A Simple Model with an Endogenous Incumbent Firm and Disruptive Entrant

*Louis Kaplow**

June 10, 2026

Preliminary Draft

Abstract

Many industrial organization models take as given the existence and capabilities of incumbent dominant firms and sometimes also disruptive entrants. This article analyzes a simple but general model in which these traits are endogenous. In basic, plausible settings, marginal changes in variables produce shifts between equilibria, generating discontinuous welfare effects that sometimes have signs opposite to those arising in a given equilibrium, which is implicitly the domain examined in much existing work. Higher investment costs or marginal operating costs for either firm, or more intense competition between them, can readily reduce welfare. Nonconvexities in firms' investment decisions and strategic interactions between the firms create discontinuities, some of which involve nonmonotonicity. The model's simplicity makes transparent robust forces associated with a variety of actions by firms and also with constraints that may be imposed by competition regulation. The present formulation can be integrated with many models of particular practices because the model's reduced-form representations of investment costs, investment recoveries, and consumer surplus are sufficient statistics for behavior and welfare.

*Harvard University and National Bureau of Economic Research. I am grateful to _____ for helpful discussions and comments and to Harvard University, including its John M. Olin Center for Law, Economics, and Business, for financial support. Disclaimer: I sometimes serve as a background consultant on antitrust matters, including for a plaintiff and two defendants in monopolization challenges; my spouse is a lawyer who has mostly represented financial services firms.

1. Introduction

Much industrial organization research and overlapping literatures on competition policy take firms' existence and capabilities as given. Such formulations are particularly common with respect to a dominant incumbent firm and sometimes are employed as well for a prominent rival, such as a competing platform (see the survey by Jullien, Pavan, and Rysman (2021)). It is, of course, widely understood that firms' existence and capabilities are a product of their prior investments, the incentives for which depend on what investment recoveries they anticipate. Moreover, it is appreciated that the prevailing legal regime, including competition regulation, may assist or constrain firms' behavior in ways that influence expected investment recoveries. Thus, both an incumbent firm's and prospective entrant's existence and capabilities are endogenous.

Many literatures nevertheless proceed as if all of the fruits of investment are given, which is not a modest assumption in light of Corrado et al.'s (2022) estimate that intangible investment alone is a sixth of GDP (more than $5 trillion annually). Whether in basic microeconomics, Tirole's (1988) industrial organization text, or surveys of sophisticated literatures (e.g., Varian (1989) on price discrimination), the model and associated diagrams often represent a monopolist with a given marginal cost and demand curve. The decisions that are analyzed are as simple as determination of the monopoly price but range across the design of price discrimination, strategic interactions, and more. These decisions, for the stipulated cost and demand curves, determine consumer and producer surplus, and also the magnitude of deadweight loss against a benchmark of price equal to marginal cost. In similar spirit, the welfare effects of contemplated legal constraints are measured by how these surplus measures would change.

A basic but often neglected point is that this notion of producer surplus, often called profits or rents, is not surplus or profit or rent in a key, economic sense. Instead, it constitutes investment recovery. For example, if producer surplus is barely sufficient to cover the investment cost, then producer surplus, profit, and rent in the ex ante sense each equal zero, making no contribution to social welfare. And if the investment recovery is slightly smaller—and that lower level is anticipated—the investment is not made, the firm never comes into existence, and consumer surplus vanishes. Total welfare equals zero, notwithstanding the shaded areas in standard diagrams that depict a situation that never arises.

In this article's simple model with discrete investment, marginal changes in parameters that affect investment recoveries (and often consumer surplus as well) will at some point shift from one equilibrium to another, which typically involves a discontinuous jump in welfare. Moreover, the sign of that jump is sometimes opposite to that of

marginal welfare effects from changing the same parameter when one takes as given being in one of the equilibria. That is, in the most basic settings, welfare as a function of many parameters is discontinuous and may also be nonmonotonic.

These observations are elementary and, at some level, well known. Yet they are often in the back of our minds or forgotten altogether when building, analyzing, and drawing lessons from sophisticated models of behavior that also determine welfare within the assumptions of the model at hand.

This article follows a different, complementary path. It presents and analyzes perhaps the simplest model one can construct that includes the basic elements of dominance and possible disruption. First, a prospective dominant incumbent firm arrives and decides whether to make an investment, followed by some period of operations wherein it earns a monopoly investment recovery if it did invest. Second, an entrant arrives and decides whether to invest, followed by the incumbent firm (if it exists) deciding whether to make a continuation investment. Finally, there is a further operations period in which one or the other firm serves the market as a monopolist, the two firms compete in some fashion, or the market is not served.

The model's simplicity makes transparent some core, robust forces. The modeling approach employs reduced forms for investment costs, investment recoveries, and consumer surplus. These representations are consistent with most standard models of investment, cost functions, demand curves, and competitive interaction. Making each reduced-form term a function of a supplemental parameter enables comparative statics for the two firms' strategic choices—particular actions in the investment or operations periods—and also most contemplated regulations that might constrain those choices (such as restrictions on predatory pricing, tying, exclusive dealing, or remedies that provide rivals access or require interoperability).

Closely related, the approach here is complementary to that in many other literatures because the results derived elsewhere—combined, as appropriate, with empirical estimates—can be inserted into this article's model. The terms for investment costs and investment recoveries are sufficient statistics for firms' behavior, and firms' strategic choices (which determine their true profits) plus the consumer surplus terms are sufficient statistics for welfare.

Many of the results may seem surprising or counterintuitive, but on reflection they are directly attributable to familiar ideas. It will be shown, for example, that higher fixed or marginal operating costs for either firm can reduce welfare, and that more intense competition when the two firms operate together can reduce welfare. These results are produced by shifts between equilibria: because investment recoveries are incentives for prior investments, anything that affects them can shift the equilibrium. Moreover, firms' strategic investment decisions are interdependent. There can be preemption (even

without any strategic behavior normally associated with this idea), and the anticipation of the other firm's investment decision feeds back on earlier investment choices. The handful of active channels in this model is sufficient to give rise to the aforementioned results and many others.

The analysis proceeds as follows. Section 2 states the model and characterizes each of the six possible equilibria, including with respect to welfare.

Section 3 presents the main analysis. First, it shows how equilibrium is determined and demonstrates that each of the six equilibria is feasible. Second, it performs comparative statics by introducing a simple parameterization. Marginal effects on welfare are described for some simple benchmarks, after which shifts between pairs of equilibria are explored. Of the fifteen possible pairings of the six equilibria, it is proved that there exist boundaries between all but two of the pairs. In the thirteen pairs of interest, there exist marginal changes in the basic parameters that shift the equilibrium between the two in the pair. In general (except in knife-edge cases), the change in welfare is discrete. Moreover, in several cases of interest, these discrete welfare changes are opposite in sign from that of the marginal welfare effects from changing the same parameter in the same direction in the neighborhood of the boundary.

Most of the cases considered, including those with discontinuous jumps associated with opposite-signed welfare effects, can plausibly arise in basic settings. Indeed, because the model is so simple and the assumptions so few, the relevant domain for the different cases is large and readily covers important cases of interest.

Section 4 explores a number of economic interpretations and applications of the basic model, including implications of different intensities of competition, preemption of other firms' investments, and investment spillovers. It then introduces extensions for multi-dimensional and continuous investment decisions, repeated games in which additional firms arrive on an ongoing basis, and uncertainty. Together, the breadth and generality are substantial and, correspondingly, the restrictions are few.

Section 5 concludes by relating the paper's model to the many literatures in industrial organization and to overlapping literatures addressed to competition policy. The discussion elaborates how existing, sophisticated work on particular behaviors and possible regulation thereof (e.g., price discrimination or platform competition) can be integrated into the present framework. It also explains how previously identified welfare effects can be misleading for two reasons. First, the failure to endogenize firms' investment decisions can readily reverse the sign. Second, even in work that takes investment to be endogenous, determinate results that have been advanced are often the product of (sometimes implicit) assumptions defining special cases, whereas under the more general formulation offered here, welfare effects are readily ambiguous in comparable settings and thus can plausibly be of opposite sign.

2. Model

The model has two periods, each with an investment and an operations subperiod. In the first period's investment subperiod, $t = 1a$, a prospective incumbent, firm I, decides whether to make an investment to come into existence. In the operations subperiod, $t = 1b$, if firm I invested, it operates as a monopolist; otherwise, demand is not served. In the second period's investment subperiod, $t = 2a$, a prospective entrant, firm E, decides whether to make an investment to come into existence. Next, firm I, if and only if it invested in the first period, decides whether to make a further investment to stay in existence; if it does not, it exits. In the operations subperiod, $t = 2b$, depending on which firms exist, one or the other operates as a monopolist, the two firms compete, or demand is not served.

$t = 1$: Incumbent I arrives
- $t = 1a$: I decides whether to invest
- $t = 1b$: I operates as a monopolist if it invested; else, nothing happens

$t = 2$: Entrant E arrives
- $t = 2a$: E decides whether to invest, after which
  - I, if and only if it invested at $t = 1a$, decides whether to make a continuation investment
- $t = 2b$: If only I exists, it operates as a monopolist
  - If only E exists, it operates as a monopolist
  - If both I and E exist, the two firms compete
  - If neither I nor E exist, nothing happens

Firm f's investment costs at time t are $F_f^t > 0$, where superscripts refer to time periods (here, always the a subperiods) and subscripts to firms. If firm f operates as a monopolist in period t (here, always the b subperiods), it receives an investment recovery of $\pi_f^t > 0$. The investment recovery earned by firm f at $t = 2b$ when both firms exist is indicated by $\pi_{f,IE}^{2b} > 0$, where the subscript refers first to the firm and then to denote the case in which both operate, and it is assumed that the competitive recovery for each firm is strictly below what its own $t = 2b$ monopoly recovery would be.[1] For each investment

[1] The assumption that $\pi_{f,IE}^{2b} > 0$ is for ease of exposition. If competition is sufficiently intense and one of the firms has sufficiently higher marginal operating costs or lower demand for its product, its investment recovery could be zero (if negative, it would cease operations), but since this term can be close to zero and

recovery there is corresponding consumer surplus, denoted CS_f^t when firm f operates as a monopolist in period t and CS_{IE}^{2b} when the two firms compete at $t = 2b$. Finally, the model abstracts from uncertainty, investment decisions are observable, all information is common knowledge, and time discounting is implicit (costs and benefits are expressed as of a common date).

