Given lower costs, the merged hospital may be willing to accept lower reimbursement rates.

---

12 The magnitude of any increase in reimbursement rates will depend on the extent to which consumers consider the merging hospitals to be substitutes for each other. If consumers do not consider them to be close substitutes, the merger may have little or no impact on the bargaining stakes. Consumers may not consider hospitals to be close substitutes in markets in which there are multiple hospitals that are not highly differentiated.
**Video Content Provider Merger.** Video content providers negotiate programming fees with multichannel video programming distributors (MVPDs). These fees depend on the extent to which the video content provider and MVPD each benefit from the video content provider participating in the MVPD’s programming network. The video content provider may benefit because participation in the MVPD’s network may increase viewership of its programs, and this may help it generate advertising revenue. The MVPD may benefit because the video content provider’s participation in its programming network may help the MVPD sell subscriptions.

A merger of video content providers may increase the fees negotiated with an MVPD by changing the bargaining stakes. Prior to the merger, if an MVPD and one of the merging video content providers do not reach an agreement, the MVPD can still contract with the other merging video content provider. After the merger, if an MVPD and one of the merged video content providers do not reach an agreement, the MVPD may not be able to contract with the other merged video content provider. Depending on the extent to which consumers consider the merging video content providers to be substitutes for each other, the prospect of offering a programming network that does not include either merging video content provider could be much worse than the prospect of offering a programming network that includes one but not the other merged video content provider. These changes in bargaining stakes may enable the merged video content provider to negotiate higher fees with the MVPD.

A merger of video content providers also may generate efficiencies that reduce their costs of developing programming. Given lower costs, the merged video content provider may be willing to accept lower fees.

**Merger of Intermediaries**

**Theory.** The impact of an intermediary merger on a negotiated input price is a bit more complicated. Given the positioning of intermediaries in the middle of supply chains, the impact of an intermediary merger on the prices the merging intermediaries negotiate with input suppliers may depend on the impact of the merger on the prices the merging intermediaries charge their own customers, and vice versa.

A merger of intermediaries may increase the importance to a supplier of reaching an agreement with the merged intermediary and decrease the importance to the merged intermediary of reaching an agreement with the supplier. These changes in bargaining stakes may create downward pressure on the input price the merging intermediary negotiates with the supplier.

---

13 In using the term MVPD, we take no position on which types of distributors are considered to be MVPDs.

14 The magnitude of any increase in fees will depend on the extent to which consumers consider the merging video content providers to be substitutes for each other. If consumers do not consider them to be close substitutes, the merger may have little or no impact on the bargaining stakes.

15 Prior to the merger, if one of the merging intermediaries does not purchase from a particular supplier, some consumers may instead purchase from the other merging intermediary. Post-merger, the merged intermediaries may negotiate with the supplier together. If the merged intermediaries do not purchase from the supplier, some of the consumers who would have substituted between the merging intermediaries may continue to purchase from the merged intermediaries, while other consumers may now substitute to another intermediary. The consumers who continue to purchase from the merged intermediaries represent sales that the merged intermediaries would have lost prior to the merger, but no longer lose after the merger if one of them ceases to purchase from the supplier. These retained sales reduce the merged intermediaries’ benefit from purchasing from the supplier, and this reduced benefit enables the merged intermediaries to negotiate a lower price with the supplier. These retained sales also represent sales that the supplier would have recaptured prior to the merger, but would not recapture after the merger. These lost sales decrease the supplier’s cost of contracting with the merged intermediaries, which also helps the merged intermediaries to negotiate a lower price with the supplier.
Returning to our hypothetical example, we now consider a merger of Johnnie's and Jerry's. Prior to the merger, if Nike does not reach an agreement with Johnnie's, some consumers will purchase Nikes from Jerry's or Sportsman's. The ability to make sales through Jerry's and Sportsman's gives Nike a stronger bargaining position when negotiating with Johnnie's, which enables Nike to negotiate a higher price with Johnnie's.

Now suppose Johnnie's and Jerry's merge and negotiate with Nike jointly. If Nike does not reach an agreement with Johnnie's/Jerry's, consumers who would have purchased Nikes from Jerry's if Johnnie's does not carry Nikes now will either purchase Nikes from Sportsman's or purchase Adidas or New Balance from Johnnie's/Jerry's. This latter group of consumers that purchases Adidas or New Balance from Johnnie's/Jerry's represents sales that Nike would have made prior to the merger, but now loses after the merger. This group also represents sales that Johnnie's would have lost prior to the merger, but now retains after the merger. These sales change the bargaining stakes—they increase the importance to Nike of reaching an agreement with Johnnie's/Jerry's and decrease the importance to Johnnie's/Jerry's of reaching an agreement with Nike. These changes in bargaining stakes will enable Johnnie's/Jerry's to negotiate a lower price from Nike.

The merger of intermediaries may also increase the merging intermediary's market power in the downstream market in which it is a seller, and this may create upward pressure on the price the merging intermediary charges in the downstream market. For example, prior to the merger, Jerry's constrains Johnnie's prices and vice versa. If Johnnie's and Jerry's merge, they no longer constrain each other's prices and this will tend to create upward pressure on their prices.

These direct effects may subsequently have countervailing effects. The direct downward pressure on the negotiated input price may create downward pressure on the price the merging intermediary charges in the downstream market, while the direct upward pressure on the price the merging intermediary charges in the downstream market may create upward pressure on the negotiated input price. For example, direct downward pressure on the price Johnnie's/Jerry's negotiates with Nike may create downward pressure on the price Johnnie's/Jerry's charges consumers, while the direct upward pressure on the price Johnnie's/Jerry's charges consumers may create upward pressure on the price Johnnie's/Jerry's negotiates with Nike.

The merger of intermediaries may also generate efficiencies that reduce the merging intermediary's incremental costs of selling its products. Any such efficiencies can create additional downward pressure on the price the merging intermediary charges in the downstream market and upward pressure on the input price the merging intermediary negotiates with the supplier. For example, if the merger of Johnnie's and Jerry's reduces its cost of selling shoes, the additional sales Johnnie's/Jerry's makes by carrying Nike are more valuable to it. This may change the bargaining stakes by increasing the importance to Johnnie's/Jerry's of reaching an agreement with Nike and this may enable Nike to negotiate a higher price with Johnnie's/Jerry's.

Thus, a merger of intermediaries may create downward or upward pressure on negotiated input prices and on the prices the merging intermediaries charge. How these countervailing effects balance out will depend on the situation.  

Health Insurer Mergers. Given the positioning of health insurers in the middle of supply chains, the impact of a merger of insurers on the reimbursement rates they negotiate with health

16 We take no position as to whether decreases in negotiated input prices should be considered cognizable efficiencies when evaluating the impact of a merger of intermediaries on competition in the downstream market.
care providers may depend on the impact of the merger on the premiums they charge their own customers, and vice versa.

A merger of insurers may create downward pressure on the reimbursement rates the merging insurers negotiate with health care providers, and upward pressure on the premiums the merging insurers charge their own customers. The downward pressure on the negotiated reimbursement rates may subsequently create countervailing downward pressure on premiums, while the upward pressure on premiums may subsequently create countervailing upward pressure on negotiated reimbursement rates. A merger of health insurers may also generate efficiencies that reduce the merging insurer’s incremental costs of selling its insurance products. Any such efficiencies can create additional downward pressure on premiums and upward pressure on negotiated reimbursement rates.

MVPD Merger. Like commercial health insurers, MVPDs are positioned in the middle of the supply chain, and so the impact of a merger of MVPDs on the programming fees they negotiate with content providers may depend on the impact of the merger on the prices they charge their own subscribers, and vice versa.

A merger of MVPDs may create downward pressure on the programming fees negotiated with video content providers and upward pressure on the prices charged to subscribers. The downward pressure on the negotiated programming fees may subsequently create countervailing downward pressure on subscriber prices, while the upward pressure on subscriber prices may subsequently create countervailing upward pressure on negotiated programming fees. A merger of MVPDs may also generate efficiencies that reduce the merging MVPD’s incremental costs. Any such efficiencies may create additional downward pressure on subscriber prices and upward pressure on programming fees.

Merger of an Input Supplier and an Intermediary

Theory. The merger of an input supplier and an intermediary—a vertical merger—also introduces additional considerations, but the same bargaining framework can be used to understand how the integration may affect the input prices the integrated supplier negotiates with now rival intermediaries and the input prices the integrated intermediary negotiates with now rival suppliers.

The integration of a supplier and an intermediary may create upward pressure on the prices the integrated supplier negotiates with rival intermediaries. By selling its input to a rival intermediary, the integrated supplier makes the rival intermediary more competitive, which may decrease the profits of the integrated intermediary. Prior to the integration, the integrated supplier does not consider this effect when negotiating with the rival intermediary. After the integration, the integrated supplier may internalize the effect. This reduces the integrated supplier’s gross benefit from contracting with the rival intermediary, creating upward pressure on the price the integrated supplier negotiates with the rival intermediary. For example, suppose Nike purchases Johnnie’s. By selling shoes to Jerry’s and Sportsman’s, Nike makes Jerry’s and Sportsman’s more attractive to consumers who want to buy Nikes. This decreases Johnnie’s profits. Prior to purchasing Johnnie’s, Nike does not consider this effect when negotiating prices with Jerry’s and Sportsman’s. After purchasing Johnnie’s, Nike may internalize the effect, which reduces Nike’s gross benefit from con-

17 If the reduction in the integrated supplier’s gross benefit is large enough, the gains-from-trade could disappear, in which case the integrated supplier might not be willing to contract with the rival intermediary.
tracting with Jerry’s and Sportsman’s and creates upward pressure on the price Nike negotiates with Jerry’s and Sportsman’s.

The integration of a supplier and an intermediary may also create downward pressure on the prices the integrated intermediary negotiates with rival suppliers. By contracting with a rival supplier, the integrated intermediary reduces the sales of the integrated supplier, which decreases the integrated supplier’s profits. Prior to the integration, the integrated intermediary does not take this effect into account when negotiating with the rival supplier. After the integration, the integrated intermediary may internalize the effect. This reduces the integrated intermediary’s gross benefit from contracting with the rival supplier, creating downward pressure on the price the integrated intermediary negotiates with the rival supplier.\(^\text{18}\) For example, by carrying Adidas and New Balance, Johnnie’s sells fewer Nikes, which decreases Nike’s profits. Johnnie’s takes this effect into account when negotiating with Adidas and New Balance after its merger with Nike. This effect reduces Johnnie’s gross benefit from carrying Adidas and New Balance and creates downward pressure on the price Johnnie’s negotiates with Adidas and New Balance.

The merger of a supplier and an intermediary may also affect competition among intermediaries. Post-integration, the integrated supplier may no longer “charge” the integrated intermediary a mark-up. This could improve the integrated intermediary’s cost position and enable the integrated intermediary to reduce the price it charges its customers. At the same time, the integrated intermediary may now take into account the effect of lowering its price on the profits the integrated supplier earns selling to rival intermediaries, and this may reduce the integrated intermediary’s incentive to reduce its price. For example, Nike may no longer charge Johnnie’s a mark-up on Nikes and this could enable Johnnie’s to reduce the prices it charges for Nikes. But if Johnnie’s reduces the prices it charges for Nikes and therefore sells more Nikes, Nike may make fewer sales to Jerry’s and Sportsman’s, and this may reduce Johnnie’s incentive to reduce the prices it charges for Nikes.

Thus, a merger of a supplier and an intermediary may create upward pressure on the prices the integrated supplier negotiates with rival intermediaries, and downward pressure on the prices the integrated intermediary negotiates with rival suppliers, and could result in intermediaries charging their customers higher or lower prices.

**Hospital-Insurer Merger.** The merger of a hospital and an insurer may affect the bargaining between the integrated hospital and what are now rival insurers—the merger may decrease the integrated hospital’s gross benefit from contracting with rival insurers, and this can create upward pressure on the reimbursement rates the integrated hospital negotiates with rival insurers.\(^\text{19}\)

The merger of a hospital and an insurer may also affect the bargaining between the integrated insurer and what are now rival hospitals—the merger may decrease the integrated insurer’s gross benefit from contracting with rival hospitals, and this decreased gross benefit can create

---

\(^{18}\) If the reduction in the integrated intermediary’s gross benefit is large enough, the gains-from-trade could disappear, in which case the integrated intermediary might not be willing to contract with the rival supplier.