These reduced-form expressions for investment recoveries and consumer surplus can be taken to represent a wide range of possible cost functions, demand functions, and models of competitive interaction. For example, if firm f has lower operating costs, that will typically be associated with higher investment recoveries for that firm when it operates, a lower investment recovery for the other firm if it also operates, and higher consumer surplus under both monopoly and competition. Similarly, greater demand typically is associated with higher investment recoveries for each firm in all cases in which it operates and greater consumer surplus, under both monopoly and competition. (Because the reduced forms can take different values in the two operations subperiods, demand could be growing, perhaps due to the development of complements, or falling, perhaps because of the development of substitutes. Likewise for operating costs, which could rise due to increasing scarcity of inputs or fall because of learning or scientific advances.) Finally, at $t = 2b$, for the case denoted IE in which both firms operate, the firms might compete intensely, collude, or anywhere in between; broadly, more intense competition typically is associated with lower investment recoveries for each firm and greater consumer surplus.

As will be discussed, these reduced forms for investment recoveries are sufficient to determine each of the firm's investment decisions at $t = 1a$ and $t = 2a$ and thus the resulting equilibrium. The equilibrium indicates which terms for consumer surplus are relevant, allowing us to compute social welfare.

It is helpful to introduce indicator variables $\mathbb{I}_f^t$ that take a value of 1 if firm f invests at time t and 0 otherwise, which decision is determined by its anticipated profits as viewed from the pertinent a subperiod in which the investment choice is made.

$$\mathbb{I}_f^t = \begin{cases} 1, & \pi_f^t > 0 \\ 0, & \pi_f^t \leq 0 \end{cases}$$

Indifference is resolved against investment.

For ease of exposition when performing comparative statics below, there will be some abuse of notation under which the derivative of an indicator variable with respect to any parameter should be interpreted as the distributional derivative when taken at the indifference point, $\pi_f^t = 0$, where the indicator function is discontinuous. The value of the

---

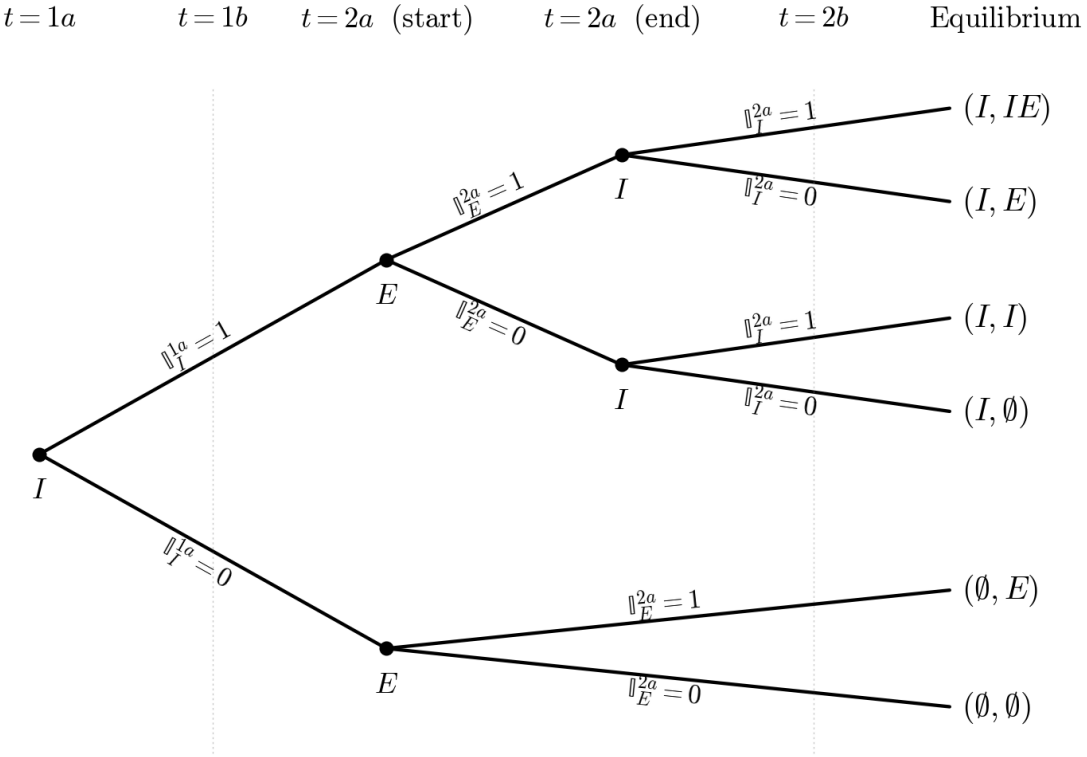
there is a fixed investment cost incurred at $t = 2a$ before firm E comes into existence or firm I decides to continue, no case of interest is removed from consideration.

distributional derivative at that point is the Dirac delta function (or the negative of that) evaluated at 0. (By the definition of the distributional derivative, this will be so even if the indifference convention would imply a value of zero had a one-sided derivative been taken on the negative side of 0, where there would not have been a discontinuity.)[2]

We can further describe the equilibria of this game using the notation $\boldsymbol{e} \in \mathcal{E} = \{(\emptyset, \emptyset), (\emptyset, E), (I, \emptyset), (I, I), (I, E), (I, IE)\}$, where each $\boldsymbol{e}$ is an ordered pair indicating whether, in each of the b subperiods in the corresponding equilibrium, we have no firm (denoted $\emptyset$), just firm I, just firm E, or both firms in operation (denoted IE). Figure 1 depicts the foregoing extensive-form game, with equilibria indicated at the right. (The payoffs associated with each equilibrium appear in Table 1, below.)

[2] The distributional derivative of an indicator function with respect to a variable equals the Dirac delta function of the difference between the value of the variable at which the derivative is taken and the variable's value at the indifference point. Ordinarily, at the discontinuity, the value of the distributional derivative of a function is given by the delta function times the magnitude of the jump, but with an indicator variable, the value of the jump simply equals one. In the comparative statics analysis below, there is some abuse of language to avoid otherwise-elaborate, repetitive statements of the precise exercise being undertaken. (As the text here indicates, such will be relevant when taking a distributional derivative at a boundary condition in the direction where the interior of the interval includes the boundary rather than in the opposite direction, where the boundary is excluded. That is, the subtleties are relevant when $\pi_f^t = 0$ at the initial condition and the marginal change in a variable reduces rather than increases profits. Had the indifference convention been reversed, then the opposite marginal change would introduce this complication.)

Figure 1. Equilibria of Extensive-Form Game



Profits in the a subperiods (as distinguished from reduced forms for investment recoveries in the b subperiods) can now be stated in fairly condensed form for all possible equilibria. Proceeding in accord with backward induction, start with firm I's continuation investment decision at the end of $t = 2a$, which will be relevant only when $\mathbb{I}_I^{1a} = 1$, that is, when firm I invested at $t = 1a$. This continuation investment will be undertaken if

$$\pi_I^{2a} = \ 1 - \mathbb{I}_E^{2a} \ \pi_I^{2b} + \mathbb{I}_E^{2a} \pi_{I,IE}^{2b} - F_I^{2a} > 0 \qquad (1)$$

If firm I invests to continue in existence at $t = 2a$, it earns its monopoly investment recovery π_I^{2b} at $t = 2b$ if firm E did not invest at $t = 2a$ but only its competitive investment recovery $\pi_{I,IE}^{2b}$ if firm E did invest. In either case, it incurs its continuation investment cost of F_I^{2a}. Given our timeline, firm I knows whether firm E invested at the beginning of $t = 2a$ when it subsequently makes this continuation investment decision at the end of that subperiod.

At the beginning of $t = 2a$, firm E invests to come into existence if

$$\pi_E^{2a} = \ 1 - \mathbb{I}_I^{1a} \mathbb{I}_I^{2a} \ \pi_E^{2b} + \mathbb{I}_I^{1a} \mathbb{I}_I^{2a} \pi_{E,IE}^{2b} - F_E^{2a} > 0 \qquad (2)$$

Firm E earns its monopoly investment recovery π_E^{2b} at $t = 2b$ either if firm I never invested at $t = 1a$ to come into existence or, if firm I did invest then, it does not make its continuation investment at $t = 2a$. But firm E receives only its competitive investment recovery $\pi_{E,IE}^{2b}$ if firm I invests both initially and to continue. In either case, it incurs its

investment cost of F_E^{2a} of coming into existence. Importantly, at the time of its decision at the beginning of $t = 2a$, firm E knows whether firm I invested at $t = 1a$ and can rationally anticipate whether, if it did, it will make its continuation investment at the end of $t = 2a$ (which decision, note, from inequality (1), may itself depend on whether firm E invests).

At $t = 1a$, firm I invests to come into existence if

$$\pi_I^{1a} = \pi_I^{1b} + \mathbb{I}_I^{2a}\pi_I^{2a} - F_I^{1a} > 0 \qquad (3)$$

Firm I earns its monopoly investment recovery π_I^{1b} at $t = 1b$ plus, if and only if it incurs its continuation investment cost at the end of $t = 2a$, its continuation profit of π_I^{2a}. And it incurs its investment cost of F_I^{1a} of coming into existence. At $t = 1a$, firm I rationally anticipates whether firm E will invest at the outset of $t = 2a$ (which decision is implicit in π_I^{2a} and depends on whether firm I invests at $t = 1a$) and whether it will make its own continuation investment at the end of $t = 2a$.

Table 1 displays the reduced-form parameters associated with each of the model's six possible equilibria. Rows indicate different equilibria and columns indicate time periods. Cells in the first column indicate the features of the pertinent equilibrium; the use of indentations in the entries mimics the structure of the corresponding game tree for the extensive form. Cells in the next four columns contain all effects pertinent to firms' investment decisions and to social welfare: investment costs, investment recoveries, and consumer surplus.

Table 1. Investment Costs and Recoveries and Consumer Surplus in each Equilibrium

	$t = 1a$	$t = 1b$	$t = 2a$	$t = 2b$
$\mathbb{I}_I^{1a} = 0$: I not invest at $t = 1a$				
$\mathbb{I}_E^{2a} = 0$: *E not invest at* $t = 2a$	0	0	0	0
$\mathbb{I}_E^{2a} = 1$: *E invests at* $t = 2a$	0	0	$-F_E^{2a}$	$\pi_E^{2b} + CS_E^{2b}$
$\mathbb{I}_I^{1a} = 1$: I invests at $t = 1a$				
$\mathbb{I}_E^{2a} = 0$: *E not invest at* $t = 2a$				
$\mathbb{I}_I^{2a} = 0$: I exits at $t = 2a$	$-F_I^{1a}$	$\pi_I^{1b} + CS_I^{1b}$	0	0
$\mathbb{I}_I^{2a} = 1$: I invests at $t = 2a$	$-F_I^{1a}$	$\pi_I^{1b} + CS_I^{1b}$	$-F_I^{2a}$	$\pi_I^{2b} + CS_I^{2b}$
$\mathbb{I}_E^{2a} = 1$: *E invests at* $t = 2a$				
$\mathbb{I}_I^{2a} = 0$: I exits at $t = 2a$	$-F_I^{1a}$	$\pi_I^{1b} + CS_I^{1b}$	$-F_E^{2a}$	$\pi_E^{2b} + CS_E^{2b}$
$\mathbb{I}_I^{2a} = 1$: I invests at $t = 2a$	$-F_I^{1a}$	$\pi_I^{1b} + CS_I^{1b}$	$-F_E^{2a} - F_I^{2a}$	$\pi_{I,IE}^{2b} + \pi_{E,IE}^{2b} + CS_{IE}^{2b}$

To elaborate, examine the row one up from the bottom, where firm I invests at $t = 1a$, firm E invests at the outset of $t = 2a$, and, upon observing this choice, firm I exits at the end of $t = 2a$. In that case, each firm's investment costs of coming into existence are incurred in the respective investment subperiods (but not firm I's cost of continuation at $t = 2a$); firm I earns a monopoly investment recovery at $t = 1b$ (and no further recovery); and firm E earns a monopoly investment recovery at $t = 2b$. Finally, for each of the periods with investment recovery, we also have the associated consumer surplus. As mentioned above, demand and firms' cost functions are implicit, in reduced form, in the expressions for investment recoveries and consumer surplus. (Higher demand or lower costs will tend to raise investment recoveries and consumer surplus.) Similarly, the nature and resulting intensity of competition when both firm I and firm E operate at $t = 2b$ is implicit, in reduced form, in the equilibrium depicted in the bottom row.