\(^{19}\) By participating in a rival insurer’s network, the integrated hospital may make the rival insurer a stronger competitor, and this may decrease the integrated insurer’s profits. Prior to the integration, the integrating hospital ignores this effect when negotiating with a rival insurer. After the integration, the integrated hospital may internalize the effect. This internalization would reduce the extent to which the integrated hospital benefits from contracting with the rival insurer, which would put upward pressure on the reimbursement rates it negotiates with the rival insurer. If the reduction in the integrated hospital’s benefit is large enough, the gains-from-trade could be eliminated, in which case the integrated hospital might not be willing to contract with the rival insurer.
downward pressure on the reimbursement rates the integrated insurer negotiates with the rival hospitals.\textsuperscript{20}

The merger of a hospital and an insurer may also affect competition among insurers. Post-integration, the integrated hospital may no longer “charge” the integrated insurer a mark-up. The removal of such a mark-up could improve the integrated insurer's cost position and enable it to reduce the premiums it charges its customers. At the same time, the integrated insurer may now take into account the effect of lowering its premiums on the profits the integrated hospital earns treating enrollees of rival insurers, and this may reduce the integrated insurer's incentive to reduce its premiums.\textsuperscript{21}

\textbf{MVPD-Video Content Provider Merger.} The merger of a video content provider and an MVPD may result in the same types of effects as a merger of a hospital and an insurer. The merger may decrease the integrated video content provider's gross benefit from contracting with what are now rival MVPDs, and this can create upward pressure on the programming fees the integrated video content provider negotiates with rival MVPDs. The merger may also decrease the integrated MVPD's gross benefit from contracting with what are now rival video content providers, and this decreased benefit can create downward pressure on the programming fees the integrated MVPD negotiates with rival video content providers. By lowering the integrated MVPD's costs and raising the rival MVPDs' costs, the merger may also affect competition among MVPDs.

\section*{Wrap Up}

A buyer and supplier should transact when there are gains-from-trade. There are gains-from-trade when the buyer's gross benefit from purchasing from the supplier is larger than the supplier's cost of selling to the buyer. When there are gains-from-trade between a buyer and supplier, a negotiated price will serve to divide the gains-from-trade. A negotiated price will tend to be higher when the buyer's gross benefit is larger, when the supplier's cost is higher, and/or when the supplier is a better negotiator than the buyer.

A merger of suppliers may affect a negotiated price by changing the bargaining stakes—increasing the importance to a buyer of purchasing from a supplier and decreasing the importance to the supplier of selling to the buyer. The impact of a merger of intermediaries on the prices negotiated with input suppliers may depend on the impact of the merger on the prices the intermediary charges its own customers, and vice versa. An intermediary merger therefore may result in either a decrease or an increase in the prices negotiated with input suppliers. Finally, a vertical merger may enable the integrated supplier to negotiate higher prices with rival intermediaries, enable the integrated intermediary to negotiate lower prices with rival suppliers, reduce the prices the integrated supplier “charges” the integrated intermediary, and could result in intermediaries charging their customers higher or lower prices.●

\textsuperscript{20} By contracting with a rival hospital, the integrated insurer may reduce the volume of patients who seek care at the integrated hospital, and this may reduce the integrated hospital's profits. Prior to the integration, the integrating insurer ignores this effect when negotiating with the rival hospital. After the integration, the integrated insurer may internalize the effect. This internalization would reduce the extent to which the integrated insurer benefits from contracting with the rival hospital, which would put downward pressure on the reimbursement rates it negotiates with the rival hospital. If the reduction in the integrated insurer's benefit is large enough, the gains-from-trade could be eliminated, in which case the integrated insurer might not be willing to contract with the rival hospital.

\textsuperscript{21} If the integrated insurer reduces its premiums, it may win enrollees away from rival insurers, which means that the integrated hospital may treat fewer enrollees of rival insurers.
The Role of the Circle Principle in Market Definition

Bryan Keating, Jonathan Orszag, and Robert Willig

The current version of the Horizontal Merger Guidelines released in 2010 articulate a less structured description of what constitutes a valid relevant market than did the prior Guidelines issued in 1992, even as the current Guidelines downplay the role of relevant markets in the analysis of mergers of sellers of differentiated products. The prior Guidelines articulated an algorithmic process that begins with a product of one of the merging parties and then adds products or geographic units in order of closeness of substitution until the hypothetical monopolist test is satisfied. The principle of adding products or geographic units to a candidate market in order of closeness of substitution is sometimes referred to by practitioners as the “circle principle” because it evokes the image of building outward from a candidate market in concentric circles.

While continuing to rely on the hypothetical monopolist test to define relevant antitrust markets, the current Guidelines abandon the prior Guidelines’ adherence to the algorithmic process for implementing that test. Nonetheless, the current Guidelines include the ideas that relevant markets normally should not exclude products that are as or more substitutable as other products that are included, and that relevant markets should usually be no larger than necessary to pass the hypothetical monopolist test. These revisions of the Guidelines have generated renewed focus on the proper method for defining relevant markets in recent merger cases, such as Pinnacle/Hershey and Aetna/Humana.

In this article, we show that there are compelling reasons, based on policy and economic analysis, for continued adherence to the circle principle in market definition. First, we articulate the underlying roles of market definition in antitrust analysis of horizontal mergers. Next, we show that for each of these roles, violations of the circle principle would lead to potentially key distortions in the ensuing assessments of the effects of the horizontal merger. Finally, we address some recent arguments that have been offered in this publication by Kostis Hatzitaskos and co-authors as support for diminishing the role of the circle principle in the market definition exercise. Indeed, we show that Hatzitaskos et al.’s claim that there are “reasonable exceptions” for ignoring or abandoning the circle principle are wrong as a matter of policy and economics.