Table 1 also indicates the level of social welfare, W^e, associated with any equilibrium e; it equals the sum of all of the entries in the associated row. For example, for the row one up from the bottom that was just discussed, we have:

$$W^{I,E} = \left(\pi_I^{1b} - F_I^{1a} + CS_I^{1b} \right) + \left(\pi_E^{2b} - F_E^{2a} + CS_E^{2b} \right)$$

For convenience, the terms in this expression are rearranged. Inside the first pair of large parentheses (corresponding to $t = 1$), we have firm I's profits (its investment recovery minus its investment cost) plus the consumer surplus its operation generates, and in the second pair (for $t = 2$), the same for firm E. Because firms invest to come into existence only if their anticipated profits are positive, we know that each of the terms for a firm's profits are positive, as is the associated consumer surplus. In that sense, each firm's existence contributes to social welfare, but (as will be emphasized below) that observation is correct only relative to a benchmark in which neither firm ever comes into existence, yielding $W^{\emptyset,\emptyset} = 0$. As will be explained, if the counterfactual is instead, for example, an equilibrium in which one of the two firms never invests but the other one does (which will often be the more relevant comparison), welfare in that equilibrium is also strictly positive, and it can be higher than $W^{I,E}$ or welfare in another equilibrium of relevance. More precisely, comparative statics analysis in section 3.2 will focus on how marginal changes in parameters may cause shifts between particular pairs of equilibria, so the welfare effect at the boundary will be the difference between welfare in those equilibria, evaluated where the shift occurs.

As a final observation on terminology, note that it is common to refer to the value of terms like π_I^{1b} and π_E^{2b} as monopoly profits (or producer surplus), but that is only apt in a world in which the $F_f^t = 0$—the world often implicit in analyses that take one or both firms' existence as given. But when investment decisions and thus existence are endogenous, the terms π_I^{1b} and π_E^{2b} (and the others for the b subperiods) are properly defined as investment recoveries. True profits, as defined above, are net of investment

costs. For example, in the $\{I, E\}$ equilibrium under discussion here, the firms' profits are given by the corresponding terms in the inner parentheses.

3. Analysis

3.1. Equilibrium Determination

To determine which of the six equilibria in the set $\mathcal{E}$ will arise as a function of the model's parameters, it is useful to restate expressions (1) – (3) for firms' investment decisions more explicitly. Seven inequalities are relevant; they are presented in an order that follows backward induction.

Beginning with expression (1) for firm I's investment decision at the end of $t = 2a$ (which is relevant only when $\mathbb{I}_I^{1a} = 1$, that is, firm I invested at $t = 1a$), we have two possibilities regarding firm E's investment decision at the beginning of $t = 2a$. If $\mathbb{I}_E^{2a} = 0$, firm I invests if and only if

$$\pi_I^{2b} > F_I^{2a} \qquad (4)$$

However, if $\mathbb{I}_E^{2a} = 1$, firm I invests if and only if

$$\pi_{I,IE}^{2b} > F_I^{2a} \qquad (5)$$

Turn next to expression (2) for firm E's investment decision at the beginning of $t = 2a$. Here, we have three possibilities to consider. For the first two: If either $\mathbb{I}_I^{1a} = 0$ or if we have $\mathbb{I}_I^{1a} = 1$ but inequality (5) does not hold (conditional on firm E's investing, firm I would not make its continuation investment), then firm E knows it will be a monopolist at $t = 2b$ if it invests, so it does so if and only if

$$\pi_E^{2b} > F_E^{2a} \qquad (6)$$

However, if $\mathbb{I}_I^{1a} = 1$ and inequality (5) holds (conditional on firm E's investing, firm I would make its continuation investment), then firm E knows it will compete with firm I at $t = 2b$ if it invests, so it does so if and only if

$$\pi_{E,IE}^{2b} > F_E^{2a} \qquad (7)$$

Finally, turn to expression (3) for firm I's investment decision at $t = 1a$. Here, we have five possibilities to consider for what firm I will anticipate subsequently. First, suppose that expression (6) fails to hold (firm E will not invest even if it expects to be a monopolist at $t = 2b$) and, similarly, inequality (4) fails to hold (firm I itself will not invest to continue even though it would be a monopolist at $t = 2b$). Then, firm I invests at $t = 1a$ if and only if

$$\pi_I^{1b} > F_I^{1a} \qquad (8)$$

Second, suppose that inequality (6) does hold (firm E invests if it expects to be a monopolist at $t = 2b$) and, moreover, that inequality (5) does not hold (firm E anticipates that, if it invests, firm I will not make its continuation investment even if firm I invested at $t = 1a$). Then, once again, firm I invests at $t = 1a$ if and only if inequality (8) holds.

Third, suppose that inequality (5) does not hold and inequality (6) also does not hold (firm E does not invest even though it anticipates operating as a monopolist at $t = 2b$ if it did), and, moreover, that inequality (4) holds (firm I invests knowing it will be a monopolist at $t = 2b$, if firm I invested at $t = 1a$). Then, firm I invests at $t = 1a$ if and only if

$$\pi_I^{1b} + \pi_I^{2b} > F_I^{1a} + F_I^{2a} \qquad (9)$$

Fourth, suppose that inequality (5) holds (firm I makes its continuation investment at $t = 2a$ even though it anticipates competing with firm E at $t = 2b$) and that inequality (7) fails (firm E does not invest at $t = 2a$, anticipating having to compete with firm I at $t = 2b$). Then, once again, firm I invests at $t = 1a$ if and only if inequality (9) holds.

Fifth, suppose that inequality (7) holds (firm E invests at $t = 2a$ when it anticipates having to compete with firm I at $t = 2b$) and inequality (5) holds (firm I will make its continuation investment when it anticipates competing with firm E at $t = 2b$). Then, firm I invests at $t = 1a$ if and only if

$$\pi_I^{1b} + \pi_{I,IE}^{2b} > F_I^{1a} + F_I^{2a} \qquad (10)$$

We can now state:

Proposition 1. Every $\boldsymbol{e} \in \mathcal{E}$ *is feasible.*

Proof: We can pick feasible values for the investment cost and investment recovery terms such that each equilibrium may arise. Let $\pi_I^{1b} = \pi_I^{2b} = \pi_E^{2b} = 100$ and $\pi_{I,IE}^{2b} = \pi_{E,IE}^{2b} = 25$. Next, let each of F_I^{1a}, F_E^{2a}, and F_I^{2a} take either a high ("H") value of 200 or a low ("L") value of 20. Triples like HHH indicate the values of the three investment cost terms in the order of the timeline, in this case all being H.

Checking inequalities (4) – (10), it is straightforward to verify that:

If HHH, then the equilibrium is $\emptyset, \emptyset$.
If HLH, then the equilibrium is $\emptyset, E$.
If LHH, then the equilibrium is $I, \emptyset$.
If LHL, then the equilibrium is I, I .
If LLH, then the equilibrium is I, E .

If LLL, then the equilibrium is I, IE .[3] ■

Consider the next to last row in the proof (the case used previously to illustrate results). The first L implies that firm I will invest at $t = 1a$ regardless. (Inequality (8) holds.) The second L implies that firm E will invest at the beginning of $t = 2a$ regardless. (Inequalities (6) and (7) both hold.) The final H implies that firm I will not make its continuation investment at the end of $t = 2a$ regardless. (Inequalities (4) and (5) both fail.)

Observe that, although it is straightforward to identify feasible cases for each equilibrium to hold, it is quite intricate (and not that illuminating) to specify the complete conditions that fully characterize each equilibrium. Note that our seven binary conditions (inequalities (4) – (10)) allow for 128 permutations, although it can be demonstrated that not all are material. More concretely, this complexity is attributable to the interdependence of the three investment decisions, as reflected both implicitly and explicitly in inequalities (1) – (3), wherein a given inequality depends on indicator variables that are determined by the others—a core feature, not a complicating bug, of the strategic environment that this model represents.[4] Moreover, these seven inequalities collectively contain eight parameters (three investment cost terms and five investment recovery terms), which means that we are defining regions in an eight-dimensional space. It is nevertheless possible to provide significant further understanding through fairly straightforward methods.

3.2. Comparative Statics

To better understand how different equilibria may arise and the welfare properties of particular equilibria and of shifts between equilibria, we will next restate the terms for investment costs and investment recoveries—and also those for consumer surplus—as functions of particular parameters. Only a handful will be introduced, each corresponding to important phenomena associated with incumbent firms and entrants.

The corresponding terms for investment costs, investment recoveries, or consumer surplus, as the case may be, are assumed to be continuously differentiable functions of any

[3] For the curious reader, HLL and HHL are moot (they do not add equilibria) because the first H tells us that firm I will not invest at $t = 1a$, regardless, so it does not have the opportunity to make (and thus will not make) a continuation investment at $t = 2a$ even if the investment cost would have been L.

[4] Viewing this aspect another way, each of the six possible equilibria may have a boundary with each of the five others. (Proposition 3 establishes that two of the possible pairs of equilibria have no boundary but all of the other pairs do have a boundary.) In addition, a particular pair of equilibria can have more than one boundary condition, as defined below (which the proof of Proposition 3 demonstrates for one of the pairs).

parameter introduced. This assumption includes the possibility that one or more terms may be constant in the variable, it only affecting other terms. Starting with investment costs, the F_f^t, let each be a function of a cost-shifter variable, ϕ. This will allow several interpretations introduced below. For example, a higher ϕ may shift up all of the F_f^t, only those for one of the firms, only initial investment costs, or only the continuation investment cost. Furthermore, let F_E^{2a} also be a function of $\mathbb{I}_I^{1a}$, which can be used to represent the possibility of investment spillovers from firm I's investment at $t = 1a$. (Under this interpretation, F_E^{2a} is falling in $\mathbb{I}_I^{1a}$. To satisfy the continuous differentiability assumption, let F_E^{2a} fall smoothly in $\alpha\mathbb{I}_I^{1a}$, where we can vary $\alpha \in [0,1]$, as introduced in section 4.1.3.)

For the investment recovery and consumer surplus terms, π_f^t and CS_f^t, respectively, let us introduce the variable κ that can (potentially) affect both, such as moving firm-time pairs in the same direction.[5] For example, if a higher κ indicates higher marginal operating costs (not explicitly modeled here) or lower demand (also implicit) for a given firm configuration in a given b subperiod, we would expect the associated investment recoveries and consumer surplus to be lower. And conversely when κ is lower. For most models of firm pricing behavior, cost functions, and demand, this will tend to be so both for the monopoly cases and when there is competition (short of extremes, such as price equal to constant marginal cost, where investment recoveries are zero regardless). These are not the only possibilities. Taking an example that will be used below, a higher κ might indicate a longer $t = 1b$ and a correspondingly shorter $t = 2b$, such as when the entrant arrives later. That would tend to raise the investment recovery and consumer surplus in the earlier period and lower those in the later period. Yet another interpretation is that, when one of the firms operates as a monopolist in some period, a higher κ might indicate an enhanced ability to price discriminate (to identify consumer valuations and prevent arbitrage), which would raise the investment recovery and have an ambiguous effect on consumer surplus.

Finally, for the two investment recovery terms and consumer surplus term for $t = 2b$ in the I, IE equilibrium, it will be convenient to introduce an additional variable, γ. (From the foregoing, one could omit this addition and further interpret the variable κ, but it will make some of the exposition below more transparent to have this additional variable.) This variable will often be interpreted as the degree of competitive intensity, so a higher γ will be associated with lower investment recoveries for each of the two firms and higher consumer surplus. It can also indicate competitive skew, where a higher γ

[5] There is some abuse of notation regarding the subscript f: standing alone it indicates the pertinent monopoly case, but it is also used (abused) here as a shorthand that includes the corresponding f, IE expressions when both firms are present and compete.

means that the interaction is relatively more favorable to firm I than to firm E, which would raise (lower) the investment recovery of firm I (firm E) and have an ambiguous effect on consumer surplus—the latter because there are many models of competitive interaction, many ways to marginally change the skew, and, importantly, different starting points (skewing slightly in favor of firm I starting from a playing field highly tilted in favor of firm E may raise rather than lower symmetry).

Marginally changing any of these newly introduced variables has two qualitatively different types of effects that we will now explore. First, taking as given some $\boldsymbol{e} \in \mathcal{E}$ (that is, assuming that one is not at a boundary of $\boldsymbol{e}$ described by the pertinent inequalities), there will be a marginal effect on firms' profits, consumer surplus, and total welfare. Second, when the initial conditions fall on the boundary between two of the $\boldsymbol{e} \in \mathcal{E}$, a marginal change will shift the equilibrium, generating a discontinuous effect on profits of the other firm, if relevant (but not necessarily for the firm that was indifferent and changed its investment decision), on consumer surplus, and on total welfare.[6] Moreover, these first-order effects can have the opposite sign of marginal effects on either side of the neighborhood of the boundary point.

To further assess the marginal effects in a more concrete manner, consider the following special case:

Benchmark Parameterizations:

$\frac{dF_f^t}{d\phi} > 0$ *for all (pertinent)* f, t

$\frac{d\pi_f^t}{d\kappa} < 0$ *and* $\frac{dCS_f^t}{d\kappa} < 0$ *for all (pertinent)* f, t

$\frac{d\pi_{f,IE}^{2b}}{d\gamma} < 0$ *and* $\frac{dCS_{IE}^{2b}}{d\gamma} > 0$ *for all* f, *and* $\frac{dCS_{IE}^{2b}}{d\gamma} > \left|\frac{d\pi_{I,IE}^{2b}}{d\gamma} + \frac{d\pi_{E,IE}^{2b}}{d\gamma}\right|$

These benchmarks are consistent with the primary interpretations above. The final assumption regarding the variable γ states that the posited increase in competitive

[6] To explain the potentially mysterious parenthetical, consider the case in which firm E is indifferent and thus (given our convention) does not invest at the beginning of $t = 2a$ when it anticipates that firm I, which invested at $t = 1a$, will make its continuation investment at the end of $t = 2a$ even if firm E invests (expression (7) holds as an equality). When firm E is marginally induced to invest (expression (7) now holds as an inequality), that discretely reduces firm I's $t = 2b$ investment recovery from π_I^{2b} to $\pi_{I,IE}^{2b}$, which in some cases induces firm I to no longer invest at $t = 1a$ (inequality (9) holds but inequality (10) fails to hold). That further switch now means that firm E enjoys a monopoly investment recovery of π_E^{2b} rather than a competitive investment recovery of $\pi_{E,IE}^{2b}$ at $t = 2b$, a first-order increase (even though the initial change in its decision to invest was a matter of indifference, when it took as given that firm I invested at $t = 1a$).

intensity raises consumer surplus by more than the two firms' combined investment recoveries fall; i.e., there is a reduction in deadweight loss.

Proposition 2. For any $\boldsymbol{e} \in \mathcal{E}$ in which the initial conditions do not entail indifference for any pertinent investment decision (i.e., none of the pertinent expressions (1) – (3) holds as an equality), then under the benchmark parameterizations, marginal effects on social welfare are:

$$\begin{aligned}
\emptyset,\emptyset: &\quad dW^{\emptyset,\emptyset}/d\phi = 0,\ dW^{\emptyset,\emptyset}/d\kappa = 0,\ dW^{\emptyset,\emptyset}/d\gamma = 0 \\
\emptyset,E: &\quad dW^{\emptyset,E}/d\phi < 0,\ dW^{\emptyset,E}/d\kappa < 0,\ dW^{\emptyset,E}/d\gamma = 0 \\
I,\emptyset: &\quad dW^{I,\emptyset}/d\phi < 0,\ dW^{I,\emptyset}/d\kappa < 0,\ dW^{I,\emptyset}/d\gamma = 0 \\
I,I: &\quad dW^{I,I}/d\phi < 0,\ dW^{I,I}/d\kappa < 0,\ dW^{I,I}/d\gamma = 0 \\
I,E: &\quad dW^{I,E}/d\phi < 0,\ dW^{I,E}/d\kappa < 0,\ dW^{I,E}/d\gamma = 0 \\
I,IE: &\quad dW^{I,IE}/d\phi < 0,\ dW^{I,IE}/d\kappa < 0,\ dW^{I,IE}/d\gamma > 0
\end{aligned}$$

Proof: Straightforward from the definition of the benchmark parameterizations and the definition of W^e as given by sum of the terms in the corresponding row in Table 1. ■

The benchmark parameterizations and Proposition 2's results are in accord with standard intuitions associated with a wide range of models (but obviously not all models and functional forms): higher investment costs or marginal operating costs reduce welfare; lower demand reduces welfare; and more intense competition raises welfare. Proposition 2 takes as given a particular equilibrium (it applies only to marginal changes in the interior). Its results are stated primarily as a reference point for the analysis that comes next, where we will contrast the discontinuous welfare effects of marginal changes in the variables that change the equilibrium. Although Proposition 3 will not invoke these benchmark parameterizations as assumptions, the proof will use them in all but a couple exceptional cases in which another parameterization is required to demonstrate the result.

It is helpful to introduce a further definition:

Boundary Condition: An expression from among those stated in (4) – (10) is a boundary condition for a pair of possible equilibria $\boldsymbol{e}_i, \boldsymbol{e}_j \in \mathcal{E}$, $i \neq j$, if (i) there exists values of ϕ, κ, or γ such that the expression holds as an equality (rather than as the stated strict inequality), and (ii) for any neighborhood around such values for any pertinent variable, there exist points in that neighborhood such that each of the two possible equilibria in the pair constitutes the equilibrium.

To elaborate (and to implicitly define the notion of a pertinent variable), consider an example involving the pair of equilibria $\emptyset,\emptyset$ and $\emptyset,E$. Following the method in the

proof of Proposition 3 to follow, consider cases in which $F_I^t > \pi_I^{1b} + \pi_I^{2b}$ for $t = 1a, 2a$, so there is no question that firm I would not make the initial or continuation investment whatever else might transpire. Hence, only the two equilibria in the posited pair are feasible. Then, expression (6) for firm E's $t = 2a$ investment decision, holding as an equality, is the boundary condition for these two equilibria. The analysis will then focus on values of variables ϕ and κ, which are pertinent if $dF_E^{2a}/d\phi \neq 0$ or $d\pi_E^{2b}/d\kappa \neq 0$, respectively, when evaluated at the stated boundary. That is, we are only concerned with neighborhoods around the values of ϕ and κ such that expression (6) holds as an equality and marginally changing the variable breaks the equality. Regardless of the sign of these derivatives (and what indifference convention we had chosen), it is thus possible to marginally change such a variable, in one direction or the other, in such a manner that each of $\emptyset, \emptyset$ and $\emptyset, E$ will be the equilibrium.[7]

Changes of any variable across any such boundary thus shift the equilibrium. Welfare will jump discontinuously at the boundary (except for some knife-edge cases, in which the sum of the welfare-relevant terms for each of the two equilibria is precisely equal—a possibility that does not arise in the comparison under consideration). The economic significance is that, as variables change across nontrivial ranges, these jumps may dominate the overall welfare impact of that discrete change in the variable. In addition to these discontinuities, some of the equilibrium shifts involve a welfare jump that is of opposite sign to that of marginal changes in the neighborhood of the discontinuity. Hence, welfare as a function of a variable will not only be discontinuous but also may be nonmonotonic when boundaries are crossed.[8]

Proposition 3.

a. Except for the two cases listed in Proposition 3.b, for any distinct pair of equilibria $\boldsymbol{e_i}, \boldsymbol{e_j} \in \mathcal{E}$, $i \neq j$, *there exists a boundary condition.*[9]

b. No boundary condition exists for the pair $\{ \emptyset, \emptyset \ , \ I, IE \}$ *or the pair* $\{ I, \emptyset \ , \ I, IE \}$.

---

[7] There is a further subtlety regarding cases in which expressions (9) or (10) are boundary conditions because changing a single variable might move two like terms in the opposite direction by precisely the same amount. This is not possible with the benchmark parameterizations, and, in any event, the definition of boundary conditions focuses on variables for which the property holds in the associated neighborhood.