4 Kostis Hatzitaskos, Nicholas Hill & Brad T. Howells, Aetna-Humana and Algorithmic Market Definition in the Guidelines, ANTITRUST SOURCE (Oct. 2017), https://www.americanbar.org/content/dam/aba/publishing/antitrust_source/oct17_hatzitaskos_10_16f.pdf [hereinafter Hatzitaskos et al.]. As stated in this article, Dr. Hatzitaskos and his co-authors consulted for the DOJ in Aetna/Humana.
5 Id. at 3.
Roles of Market Definition in Antitrust Analysis of Horizontal Mergers

The 2010 Guidelines continue to rely on the hypothetical monopolist test as the primary mechanism for “evaluat[ing] whether groups of products in candidate markets are sufficiently broad to constitute relevant antitrust markets.”6 The Guidelines note that, depending on the specific details of individual mergers, it is possible that (1) the hypothetical monopolist test may exclude from the relevant market certain products that are at least partial substitutes for the products of the merging parties and (2) the hypothetical monopolist test may be satisfied by multiple candidate markets.7 In an effort to provide further clarity, the Guidelines provide two points of further guidance:8

If the market includes a second product, the Agencies will normally also include a third product if that third product is a closer substitute for the first product than is the second product . . . when the Agencies rely on market shares and concentration, they usually do so in the smallest relevant market satisfying the hypothetical monopolist test.

There are compelling reasons, based on policy and economic analysis, for continued adherence to the circle principle in market definition.9

The 2010 Guidelines continue to rely on the hypothetical monopolist test as the primary mechanism for “evaluat[ing] whether groups of products in candidate markets are sufficiently broad to constitute relevant antitrust markets.”6 The Guidelines note that, depending on the specific details of individual mergers, it is possible that (1) the hypothetical monopolist test may exclude from the relevant market certain products that are at least partial substitutes for the products of the merging parties and (2) the hypothetical monopolist test may be satisfied by multiple candidate markets.7 In an effort to provide further clarity, the Guidelines provide two points of further guidance:8

If the market includes a second product, the Agencies will normally also include a third product if that third product is a closer substitute for the first product than is the second product . . . when the Agencies rely on market shares and concentration, they usually do so in the smallest relevant market satisfying the hypothetical monopolist test.

The first point articulates a version of the circle principle. We refer to the second point as the “smallest market principle.”

The prior version of the Guidelines laid out a specific algorithmic process for implementing the hypothetical monopolist test. In that process, products are to be added to a candidate market in order of “next best substitute” until the candidate market satisfies the test.9 Such a process, as a conceptual template rather than as a specific recipe for doing the analysis, automatically gave results that adhered to both the circle and the smallest market principles.10

While providing guidance that indicates continued adherence to the principles underlying this algorithmic process, the 2010 Guidelines no longer articulate this specific algorithmic process. This omission has sparked new debate about the proper method for analyzing relevant markets. For example, in the Aetna/Humana case, the Court implied that, because the algorithmic process is no longer explicit, the Guidelines must also endorse the notion that in some circumstances it can be appropriate to analyze a relevant market that excludes one (or more than one) product that is possibly a closer substitute to the reference product than are some of the other products included in the market.11

A possible reasonable explanation for this abandonment of the algorithmic process might be that such a process seems impractically mechanical and demands granular quantitative measurements of substitutability. Valid application of the antitrust laws might founder if antitrust agency decision-makers or court fact-finders were to read the Guidelines as requiring that cases against horizontal mergers carry the burden of proving their market definitions based on a precise application of the full algorithmic process. However, even if the principle cannot be applied with perfect algorithmic precision by practitioners working with limited data, the 2010 Guidelines do not and should not abandon the economic principle that relevant markets should include a product if it is a closer substitute to the reference product than some of the other products included in the market.

---

6 2010 Guidelines, supra note 1, § 4.1.1.
7 Id. § 4.1.
8 Id. § 4.1.1.
9 1992 Guidelines, supra note 2, §1.1 & n.9.
10 While the prior version of the Guidelines described the general process for implementing the hypothetical monopolist test, they did not, for example, prescribe specific empirical methods for determining which alternative would be the “next best substitute.” Id. §1.1.
To understand why this principle is still required for valid market definition, it is essential first to articulate the underlying roles of market definition in antitrust analysis of horizontal mergers. There are three:

**Role 1:** to establish the universe of sources of competition for the merging parties that will be subjected to more searching analyses via various methods, including the parties' documents, historical experience, and industry analysts' views;

**Role 2:** to establish the universe of suppliers that will underlie measures of market shares and change in concentration due to the merger; and

**Role 3:** to establish the universe of products, geographic regions, and rival firms that will be analyzed in the next step of competitive effects analysis utilizing methods such as upward pricing pressure (UPP) and merger simulation.

For each of these roles, abandoning a market that is consistent with the circle principle in favor of one that is more consistent with "intuitive" or "natural" groupings would lead to potentially key distortions in the ensuing assessments of the effects of a horizontal merger.

**Abandoning the Circle Principle Leads to Distorted Competitive Effects**

To see why, consider a hypothetical merger between two manufacturers of sports cars: SC1 and SC2. In our hypothetical, there are three brands of sports cars (SC1, SC2, and SC3) with substantially equal sales. However, the three brands of sports cars differ in their appeal to consumers. SC2 and SC3 are highly substitutable for each other (e.g., both are luxury sports cars), SC2 and SC3 are both weak substitutes for the stripped-down sports car (SC1), and numerous non-sports cars (i.e., sedans) are stronger substitutes for SC1 than are the luxury sports cars SC2 and SC3.

Intuition and perhaps even qualitative and documentary evidence might indicate that the most sensible delineation of the relevant market in which to analyze the effects of the merger of SC1 and SC2 consists of all the sports cars. Indeed, a hypothetical monopolist would find it profitable to raise the average prices of the sports cars since SC2 and SC3 are close substitutes for each other, and SC1 is somewhat substitutable too. Thus, the sports car market passes the hypothetical monopolist test, and the merger of SC1 and SC2 would be presumptively anticompetitive since their sales total approximately 66 percent of the market, with a delta HHI of 2222 and a premerger HHI of 3333.