[8] Each of the terms in the boundary conditions are assumed to be continuously differentiable in all the variables, but clearly welfare is not differentiable at the boundary due to the discontinuity. Nevertheless, if we use the distributional derivative (discussed above, after the indicator variables were introduced), we can integrate the derivative over the range of the variable to state the total change in welfare.

[9] It is not claimed that the boundary condition is unique. Indeed, the proof for the pair of equilibria $\{ \emptyset, E \ , \ I, I \}$ demonstrates that there exist two qualitatively different boundary conditions.

c. For equilibrium pairs in Proposition 3.a,

(i) The effect on welfare of switching equilibria is first-order relative to marginal effects in either equilibrium in the pair in the neighborhood of the associated boundary, except in knife-edge cases in which welfare (evaluated at the values of the pertinent variables at the boundary point) is equal in the two equilibria.

(ii) For some of the pairs, the discrete effect on welfare from changing a variable in a manner that switches the equilibrium is of opposite sign to the marginal effect from changing the same variable in the same direction in the neighborhood of the boundary (excluding the boundary itself).

The proof for each of the fifteen possible equilibrium pairs appears in the Appendix. The above discussion of the definition of a boundary condition illustrates that analysis for the first case considered there. The focus is on identifying a boundary condition for each pair of equilibria or, for the two exceptions, showing that no boundary exists.

4. Interpretation and Extensions

4.1. Economic Interpretation of the Model's Parameters

4.1.1. Competition: γ. The variable γ was introduced into the two investment recovery terms and consumer surplus term for $t = 2b$ in the I, IE equilibrium. Focus on the interpretation of γ as indicating the intensity of competition.[10] Consider the range from a very low level—perhaps perfect coordination on the joint profit-maximizing prices—to a very high level—which might be Bertrand competition in price or, for some cases such as with differentiated products, even lower prices that might be imposed by a regulator. We will follow the benchmark parameterization by assuming that $d\pi^{2b}_{I,IE}/d\gamma < 0$, $d\pi^{2b}_{E,IE}/d\gamma < 0$, $dCS^{2b}_{IE}/d\gamma > 0$, and $\frac{dCS^{2b}_{IE}}{d\gamma} > \left|\frac{d\pi^{2b}_{I,IE}}{d\gamma} + \frac{d\pi^{2b}_{E,IE}}{d\gamma}\right|$.

If we take as given that both firms make the requisite investments to be in operation at $t = 2b$ so that we are in equilibrium I, IE at an interior point (i.e., none of the boundary conditions hold), we have the familiar result that more intense competition raises not only consumer welfare but also total welfare. As competitive intensity changes, however, a boundary may be crossed, switching the equilibrium. The three pertinent

[10] Additional, plausible cases of economic interest arise when γ indicates the degree of skew between the two firms. The supplemental implication is that changing the skew makes the more favored firm more likely to invest and the other less likely, as reflected in the boundary conditions associated with expressions (5), (7), and (10) holding as equalities.

possibilities (boundary conditions associated with expressions (5), (7), and (10)) are analyzed in the proof of Proposition 3, so the discussion here will be informal and intuitive. Much can be gleaned by considering the following sequence, which formally is feasible and also is plausible in that it depends only on a few assumptions that could readily hold in realistic settings. First, when competitive intensity is extremely weak, I, IE is the equilibrium because the high investment recoveries for each firm are sufficient to induce firm E to invest at the beginning of $t = 2a$ (even though it knows that firm I will make its continuation investment) and to lead firm I to make its continuation investment at the end of $t = 2a$ (even though it knows that firm E invested and thus will operate at $t = 2b$). Second, when competitive intensity is intermediate, firm E no longer invests at the beginning of $t = 2a$ because it knows that firm I will earn sufficient investment recovery if it did to nevertheless make its continuation investment at the end of $t = 2a$ (firm I having a lower continuation investment cost than firm E's investment cost of coming into existence) and that the resulting competitive investment recovery is now insufficient to cover firm E's investment cost. Third, if competition is very intense, neither firm makes much (or, in the extreme, any) investment recovery, which could give rise to a setting in which firm E invests at the beginning of $t = 2a$, knowing that this commitment renders firm I's continuation investment unprofitable at the end of $t = 2a$ (even though firm I has a fairly low continuation investment cost), so firm E can expect to operate as a monopolist at $t = 2b$. (There are, of course, other possibilities of interest.)

Note first that we have nonmonotonicity with respect to firm E's investment decision to come into existence. Weak or intense competition (a low or high γ) induce firm E to invest whereas intermediate intensity (medium levels of γ) does not. We also have (holding all other parameters constant) a situation in which firm E may enter and compete with firm I, where firm E's entry displaces the incumbent firm, and where firm E does not enter at all.

A further remark about these cases is that there is no close connection between how efficient firm E is in comparison to firm I and whether firm E finds it profitable to invest to come into existence. When competitive intensity is low, firm E may find it profitable to invest even though it is in every sense less efficient than is firm I, because its investment recovery is large enough to cover even substantial investment costs and, despite higher marginal operating costs, it expects to earn a large investment recovery. When competitive intensity is intermediate, firm E may find investment to be unprofitable even though it has lower marginal operating costs than does firm I, because investment recoveries are only modest, but still are large enough to cover firm I's low continuation investment cost. (Note, however, that one might deem firm I to be more efficient precisely on that account, although much discussion of firms' relative efficiency in this domain focuses on marginal operating costs, ignoring fixed costs.) Here, we might say

that firm I excludes firm E by its very decision to come into existence, which, moreover, firm I may be induced to do solely by its $t = 1b$ investment recovery, when that covers its $t = 1a$ investment cost (that is, inequality (8) holds). Finally, when competitive intensity is high, firm E may be less efficient than firm I is but, because firm E moves first in $t = 2a$, it preempts firm I, knocking a more efficient firm out of the market. In this case, it is firm E's very existence that excludes firm I. (Extensions below will consider cases in which firm I can upset this preemption strategy through its own, earlier preemptive investment.)

Note further that, in the case with intense competition, firm E may not only preempt firm I from making its continuation investment at $t = 2a$, but as a consequence further induce firm I not to come into existence at $t = 1a$, which happens when inequality (8) does not hold (that is, when firm I's investment recovery at $t = 1b$ is insufficient to cover its initial investment cost at $t = 1a$). In this case, the market is not served at all at $t = 1b$. In all, we can see that the fundamental forces that determine incentives for firms to invest and come into existence (or to continue in existence) are not well aligned with having the more efficient firm (or any firm) serving the market. A further irony is that more intense competition hurts rather than helps consumers at $t = 2b$ because its anticipation eliminates competition, leaving them in this case with firm E operating as a monopolist. (Under other assumptions, it would leave them with firm I as the monopolist.)

Finally, it is necessary to consider welfare explicitly because the more efficient firm winning is not a good proxy for welfare being higher. It should be obvious that firms' investment decisions often fail to maximize welfare, and that more intense competition in particular may reduce welfare once account is taken of its effects on investment incentives at $t = 2a$ and $t = 1a$. For example, when more intense competition (moving from a low to intermediate level) results in firm E not investing at $t = 2a$, we have that the prospect of higher competitive intensity results in monopoly. And, likewise, when very intense competition leads firm E to displace firm I. Hence, the prospect of more intense competition can readily reduce consumer surplus at $t = 2b$. And when firm I is induced not to invest at $t = 1a$, consumer surplus at $t = 1b$ is eliminated entirely.

Moreover, consumer surplus is not total welfare, which also includes firms' profits (that is, their investment recoveries net of their investment costs). This component of welfare involves further misalignment between firms' decisions and welfare. Even some of the simplest cases are at best ambiguous, which is familiar from Spence (1976), Dixit and Stiglitz (1977), and Mankiw and Whinston (1986), addressing whether incentives for a firm to enter a market may be socially excessive or insufficient. When intensity (or any other variable) changes marginally so that investment is just induced, the investing firm, being at its indifference point, earns no profits (ignoring further effects on the other firm's

investment decision). If the other firm invests and operates regardless, its investment recovery falls, reducing its profits (the business-stealing externality). That loss reduces total welfare, and this reduction may but need not exceed the rise in consumer surplus. In addition, we have seen that in basic cases the profit reduction leads the other firm not to invest, resulting in monopoly or a market not being served, in which event consumer surplus falls rather than rises. In all, the resulting equilibrium may readily fail to be the one that maximizes welfare, and whether the better firm operates at $t = 2b$ is not a good indicator of welfare being higher rather than lower as a result.

4.1.2. Investment costs: ϕ. The variable ϕ was introduced into the expressions for firms' investment costs, F_I^{1a}, F_I^{2a}, and F_E^{2a}. In our benchmark parameterization, we have $dF_I^{1a}/d\phi > 0$, $dF_I^{2a}/d\phi > 0$, and $dF_E^{2a}/d\phi > 0$. Hence, as Proposition 2 states, marginally increasing ϕ reduces welfare in any equilibrium in which at least one firm makes an investment. However, the proof of Proposition 3 makes clear that a variety of shifts between equilibria are possible, and it will be apparent that some will discontinuously boost welfare.

It eases exposition to focus on reductions in ϕ, that is, falls in investment costs, where marginal effects are favorable. If we begin in the equilibrium $\emptyset, \emptyset$, a marginal reduction in ϕ that brings at least one firm into existence raises welfare: the investing firm is indifferent and thus earns no profits either way, but consumer surplus is created. Similarly, if we begin at the boundary condition of $I, \emptyset$ that is associated with expression (4) holding as an equality and the switch is to I, I, there again would be no incremental profit (here, to firm I, which was indifferent) but consumer surplus is created at $t = 2b$.

But there are other plausible cases. For example, suppose we begin at equilibrium I, I, but are at the boundary condition for equilibrium $\emptyset, E$ associated with expression (6) holding as an equality.[11] If there is a marginal reduction in just firm E's investment cost, firm E's decision to invest, in anticipation, induces firm I not only to refrain from its continuation investment at $t = 2a$ but also its initial investment at $t = 1a$. This eliminates CS_I^{1b} and may raise or lower consumer surplus at $t = 2b$ because either CS_I^{2b} or CS_E^{2b} may be larger. Effects on profits are subtle but, following the analysis in the proof, the effect can be shown to be negative (firm E being at its indifference point does not profit from investing rather than abstaining, and firm E's investment induces firm I no longer to invest at all, whereas it had strictly positive profits when firm E abstained from investing and it thus invested in both of the a subperiods). Regardless of these latter points, it is easy to imagine cases in which the dominant welfare effect from a reduction in

[11] As the proof demonstrates, there exist two boundary conditions associated with this pair of equilibria; the discussion in the text examines the one in which firm E is indifferent at the associated boundary.

firm E's investment cost that switches the equilibrium from I, I to $\emptyset, E$ would be the loss of CS_I^{1b}, a large reduction in both consumer surplus and total welfare.