However, because the proposed sports car market in this example violates the circle and the smallest market principles, this analysis is incorrect. The merger would be competitively benign, and certainly should not be viewed as presumptively anticompetitive, because the independent SC3 would thoroughly discipline the price of SC2 after the merger, and the numerous non-sports cars would discipline SC1. Adherence to the circle principle would place the numerous non-sports cars into the relevant market because they are closer substitutes for SC1 than are SC2 and SC3, and that alone would change the conclusion from an analysis of concentration within the delineated relevant market.

Furthermore, the market made up of all cars would not be consistent with the smallest market principle. After all, SC1 and non-sports cars would together constitute a relevant market without the luxury sports cars, and SC2 and SC3 would together constitute a relevant market of luxury sports cars. It is important to recognize that in neither of these relevant markets is the merger between SC1 and SC2 a presumptively anticompetitive horizontal merger, since the parties' products are not together in any properly delineated relevant market under the circle and the smallest market principles. Hence, it is clear from this example how critical adherence to these principles of market definition can be to the reasonable reliability of antitrust assessments of horizontal mergers that make use of the analytic tools of relevant markets and concentration.
Our hypothetical raises a related issue: how should one aggregate the market shares of the non-sports car manufacturers? The aggregation issue arises most importantly in the context of the circle principle, but also implicates the smallest market principle as it interacts with the circle principle.

Suppose there are many small car manufacturers with differentiated brands, all very substitutable for each other and, in aggregate, substitutable for the stripped-down sports car brand, SC\text{1}. One the one hand, if one starts the algorithmic process with the stripped down SC\text{1}, the closest substitute may be one of the luxury sports cars, because each of the non-sports car brands, treated separately, is small. The next addition to the putative market would be the other luxury sports car, and now the hypothetical monopolist would raise the prices of the luxury sports cars, including merging party SC\text{2}, and so this market of all the sports cars would satisfy the hypothetical monopolist test, without any non-sports cars included. Furthermore, it would satisfy the circle principle, again due to how small each non-sports car brand is. On the other hand, if one were to aggregate sedans to account for their collective competitive constraint on SC\text{1}, one would arrive at a market of SC\text{1} and sedans, thus yielding very different conclusions about the merger.

Moreover, a candidate market that starts with SC\text{2} would yield a market of SC\text{2} and SC\text{3}. This is where the circle principle and the smallest market principle interact: a putative market of all sports cars is too big, and application of the smallest market principle tosses out SC\text{1} (when starting with SC\text{2}) and leaves only the luxury sports cars in a market that passes the hypothetical monopolist test, is consistent with the smallest market and the circle principles, and does not include both merging parties. The logical process of applying the market definition principles shows the importance of considering aggregation, which here leads to the correct relevant market of SC\text{1} and sedans, on the one hand, and SC\text{2} and SC\text{3}, on the other hand. In this case, the proposed merger of SC\text{1} and SC\text{2} is not horizontal in the relevant markets.

In part to avoid some of the challenges with market definition, the 2010 Guidelines favor certain quantitative measures for analyzing the competitive effects of differentiated products mergers that do not rely directly on market definition. The key method relies on measures of upward pricing pressure that take as inputs diversion ratios between and margins of the merging parties’ products. In many cases, a standard assumption is that diversion ratios can be calculated from shares (as the destination product’s share divided by one minus the origin product’s share). However, to the extent shares reflect the above discussed defects in the definition of the market, such measures of diversion ratios will be distorted as well. Of course, more sophisticated measures of diversion ratios that account more directly for the closeness of substitution between different products and thus do not rely on shares would be immune to such concerns. But often, data directly measuring consumers’ substitution preferences are unavailable, making it equally hard to properly define a market and analyze competitive effects.

12 This issue also arose in the Aetna/Humana merger. See Aetna, 240 F. Supp. 3d at 49–55.
13 While the hypothetical monopolist over all sports cars could profitably impose a small but significant and non-transitory increase in the prices (SSNIP) of SC\text{2} and SC\text{3}, it could not profitably raise the price of SC\text{1} because SC\text{1} faces competition from sedans. Thus, including SC\text{1} in the relevant market violates the smallest market principle.
14 The issue of aggregation is complicated and aggregating too many brands may itself run afoul of the smallest market principle. The logic is that inclusion in the putative relevant market of all of the non-sports cars may be unnecessary to yield significant market power to the hypothetical monopolist. Note that even if one aggregates the non-sports car manufacturers for market definition purposes, one should not aggregate them for concentration and competitive effects analyses.
15 2010 Guidelines, supra note 1, § 6.1.
Response to Arguments to Diminish or Abandon the Circle Principle

Given the importance of adhering to the circle and smallest market principles to avoid distorting the analysis of competitive effects, it is mistaken to conclude that the current Guidelines’ abandonment of the algorithmic process of market definition is equivalent to the abandonment of the circle principle. While it is true that the algorithmic process does incorporate the circle principle, it would be a logical error to conclude that by abandoning the algorithmic process for market definition, the 2010 Guidelines similarly support abandoning the economic principles underlying the process.

In a recent article that assesses the impact of the 2010 Guidelines on the antitrust analysis of relevant markets, Hatzitaskos et al. seem to suggest that by abandoning the algorithmic process for market definition, the 2010 Guidelines open the door to ignoring the circle and smallest market principles for what they term “reasonable exceptions.” They do not explain what analysis should be relied upon to identify the circumstances making reasonable exceptions appropriate. They assert that the 2010 Guidelines “better reflect how product markets are defined in practice.” They write that the 2010 revisions provide the agencies and courts the ability to “focus on testing candidate markets that constitute intuitive, natural groupings of products that are based upon documentary and other qualitative evidence and can be tested using available data.” They conclude that the 2010 Guidelines endorse the notion that there are “reasonable exceptions” to the algorithmic approach and that “following the algorithmic approach can…lead to counterintuitive markets that look gerrymandered . . . particularly when data on closeness of substitution are incomplete.” There are a number of significant flaws in this analysis.