There are other cases of interest, but the broad lesson is clear. Presumptively beneficial reductions in investment costs can, in simple, plausible cases, switch between two equilibria, and in some cases that switch can significantly reduce welfare even though marginal effects taking as given remaining in any particular equilibrium are always favorable (if any firm operates).

4.1.3. Investment cost spillovers: $F_E^{2a}(\mathbb{I}_I^{1a})$. This parameterization (not used in establishing Propositions 2 and 3) constitutes one way to introduce investment spillovers, here from firm I making its $t = 1a$ investment to firm E enjoying lower $t = 2a$ investment costs.[12] To explore this phenomenon, consider a special case: $F_E^{2a} = (1 - \alpha \mathbb{I}_I^{1a}) F_I^{1a}$, where $\alpha \in [0,1]$. If firm I does not invest at $t = 1a$, there are no spillovers and, moreover, firm E's $t = 2a$ investment cost is the same as firm I's $t = 1a$ investment cost was. If firm I does invest at $t = 1a$, however, firm E's $t = 2a$ investment cost is lower by the fraction α, indicating perhaps the portion of the intangible assets firm I created that firm E can use for free. Most familiar, firm I may have developed and obtained approval for a new drug; if there is no patent protection or that protection (or exclusivity period) expires, firm E may not have to bear any of the investment costs associated with the drug's discovery and approval. But, much more broadly, firm I's very success may greatly reduce firm E's exploration costs to identify a bountiful market and an effective production and distribution strategy, all of which may be publicly visible. More subtly, firm E might copy or even steal some of firm I's methods or trade secrets. Or a regulator may require firm I to provide access to some of its assets (particularly intangible ones) to firm E for free or on favorable terms.

In all such cases, we have a familiar tradeoff from the patent literature that is readily seen in this model. Take as our baseline the case in which $\alpha = 0$ (complete appropriability) and consider marginal increases in α throughout its range. Taking the equilibrium as given (no equilibrium shifts), these increases are irrelevant if firm E does not invest at $t = 2a$, but they strictly increase welfare when firm E does invest because of the reduction in the investment cost it incurs, assuming that firm I invested at $t = 1a$.

A higher α also, at some point, can readily shift from one equilibrium to another. There are several possibilities. One is where the equilibrium shifts from I, I to I, IE. This shift raises consumer surplus at $t = 2b$ while lowering overall investment recovery

[12] Investment cost spillovers are just one possibility. Obviously, firm I's $t = 1a$ investment and its $t = 1b$ operations could generate spillovers that reduce firm E's $t = 2b$ operating costs, making it more likely to invest at $t = 2a$ or a tougher competitor at $t = 2b$ regardless, both of which have feedbacks on firm I's investment decisions that are similar to those examined in the text to follow. Investment cost spillovers can also be negative, such as when firm I obtains patents that raise firm E's investment cost.

(taking the two firms together) at $t = 2b$, with the former generally larger due to the reduction in deadweight loss. Note, however, that we again may have a reduction in total welfare because the shift means that F_E^{2a} is now incurred (albeit at a marginally lower level) rather than not. As discussed before, in plausible cases this latter effect is larger so total welfare falls.

But more momentous changes can happen. Most worrisome is when the equilibrium shifts from I, I to $\emptyset, \emptyset$, losing the entirety of consumer surplus in both of the b subperiods and also all of firm I's profits, which may have been strictly positive. It is straightforward how this shift can arise. As α increases, F_E^{2a} falls. Take one of the possibilities of interest: the boundary associated with expression (6) holding as an equality (firm E is indifferent whether to invest at $t = 2a$ when it expects to operate as a monopolist at $t = 2b$).[13] Suppose further that inequality (5) fails (when firm E invests at the beginning of $t = 2a$, that leads firm I to desist from making its continuation investment at the end of $t = 2a$). Finally, when this happens and inequality (8) also fails (firm I does not earn enough investment recovery at $t = 1b$ to cover its investment cost at $t = 1a$), then we have the posited shift between the two equilibria. Note further that, in addition to losing consumer surplus in both of the b subperiods and all of firm I's profits in the I, I equilibrium, no actual spillovers materialize. They only arise on a path that is not an equilibrium, but the anticipation thereof nevertheless has the investment suppression effect. This is a familiar idea associated with patents (particularly for pharmaceuticals), but the applications are much broader given the existence of large spillovers in many realms that are not governed by patent law or other effective protection of the originally investing firm's intangible assets (which, recall, include identifying a market opportunity and a general means of meeting it).

4.1.4. Investment recoveries and consumer surplus: κ. The variable κ was introduced in the reduced-form expressions for investment recoveries and consumer surplus to allow for a number of additional phenomena, particularly those associated with marginal operating costs and demand, each of which move associated investment recoveries and consumer surplus in the same direction in standard formulations. A basic variation is that lower operating costs for one firm tends to reduce the investment recovery for the other firm when the equilibrium associated with $t = 2b$ has both firms operating, that is, in equilibrium (I, IE).

Marginal changes in κ, which for behavior have their impact through investment recoveries, have similar effects to marginal changes in ϕ, which affect behavior through investment costs. Each of our investment conditions, inequalities (4) – (10), which also

[13] The other subcase of interest is when the relevant boundary is associated with expression (7) holding as an equality.

serve as boundary conditions between pairs of equilibria in the proof of Proposition 3, have investment recovery terms (only) on the left side and investment cost terms (only) on the right side. Hence, we can identify one-to-one relationships between various parameterizations and marginal changes in the variables that generate the same discrete effects at the same boundaries. The only difference is that, here, marginal changes in κ also directly affect consumer surplus in a given equilibrium—effects that, in our benchmark parameterization, have a marginal effect on welfare of the same sign as the effect through investment recoveries and thus profits. In all, the analysis is the same as before, with marginal effects within a given equilibrium of one sign and discontinuous effects from shifting between two equilibria that may be of the opposite sign. Hence, it is plausible that lower marginal operating costs or stronger demand reduce consumer and total welfare.

Outside our benchmark parameterization for κ, additional cases of interest arise. Suppose that firm f at time t is a monopolist, and let an increase in κ be associated with a rise in the ability to (profitably) price discriminate. That is, $d\pi_f^t/d\,\kappa > 0$ and the sign $dCS_f^t/d\,\kappa$ is ambiguous, so the sign of the effect on welfare is also ambiguous. See Varian (1989). This standard welfare characterization is, of course, incomplete once one accounts for endogenous investment decisions. The term π_f^t represents investment recovery, not profit, so it (without subtraction of the associated investment cost) is not welfare. (Nevertheless, a marginal change in π_f^t in a given equilibrium does constitute a marginal change in welfare.) More important, easier (or more permission for) price discrimination, by raising investment recovery in the associated equilibrium, may make that equilibrium more likely to arise due to the anticipatory effect on the incentive to invest. Taking the simplest case, consider the pair of equilibria $\emptyset,\emptyset$ and $I,\emptyset$. At the boundary condition associated with expression (8) holding as an equality, a marginal increase in the ability of firm I to price discriminate at $t = 1b$ induces it to invest at $t = 1a$. No matter how large the investment recovery π_I^{1b} (which is now marginally higher), firm I's profit equals zero because it is indifferent. By contrast, there is a discrete increase in both consumer surplus and total welfare of CS_I^{1b}, which can be quite large, regardless of whether, given that the equilibrium is $I,\emptyset$, the marginal effect on consumer surplus or total welfare would be positive or negative. Other cases could be examined, but the broad message is the same: standard welfare analysis of price discrimination, which takes the monopolist's existence and capabilities as given, is incomplete and can be quite misleading, specifically, of opposite sign and different order when investment is endogenous. Indeed, like in the discussion of investment spillovers, a limitation on price discrimination that would, in the given equilibrium, raise consumer surplus can instead extinguish consumer surplus entirely.

Consider also cases in which a change in the variable κ helps one firm and hurts the other. There are many possibilities, but for concreteness consider: $d\pi^{2b}_{f,IE}/d\,\kappa > 0$ and $d\pi^{2b}_{\sim f,IE}/d\,\kappa < 0$, where $\sim f$ denotes the other firm. Circumstances may become differentially more favorable to firm f for exogenous reasons or, as will be discussed below, because of some action of firm f or perhaps some regulatory restriction on some action of firm $\sim f$. The effects on total investment recovery, consumer surplus, and welfare at $t = 2b$ assuming that we are in the equilibrium I, IE are ambiguous without further specification, but it is helpful to have in mind a case in which consumer surplus and total welfare fall when there is a marginal increase in κ. From the prior analysis, this marginal effect may induce firm E not to invest at $t = 2a$, which reduces competition and welfare at $t = 2b$ relative to the posited counterfactual. We also know that welfare could nevertheless rise because firm E's $t = 2a$ investment cost is saved. But that is not the end of the story: the baseline in which both firms operate at $t = 2b$ may not have been an equilibrium because firm I may not have the requisite investment incentives. Hence, the posited $t = 2b$ welfare reduction in equilibrium I, IE may have been merely hypothetical, but this off-equilibrium change could induce investment, including by firm I at $t = 1a$, which could raise consumer surplus and welfare substantially. This is not the only possibility, but a plausible one in some settings. (Similar logic is applicable to extensions below, such as ones in which firm f has an investment choice that embeds the posited effect in the operations subperiod in its earlier investment decision; that makes it a credible commitment and thus something firm f might do even if the direct effect of its action, at the margin, during $t = 2b$, would be to reduce firm f's investment recovery in that period.)

4.2. Extensions

4.2.1. Investments. In the model, firms have only a discrete choice whether to invest to come into existence, and firm I at the end of $t = 2a$ has a further discrete choice whether to invest to continue in existence. Each of these decisions could be generalized to have firms choose $i^t_f \in \mathcal{I}^t_f$, where the choices are vectors with elements that may be continuous. Nonconvexities would still be allowed, including for the choice to come into or remain in existence. We would now have $F^t_f(i^t_f)$, $\pi^t_f(i^t_f)$, and $CS^t_f(i^t_f)$ (where, as above, with some abuse of notation, the latter two also include the associated IE equilibria at $t = 2b$ in which both firms operate). Firm f at time t chooses $i^t_f \in \mathcal{I}^t_f$ to maximize its profits (total investment recovery net of investment costs) in light of what has already occurred or it expects to happen, as applicable.

This extension makes the problem much more complex, but the basic forces remain. The primary difference is that, in addition to shifts between equilibria, changes in variables (those introduced in our parameterization) now may cause marginal investment adjustments by the firms. Many of these adjustments would be similar in character to the discrete jumps and thus sometimes would be of opposite sign to direct effects. Stronger demand tends not only to make the existence of a firm more likely but also to induce it to invest more if it does exist. A greater ability to price discriminate may induce greater investment in product quality. Larger spillovers (a higher α) will tend to reduce firm I's level of investment at $t = 1a$, even short of cessation. Hence, as we contemplate marginal changes in parameters that have various effects, the overall marginal effect, within a given equilibrium, may be magnified, attenuated, or reversed.[14] This observation broadens the central lesson: neither firms' existence nor their capabilities should be taken as given rather than understood as being endogenous to the environment and the setting of policy instruments.