First, Hatzitaskos et al. suggest that both the algorithmic process of market definition and the circle principle require that there must be a single market definition for a particular proposed merger. The authors then note that because the 2010 Guidelines recognize that there may be multiple valid relevant markets for analysis of a horizontal merger, the 2010 Guidelines must therefore reject the circle principle. Specifically, Hatzitaskos et al. state that the Guidelines’ acceptance of multiple relevant markets is “consistent” with the existence of “reasonable exceptions to such an algorithmic approach.”

The prior version of the Guidelines did not provide that there must be a single valid market definition for a particular proposed merger. Nor does the use of the algorithmic process for the delineation of a relevant market dictate that there be a single valid market definition. The Guidelines state that, starting with any one (and eventually each one) of each of the two parties’ products, the algorithmic process should proceed. A different relevant market could be the outcome of the algorithmic process using each one of the merging parties’ products as a different starting point.

16 Hatzitaskos et al., supra note 4, at 3.
17 Id. at 2.
18 Id. at 2.
19 Id. at 3.
20 Id.
21 Specifically, Hatzitaskos et al. state that the Guidelines’ acceptance of multiple relevant markets is “consistent” with the existence of “reasonable exceptions to such an algorithmic approach.” Id.
Even the smallest market principle does not lead to a single relevant market, since it does not reject the larger of two incompletely overlapping collections of products or geography. It could be the case that starting with Party A’s product, and working through a formal or informal process of adding substitutes until the hypothetical monopolist test is satisfied, this relevant market includes Party B’s product but does not exhibit a concerning increase in concentration due to the merger, while starting with Party B’s product leads to a different relevant market that does exhibit a concerning increase in concentration due to the merger. Each of these two relevant markets would satisfy the circle and the smallest market principles, and yet be different from each other, incompletely overlapping, and suggest different measurements of concentration. Thus, it is incorrect to suggest that one has to abandon the circle principle to allow for the possibility of multiple valid relevant markets for the assessment of a particular horizontal merger.

Second, Hatzitaskos et al. mistakenly argue that the practical reasons for the 2010 Guidelines to back away from the requirement that markets be defined algorithmically are also reasons for diminishing or abandoning the circle principle. As the 2010 Guidelines explicitly acknowledge, starting with generally articulated industry usage or with Brown Shoe factors is often a more practical initiation into market definition than employment of the algorithmic process. While using such less formal information can circumvent the need to establish a highly detailed matrix of cross-elasticities of demand from all-too-scarce data for precise scientific measurement, it does not stop the circle principle from being a possibly critical feature of market delineation. This is articulated in the 2010 Guidelines: “the Agencies will normally also include a third product if that third product is a closer substitute for the first product than is the second product.” In the hypothetical car example illustrated earlier in this article, the “first product” would be SC₁ (stripped down sports car), “the second product” would be SC₂ (luxury sports car) and the “third product” would be one of the sedans, which should have been included in the defined relevant market, but may not be, if the circle principle is not followed.

Example 6 in the 2010 Guidelines clarifies this idea further, concluding that the third product will “normally be included even though Products A and B together satisfy the hypothetical monopolist test.” Consider again our hypothetical merger of sports car manufacturers. In that example, defining a market following industry segmentation and commonly understood labels would omit non-sports cars from the market definition even though consideration of actual substitution behavior showed that non-sports cars (particularly sedans) were more substitutable for one of the parties’ overlapping products than some of the other products with the pertinent “sports car” label. In such cases, the circle principle assures that non-sports cars are included, so as to avert gerrymandering and misrepresentation of the set of products that are truly relevant for competitive assessment of the merger.

For these reasons, it is flawed as a matter of policy and economic analysis to ignore or abandon the circle principle for “reasonable exceptions” that cannot be identified in a disciplined analytic fashion. Indeed, as we show above, the continued adherence to the circle principle in market definition helps to achieve analytically correct relevant market definition.

24 2010 Guidelines, supra note 1, § 4.1.1.
25 Id.
26 If this stated substitution pattern is truly measurable and is the way as stated, then it is possible that the competitive effects analysis would utilize this factor and still reach the correct conclusion. However, if such data are unavailable, defining relevant markets without the circle principle, especially when there is no disciplined approach to identifying “reasonable exceptions” would distort the analysis.
Book Review
Competition Policy and Collusion

Joseph E. Harrington, Jr.
The Theory of Collusion and Competition Policy
The MIT Press, 2017

Reviewed by Wentong Zheng

Joseph Harrington’s new book, The Theory of Collusion and Competition Policy, surveys economic research on competition policy and collusion. It addresses two broad categories of questions: First, what is the impact of competition law and enforcement on whether firms collude, how long they collude, and how much they collude? Second, what is the optimal design of competition law and enforcement?1 While the focus of the book is theoretical and most of the research surveyed in the book does not concern how courts and enforcement authorities decide antitrust cases involving collusion, the findings presented are of potential interest to economists, judges, and legal scholars who want to better understand the impact and optimal design of competition policy. Professor Harrington’s book also reveals gaps in the existing economic literature and points to areas for future research.

Harrington has had a long and distinguished career in competition law and economics. He is currently the Patrick T. Harker Professor of Business Economics and Public Policy at the Wharton School of Business, University of Pennsylvania. He is the author of more than 75 articles and two textbooks. He is an expert on collusion and cartels, and his research has been presented before numerous competition authorities around the world.

This book is his latest work on collusion. The intended audience for the book is “economics scholars who seek to contribute to our understanding of how unlawful collusion operates and the proper design of competition law and enforcement . . . .”2 Nearly all of the research reviewed is mathematically oriented; a good understanding of basic economic modeling techniques is therefore necessary to fully appreciate the depth and breadth of the book. People with no mathematical background, however, can still benefit from the book, not the least because Harrington does an excellent job paraphrasing what the economic models mean in plain English.

After beginning the book with a brief description of the different notions of collusion employed in economics and in law, as well as the basic modeling techniques used in the book, Harrington presents a number of insights from existing economic research on how competition law and enforcement affect collusive behavior. Represented by the probability of conviction and the magnitude of the penalties imposed upon conviction, competition policy is shown to have robust effects on the operations of cartels. The effects could be direct, such as when active cartels are

---

2 Id. preface.
shut down. Competition policy could also have indirect deterrence effects when it deters cartels from forming. It could also have indirect price effects when the cartels that are not shut down or deterred are forced to set lower collusive prices. All in all, Harrington demonstrates that economic research has shown that competition policy has the desired effects of causing cartels to form in fewer markets and with shorter durations, reducing the size of the largest stable cartel, and increasing the size of the smallest stable cartel.

With respect to pricing, Harrington shows that competition policy is predicted to lead to less dispersion of collusive markups as it reduces the frequency of low markups and high markups. The effects of competition policy on collusive prices are uncertain, however, depending on whether the equilibrium condition at the current price is binding—that is, whether the cartel would become unstable or unprofitable at a higher price. Finally, if the probability of detection is sensitive to the price path, then the presence of competition policy is shown to cause firms to raise prices gradually upon cartel formation, as opposed to significant one-time price increases.

Harrington then discusses the optimal design of competition policy in light of its impact on collusive behavior. If the decision to prosecute a cartel is made on the basis of observed prices and quantities, then optimal enforcement requires the prosecutor to focus on collusion at high price levels and to minimize enforcement costs. In the event of prosecution and conviction, the socially optimal penalty should be based on the amount of overcharge multiplied by the quantity that would have been sold at the competitive price. This prescription differs from how damages are assessed in private actions, where the actual quantity, instead of the but-for quantity, is used.

Harrington also presents findings from economic research on optimal design of competition policy that takes into account strategic considerations on the part of customers and firms. For example, when customers are aware of a cartel and make purchases in anticipation of the possibility of collecting damages, the higher demand from customers induces the cartel to price higher, making the customers no better off than if there were no damages. For another example, because a standard method for estimating damages uses the post-cartel price as the estimate for the but-for price, this induces firms to price above the competitive price in the post-cartel period in an effort to lower estimated damages.

The policy issue for which theoretical research has been most extensive and for which Harrington provides most coverage in his book is the effectiveness value of leniency programs. A leniency program creates incentives for individual members of a cartel to apply for leniency out of concerns that the prosecutor will detect the cartel. It also creates incentives for individual members of a cartel to apply for leniency before other members do. These two incentives reinforce each other in inducing firms to apply for leniency. In terms of its impact on cartel stability, a leniency program has three effects. It could make a cartel less stable because it increases the payoff from cheating for individual members of the cartel by reducing the expected amount of penalty. But it could also make a cartel more stable because a sufficiently generous leniency program could make every member of the cartel apply and, as a result, reduce the amount of expected penalty for the entire cartel. Yet another effect of leniency programs is to cause firms to “race to the courthouse” to apply for leniency, which increases expected penalties and lowers the payoff from collusion. Generally, the destabilizing effects outweigh the stabilizing effects, and collusion becomes more difficult as leniency programs get more generous.

Regarding the optimal design of leniency programs, the economic research surveyed in the book largely recommends granting maximal leniency to the first leniency applicant before the commencement of an investigation, allowing firms to apply for leniency even after an investigation has already started, and not limiting eligibility for leniency for ringleaders and recidivists.
As Harrington points out, there is a fundamental mismatch between the kinds of collusion studied in the economic literature and the kinds of collusion deemed unlawful in law. In economics, collusion is “when firms in a market coordinate their behavior for the purpose of producing a supra-competitive outcome.” This coordination could be achieved by “implicit or explicit understanding among firms that ties future rewards and punishments to current behavior,” without the firms communicating with one another. In law, however, an unlawful agreement is not just a mutual understanding among firms to restrain competition, but the achievement of that mutual understanding must be through some form of communication. Harrington describes three categories of collusion: (1) explicit collusion, reached through explicit exchange of assurances of common actions; (2) tacit collusion, achieved by one firm announcing its plan to raise prices at an industry gathering, followed by parallel behavior indicating mutual beliefs that all firms will raise their prices; and (3) conscious parallelism. He observes that the first two categories of collusion are recognized in law as unlawful because they involve communication, while the third category is not considered unlawful because there is no communication.

Harrington’s view of the distinction between lawful and unlawful collusion represents the mainstream view. But it is not the only possible view. After all, conscious parallelism also involves communication—communication through actions. An alternative view of the legal standard for ascertaining unlawful collusion is that the law treats a particular form of collusion as lawful not because there is no communication, but because the communication at issue serves efficiency purposes. Yet another view is that the law ascertains unlawful collusion based on whether the communication at issue allows the alleged coconspirators to have constructive knowledge of one another’s conspiratorial intent. As an indication of the confusing state of the matter, Harrington

3 For discussions of the leniency programs in the United States, the European Union, and member states of the EU, see Wouter P.J. Wils, Leniency in Antitrust Enforcement: Theory and Practice, 30 World Competition 25 (2007).
4 HARRINGTON, supra note 1, at 1.
5 Id. at 2.
6 Id.; see also In re High Fructose Corn Syrup Antitrust Litig., 295 F.3d 651, 654 (7th Cir. 2002) (“[I]t is generally believed . . . that an express, manifest agreement, and thus an agreement involving actual, verbalized communication, must be proved in order for a price-fixing conspiracy to be actionable under the Sherman Act.”).
7 An example of such explicit collusion is found in United States v. Socony Vacuum Oil Co., 310 U.S. 150 (1940), where the defendant oil producers entered into an “informal gentlemen’s agreement or understanding whereby each undertook to perform his share of the joint undertaking” to effectuate an increase in spot and retail gasoline prices. Id. at 179.
8 As noted below, courts usually refer to this kind of illegal collusion as “tacit agreement,” not “tacit collusion.” See infra note 13. An example of such illegal collusion is found in United States v. Foley, 598 F.2d 1323 (4th Cir. 1979), where all the defendants announced their plans to raise prices regardless of what others would do, and then raised their prices. The court in Foley held that this sequence of events was sufficient for the jury to find a conspiracy among the defendants. Id. at 1331.
9 HARRINGTON, supra note 1, at 6.
10 Louis Kaplow also believes that the law penalizes certain forms of interfirm communications, although he is critical of the communications requirement. See Louis Kaplow, An Economic Approach to Price Fixing, 77 Antitrust L.J. 343, 344 (2011).
11 See William H. Page, Tacit Agreement Under Section 1 of the Sherman Act, 81 Antitrust L.J. 593, 614 (“[T]acit agreement means interdependent actions coordinated by prior communications of competitive intentions, in circumstances in which the communications do not provide any benefit to other audiences, especially consumers.”).
him self uses the wrong terminologies in describing the different categories of collusion: Courts indeed refer to conscious parallelism as tacit collusion, and refer to what he calls tacit collusion as tacit agreement.\textsuperscript{13}