A natural application is to consider how firm I's marginal investment choice at $t = 1a$ (in capacity or in innovation to reduce marginal operating costs) may preempt firm E at the beginning of $t = 2a$, and similarly how firm E's marginal investment choice at that time may preempt firm I's continuation investment at the end of $t = 2a$. The former is in the spirit of the models and analysis in Spence (1977, 1979), Dixit (1979, 1980), and Ordover and Willig (1981). The analysis is largely the same as that above, where it was observed that these firms' discrete investment choices have preemptive effects. Before, there need not have been a preemptive motive. When inequality (8) holds, firm I invests at $t = 1a$ even if its only investment recovery would be at $t = 1b$; nevertheless, that choice might (along with its anticipated continuation investment) induce firm E not to invest at the beginning of $t = 2a$.[15] Likewise, when inequality (7) holds, firm E invests at the beginning of $t = 2a$ even if firm I would have made its continuation investment at the end of $t = 2a$; nevertheless, that choice might induce firm I not to make that investment. Here, we add the possibility of marginal increases in investment that may induce the later firm to invest less or not at all, motivating a higher initial investment level than otherwise. The most important question, however, is one not directly addressed in the earlier literature: the effect on welfare. For example, Spence (1977)—if one takes the closest mapping to the model here—shows that, when firm E is preempted (it does not invest at all), firm I's $t = 1a$ investment level is socially excessive in the ex post sense that the welfare gain at $t = 2b$ (when firm I operates as a monopolist at a correspondingly low

[14] Relatedly, welfare now may be nonmonotonic in a variable even within a given equilibrium.

[15] As the prior analysis indicates, however, firm I's lack of a preemptive motive does not guarantee that firm I's $t = 1a$ investment raises welfare. This observation further supports the point that there is not a close nexus between traditional understandings of preemption and implications for welfare.

quantity) does not cover the increased investment cost. But the full welfare calculus also includes both the effects at $t = 2a$ and, importantly, the effects of deterring firm E's investment, which we have seen can have either sign.[16] (In many settings, there may be additional welfare effects from greater investment spillovers, both to the sector under examination, which may be moot if firm E never comes into existence, and to other sectors.) Hence, to analyze the full welfare effects of preemptive investment (or of regulations), one must employ the approach taken here.

More broadly, a firm might use investment in an a subperiod to embed the effects of an action it might otherwise wish to undertake in a later, b subperiod because investment involves commitment and because regulation may restrict the later action but not reach the earlier investment choice.[17] Again, we must keep in mind that the welfare effects of such actions are ambiguous at this level of abstraction.[18]

Additional interpretations and applications are available. Sticking with preemption, to the extent that firm I can make its continuation investment not at the end of $t = 2a$ but earlier, before firm E arrives, it may thereby preempt firm E in another way while also eliminating firm E's ability to preempt firm I through firm E's investment at the beginning of $t = 2a$, as the model's original timeline dictates. Also, some investments may change the duration of the time periods. In this regard, $t = 1b$ may be interpreted as the duration of firm I's first-mover advantage, but that duration may itself be endogenous to firm I's $t = 1a$ investment choices. Some investment choices may be harder for firm E to copy, emulate, or otherwise invent around. Firm I might create operations that reduce the flow of spillovers, delaying the point at which firm E's investment cost falls far enough to make investment profitable. Other possibilities can be included as well.

---

[16] Even focusing just on effects in the b subperiods taking preemption as effective, we need to keep in mind that positive effects on consumer surplus are external to the firm and thus underweighted in the preempting firm's profit calculus, even though its business-stealing from the preempted firm is overweighted relative to the social calculus.

[17] Regarding the latter, if a legal restriction on predatory pricing binds when price is below some measure of the firm's cost in the b subperiod (such as marginal or average variable cost), raising subperiod a investment may lower that threshold and thus relax this legal constraint. Similarly, under an "as-efficient competitor" test, if a firm's subperiod a investment makes it more efficient in the relevant legal sense in subperiod b, then more aggressive behavior is permitted at that time.

[18] Ordover and Saloner (1989), Kaplow and Shapiro (2007), and others have explained that legal tests for predation and analyses thereof have been largely disconnected from the underlying economics literature on the subject, although the emphasis of this criticism was on positive analysis (such as relating to credibility). Scheffman (1981) criticized Ordover and Willig's (1981) for lacking a conceptual foundation for their proposed test. Most relevant is subsequent work that applies some of these predation concepts and finds in simulations that restricting dominant firms' actions in accord with the stated principle has an ambiguous effect on welfare. See Besanko, Doraszelski, and Kryukov (2014, 2020). The analysis here supplies the underlying intuition for such results.

4.2.2. Repeated game. The model has two firms arriving in sequence, with everything ending at the close of the second period. It is appealing to embed this framework in a repeated game in which, over an indefinite period of time, firms arrive, explore, make investment decisions (including decisions to cease), and operate, with each firm rationally anticipating subsequent arrivals, where later firms will likewise make their own forward-looking, profit-maximizing decisions. Models in similar spirit include Hopenhayn (1992), Ericson and Pakes (1995), and Melitz (2003). This type of extension enables richer strategic interactions. For example, at the beginning of $t = 2a$, firm E contemplates not only whether firm I, if it came into existence at $t = 1a$, will make its continuation investment at the end of $t = 2a$, but also whether additional firms may arrive and invest in later periods. At that point, firm E is no longer the potential disruptive entrant but itself may be a dominant incumbent firm. It will also, therefore, worry about spillovers resulting from its own investment decisions and not just its ability to benefit from firm I's $t = 1a$ investment.

All firms also will contemplate ongoing investments, not just to remain in existence but to expand, improve or modify product offerings, change production processes and distribution methods, and so forth as a consequence of its own learning, changes in demand, and scientific advances (and spillovers from other firms, including firms in other sectors). Indeed, the most successful firms, which often have invested tens or hundreds of billions of dollars, started very small with a modest initial investment. Those investments created real options, the importance of which arises from uncertainty, considered momentarily. Positive realizations lead firms to make further investments, and much of the recovery from earlier investments are the recoveries that the now-enabled later investments will generate.

The simple model's core forces remain, and a benefit of the more elementary approach adopted here is to make those forces clear.[19] Relatedly, it is important not to lose sight of the analogue to $t = 1a$ in the present formulation, the beginning of time. Often, analysis focuses on the steady state (or stationary ergodic processes). But initial investments can be highly consequential in themselves, and as initial conditions they may influence the steady state. Moreover, periods are often best understood as years or decades long, with the sector perhaps becoming obsolete after a modest number of periods (rather than thousands). As a consequence, welfare effects of various actions and regulations can be quite different, both much better and far worse, than may otherwise appear if one does not explicitly incorporate the initial investment decision and trace the

[19] To compare, Gowrisankaran (1999) and Mermelstein et al. (2020) develop models with a repeated game in which horizontal mergers can be assessed, and they obtain results similar to those in Rasmusen's (1988) two-period model.

subsequent evolution. This point is particularly obvious when early investments generate large spillovers, but the analysis throughout indicates that, even without spillovers, decisions at $t = 1a$ can have large effects.

4.2.3. Uncertainty. The model here is deterministic and common knowledge is assumed. Simply adding uncertainty may appear to change little because expected values can be substituted, but even modest alterations affect the strategic environment. For example, if firm I is uncertain at $t = 1a$ about the value of parameters at $t = 2a$, including how the value of $t = 2b$ parameters will be estimated at that time, then its investment decision is notably more complex. For example, a decision to invest may probabilistically preempt firm E, and (drawing on the extension regarding broader investment choices), marginal changes in some aspects of firm I's $t = 1a$ investment vector may marginally influence that probability. Similar results would follow if firm I's uncertainty at $t = 1a$ involved firm E's type (its investment cost and investment recovery). Nevertheless, the simple model's core forces remain, so the analysis of both behavior and welfare would proceed similarly.

As a practical matter (affecting relevant magnitudes), uncertainty is highly consequential in many settings of interest. Most highly innovative investments—to create new sectors or disrupt existing ones through alternative approaches—fail, substantially or completely. Relatedly, uncertainty and failure are powerful forces affecting investment spillovers. When the earliest firms expend substantial sums exploring demand, production and distribution methods, and more, later-arriving firms observe much of this. Even if many of the failures are invisible, many traits of successful ventures are apparent or can be discerned with modest effort (through reverse engineering). This factor is associated with what some refer to as a second-mover advantage, and it may help explain why many of the most successful firms in modern technology sectors are not the initial innovators. This phenomenon also occurred throughout industrial history, such as with the automotive industry over a century ago. That is, significant uncertainty may substantially raise F_I^{1a} but F_E^{2a} may be much lower because α is high. Accordingly, if firm I is unable to find ways to reduce α along the way, it may invest far less or not at all at $t = 1a$.

5. Conclusion

This article analyzes a simple model in which a prospective dominant incumbent firm and a subsequent potential entrant each decide whether to invest to come into existence rather than taking one or both firms' existence as given, which is done in much of the literature. These investment choices depend on the firms' anticipated investment recoveries. Hence, a firm's investments are influenced by exogenous changes, government regulation that assists or limits the firm's own actions, and actions taken by the other firm. Moreover,

investment choices often involve nonconvexities, which is a central consideration when analyzing dominant firms and rivals that might pose serious challenges. As mentioned in the introduction, intangible investments alone are trillions of dollars per year. Single leading firms in important industries invest tens or even over a hundred billion dollars per year.

In this setting, marginal changes that are the focus of many analyses at some point cause discontinuous shifts, here, between equilibria. Not only does welfare change discontinuously, but in important cases, the sign of those changes is opposite to that of marginal welfare effects when one takes firms' existence (the equilibrium) as given. It is easy to construct plausible examples in which higher fixed or marginal operating costs for the incumbent or the entrant raises social welfare, and where more intense competition when both firms operate reduces welfare. Entire literatures' central conclusions on welfare (e.g., on price discrimination) are importantly incomplete and sometimes misleading. The reason is not that one can construct complex, subtle cases that may be the exception to the rule. Instead, robust forces readily generate these outcomes in a basic model with no special assumptions once firms' existence and capabilities are taken to be endogenous. In respects, these lessons are elementary and widely known. Nevertheless, much analysis relegates them to the background where often they are forgotten entirely, especially by outsiders who do not appreciate the limitations of a body of work.

The simple model presented here, even without the extensions, has a great deal of generality due to its use of reduced forms for firms' investment costs and investment recoveries, and also for consumer surplus, all of which are consistent with standard cost functions, demand curves, and models of competitive interaction. These reduced forms are sufficient statistics for both behavior (equilibrium determination) and welfare. This feature enables integration of the present analysis with substantial and often quite sophisticated literatures in industrial organization that do not endogenize firm origins.