Despite the confusion, Harrington makes an important point: the economic literature overly focuses on noncompetitive outcomes, but not the process through which the noncompetitive outcomes are reached. In specifying what unlawful collusion is, “almost all economic models do not draw the distinction that exists in the law with regard to different categories of collusion. The approach is instead to focus on a particular class of equilibria and presume that acting according to them is unlawful.”\textsuperscript{14} As a result, very few economic studies address the question of when competition policy induces firms to choose collusion with less effective yet lawful means of communication.\textsuperscript{15}

Another gap identified by Harrington is the lack of economic research on how to define liability and evidentiary standards.\textsuperscript{16} Questions relevant to liability and evidentiary standards include: Should only explicit collusion be illegal? Should “economic evidence” based on market data be sufficient to prove unlawful collusion, or must there be evidence of communication?\textsuperscript{17} Harrington laments that “to my knowledge there is no formal analysis that seeks to determine the socially optimal definitions of liability and evidentiary standards.”\textsuperscript{18} As a result, he has to leave this important issue out of his book because of the lack of work in this area.

The questions ignored by economic research, however, are the very questions for which the courts need the most guidance. For instance, courts have identified a number of “plus factors” that tend to exclude the possibility of independent actions and suggest collusive behavior,\textsuperscript{19} but have struggled with assigning probative values to the plus factors.\textsuperscript{20} In addition, courts have imposed different evidentiary standards in different procedural stages of lawsuits against unlawful collusion.\textsuperscript{21} How do these evidentiary standards impact social welfare? Unfortunately, the existing economic research sheds no lights on these important questions.

In the final chapter of the book, Harrington identifies some important areas for future research. One glaring omission from the current literature is studies on corporate managers’ incentives to collude. As Harrington points out, “The presumption in most research is that a manager acts in the best interests of shareholders by taking full account of the additional profits delivered by forming

\textsuperscript{13} See Page, supra note 11, at 595.
\textsuperscript{14} HARRINGTON, supra note 1, at 13.
\textsuperscript{15} Id. at 49–50.
\textsuperscript{16} Id. at 61.
\textsuperscript{18} HARRINGTON, supra note 1, at 61.
\textsuperscript{19} Examples of such plus factors include: proof that the parallel conduct is contrary to each defendant’s apparent self-interest, proof of defendants’ motives for entering into the alleged conspiracy, artificial standardization of products, and price increases during times of low demand.
\textsuperscript{21} In Theatre Enterprises, Inc. v. Paramount Film Distributing Corp., 346 U.S. 537 (1954), the Supreme Court held that evidence of parallel conduct alone does not compel a finding of an unlawful agreement. In Matsushita Electrical Industrial Corp., Ltd. v. Zenith Radio Corp., 475 U.S. 574 (1986), the Court held that evidence of parallel conduct alone would not survive a defendant’s motion for summary judgment or for a directed verdict. Id. at 588. Most recently, in Bell Atlantic Corp. v. Twombly, 550 U.S. 544 (2007), the Court held that an allegation of parallel business conduct and a bare assertion of conspiracy alone would not survive a defendant’s motion to dismiss for failure to state a claim under Section 1 of the Sherman Act. Id. at 556.
a cartel as well as the possible corporate penalties that may be imposed.”

22 But because of agency issues, managers’ interests do not always coincide with the interests of shareholders. They may collude too little, or too much, from the point of view of shareholders. Since managers are the individuals who are actually colluding, research on the impact of their personal or career concerns on collusion is desperately needed to gain a full understanding of companies’ collusive behavior.

On the opposite side of corporate managers, antitrust enforcement officials are also motivated by personal and career concerns in carrying out their duties. Harrington observes that “[p]resumably, members of the [competition authority] are . . . interested in maximizing their perceived performance while taking into account any personal costs.”

23 Such personal and career concerns may impact antitrust enforcement in several ways. First, antitrust enforcement officials may believe that their income and future job prospects “are most impacted by how many cartels are prosecuted, how many cases are won, and how much in fines is collected.”

24 While Harrington does not specify the exact mechanisms, these effects could be caused by enforcement officials’ incentives to increase enforcement activities to build up their human capital that would make them more attractive to prospective employers upon leaving the government. Or they could be caused by enforcement officials’ incentives to increase enforcement activities to expand the market demand for their expertise in defending enforcement actions upon leaving the government.

25 Second, enforcement officials may “put more weight on shutting down cartels than on deterring cartels from forming—as the former is more observable than the latter—and may avoid difficult cases (though the pursuit of them could enhance deterrence).”

26 Additional research is needed to understand how enforcement officials’ personal and career concerns may make enforcement activities depart from the socially optimal levels and how better rules could lead to better welfare outcomes.

Harrington concludes his book by circling back to the largest gap in the current economic literature on collusion: the lack of formal modeling that takes account of different liability regimes and evidentiary standards. Harrington observes that “[m]odels need to factor in various legal regimes and how illegality depends on the manner in which firms collude and not just on whether they collude.”

28 In a somewhat daunting call for action, Harrington states: “The integration of liability and evidentiary standards in economic models of collusion is a vast unexplored territory. Until we venture into it, a fundamental question in the area of competition law and enforcement will remain open: What is the socially optimal definition of unlawful collusion?”

---

22 HARRINGTON, supra note 1, at 107.
23 Id. at 108.
24 Id.
26 Id. at 1281.
27 HARRINGTON, supra note 1, at 108.
28 Id. at 109.
29 Id. at 110.