Some illustrations were developed in section 4. Analyses of the welfare effects of price discrimination in a variety of settings (regarding demand and feasibility of various strategies) can be understood as identifying the determinants of the investment recovery and consumer surplus in any b subperiod in which a single firm operates as a monopolist. The analysis of producer surplus in conventional work is, here, an assessment of the firm's investment recovery, which in turn dictates its preceding a subperiod incentives to invest. This simple observation, which is critical when the firm's existence and, in the extension on investments, its capabilities (cost function, demand curve) are taken to be endogenous rather than given, enables one to solve for equilibrium behavior and welfare, including important cases in which effects on both consumer and total welfare have signs opposite to those in static analysis.

Much literature in industrial organization takes firms' existence and capabilities as given—especially of what is here referred to as the incumbent firm—and examines how various actions by firms and possible regulations thereof affect behavior and welfare. Such analysis is necessary but incomplete. It can often be linked to a model like the present one to provide a richer and often qualitatively different understanding of the subject.

In this regard, an important feature of the present model's generality—really, economic interpretation—should be emphasized. Firms' actions in the b subperiods in which they operate—as monopolists or as competitors in the I, IE equilibrium at $t = 2b$—were taken to be implicit in the reduced form expressions for investment recoveries and associated consumer surplus. That is, it is assumed that, however many decisions firms may have to make and whatever their complexity (including strategic interactions), these are made so as to maximize the firm's subperiod b investment recovery, and that level of recovery is the value of the associated reduced-form expression.

It is thus straightforward to embed any such actions from another model into the present one. Moreover, by using the parameterizations that introduced additional variables into these reduced forms, we can similarly apply the comparative statics analysis to all manner of regulation. Turning again to the case with price discrimination, it was explained how we might parameterize a firm's ability to price discriminate using the variable κ. In the same fashion, we can represent the tightening of a restriction on the use of price discrimination as a reduction in κ.

Taking another example, features of the protection of intellectual property rights and many forms of contemplated competition regulation can be represented as changes in the level of α, which indicated the degree of spillovers from firm I's $t = 1a$ investment. Note further that much intellectual property protection is via forms of self-help, such as by erecting firewalls, boosting cybersecurity, distributing valuable process knowledge across many workers, and simply refusing to sell the firm's intellectual property to rivals, much less to give it away for free. Much self-help is intertwined with the prevailing legal regime. The state sometimes assists (law enforcement may be deployed against hackers and physical invaders), and it sometimes leaves firms alone (it neither subsidizes nor prohibits internal cybersecurity, and it generally permits firms to choose whether to license their intellectual property and if so, at what royalty, if any). But the state also sometimes regulates contracts (such as by restricting the enforcement of noncompete clauses in employment contracts), and it sometimes may force access, such as through legislation (the Digital Markets Act in Europe) or an antitrust degree that requires access or forms of interoperability that explicitly or implicitly raise α, that is, increase the level of investment spillovers.

Additional possibilities are suggested in the discussion of preemption as part of the extension on investments. Regulation modeled as changes in the variable κ can also

incorporate rules prohibiting predatory pricing, tying, or exclusive dealing arrangements. Such regulation restricts the targeted firm's choice set in the b subperiod and thereby alters its own investment recovery and that of rival firms, and also consumer surplus. The model makes clear that welfare analysis of such regulation that is confined to the b subperiod is incomplete, for the effects on investment recoveries of the two firms influence both firms' incentives to invest (sometimes in an interacting manner), which in our simple model may shift the equilibrium, which has large and sometimes opposite-signed effects on welfare. The determinants of welfare effects in a variety of settings, moreover, often have little nexus to factors emphasized in assessments of competition regulations.

Literatures in industrial organization and those focused on competition regulation in particular—which typically examine particular practices and regulations in a specified environment—can usefully be combined with the sort of model developed here. Some literature already takes one or both firms' existence and capabilities to be endogenous. The primary message for such analyses is that, to the extent that determinate effects on behavior or welfare are identified, that definiteness may well be the product of assumptions that limit consideration to a special case. Despite the present model's simplicity, we saw that different plausible values of even a single reduced-form term can shift the equilibrium and have large welfare effects, sometimes opposite to what is usually contemplated—that is, when the same variable is changed in the same direction taking firms' existence (the equilibrium) as given.

References

Besanko, David, Ulrich Doraszelski, and Yaroslav Kryukov. 2014. "Efficiency or Predation: What Drives Pricing When There Is Learning-By-Doing?" *American Economic Review* 104: 868–897.

_____, _____, and _____. 2020. "Sacrifice Tests for Predation in a Dynamic Pricing Model: Ordover and Willig (1981) and Cabral and Riordan (1997) meet Ericson and Pakes (1995)." *International Journal of Industrial Organization* 70: 102522.

Corrado, Carol, Jonathan Haskel, Cecilia Jona-Lasinio, and Massimiliano Iommi. 2022. "Intangible Capital and Modern Economies." *Journal of Economic Perspectives* 36(3): 3–28.

Dixit, Avinash K. 1979. "A Model of Duopoly Suggesting a Theory of Entry Barriers." *Bell Journal of Economics* 10: 20–32.

_____. 1980. "The Role of Investment in Entry-Deterrence." *Economic Journal* 90: 95–106.

_____ and Joseph E. Stiglitz. 1977. "Monopolistic Competition and Product Diversity." *American Economic Review* 67: 297–308.

Ericson, Richard, and Ariel Pakes. 1995. "Markov-Perfect Industry Dynamics: A Framework for Empirical Work." *Review of Economic Studies* 62: 53–82.

Gowrisankaran, Gautam. 1999. "A Dynamic Model of Endogenous Horizontal Mergers." *Rand Journal of Economics* 30: 56–83.

Hopenhayn, Hugo A. 1992. "Entry, Exit, and Firm Dynamics in Long Run Equilibrium." *Econometrica* 60: 1127–1150.

Jullien, Bruno, Alessandro Pavan, and Marc Rysman. 2021. "Two-sided Markets, Pricing, and Network Effects." In Kate Ho, Ali Hortaçsu, and Alessandro Lizzeri, eds. *Handbook of Industrial Organization* 4: 485–592. Amsterdam: Elsevier.

Kaplow, Louis, and Carl Shapiro. 2007. "Antitrust." In A. Mitchell Polinsky & Steven Shavell, eds. *Handbook of Law and Economics*, vol. 2: 1073–1225. Amsterdam: Elsevier.

Mankiw, N. Gregory, and Michael D. Whinston. 1986. "Free Entry and Social Inefficiency." *Rand Journal of Economics* 17: 48–58.

Melitz, Marc J. 2003. "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity." *Econometrica* 71: 1695–1725.

Mermelstein, Ben, Volker Nocke, Mark A. Satterthwaite, and Michael D. Whinston. 2020. "Internal Versus External Growth in Industries with Scale Economies: A Computational Model of Optimal Merger Policy." *Journal of Political Economy* 128: 301–341.

Ordover, Janusz A., and Garth Saloner. 1989. "Predation, Monopolization, and Antitrust." In Richard Schmalensee and Robert Willig, eds. *Handbook of Industrial Organization* 1: 537–596. Amsterdam: Elsevier.

_____ and Robert D. Willig. 1981. "An Economic Definition of Predation: Pricing and Product Innovation," *Yale Law Journal* 91: 8–53

Rasmusen, Eric. 1988. Entry for Buyout. *Journal of Industrial Economics* 36: 281–299.

Scheffman, David T. 1981. "Comments on 'An Economic Definition of Predatory Product Innovation.'" In Steven C. Salop, ed. *Strategy, Predation, and Antitrust Analysis*, 397–413. Washington, DC: Federal Trade Commission.

Spence, Michael. 1976. "Product Selection, Fixed Costs, and Monopolistic Competition." *Review of Economic Studies* 43: 217–235.

_____. 1977. "Entry, Investment and Oligopolistic Pricing." *Bell Journal of Economics*: 8: 534–544.

_____. 1979. "Investment, Strategy and Growth in a New Market." *Bell Journal of Economics* 10: 1–19.

Tirole, Jean. 1988. *The Theory of Industrial Organization*. Cambridge, MA: MIT Press.

Varian, Hal. 1989. "Price Discrimination." In Richard Schmalensee and Robert Willig, eds. *Handbook of Industrial Organization* 1: 597–654. Amsterdam: Elsevier.

Appendix

Proof of Proposition 3: For each pair of equilibria for which a boundary condition exists, Proposition 3.c is straightforward (and thus will not be repeated for each pair). The welfare jump at the boundary for any pair of equilibria is given by the difference between the sums of the two corresponding rows in Table 1. When one of the equilibria is $\emptyset, \emptyset$, where welfare is 0, welfare in the other equilibrium in the pair is necessarily higher because strictly positive. For all of the other pairs (where a boundary condition exists), welfare in each equilibrium in the pair has at least one term not in welfare for the other pair, so it is obvious that, except in the knife-edge case, there will be a jump and, moreover, there will exist values for which the jump can be in either direction and thus of opposite sign to the welfare effect of the marginal change of the variable. In this regard, note further that all equilibria pairs differ in their consumer surplus term, which are not themselves part of any boundary condition.

Turn now to each of the fifteen possible pairs of equilibria. The approach, following that used in the illustration for the definition of a boundary condition, is to identify values of one or more parameters that rule out all but the two equilibria in the pair and then to show that at least one of expressions (4) – (10) is a boundary condition for that pair. (For each of the two exceptions, it is instead shown that there cannot be a boundary for that pair of equilibria.)

$\{ \emptyset, \emptyset \ , \ \emptyset, E \}$: Suppose that the F_I^t are prohibitively high. (Throughout, such statements should be interpreted to mean that $F_I^t > \pi_I^{1b} + \pi_I^{2b}$ for $t = 1a, 2a$, so there is no question that firm I would not make the initial or continuation investment whatever else might transpire—and similarly for analogous statements.) Then, expression (6) for firm E's $t = 2a$ investment decision, holding as an equality, is the boundary condition. (Here and throughout, because there are no restrictions on the reduced-form expressions for investment costs and for investment recoveries, or on how each may depend on the variables it is a function of, existence will be obvious and not remarked upon further.) For our benchmark parameterizations, for example, a marginal decrease in ϕ or κ will reduce the investment cost F_E^{2a} or raise the investment recovery π_E^{2b}, respectively, breaking indifference, inducing firm E to invest, and shifting the equilibrium from $\emptyset, \emptyset$ to $\emptyset, E$. (And a marginal increase would break indifference in favor of not investing, although under our indifference convention, we already had that case.) Note that the marginal effect on welfare in the neighborhood of the boundary is zero for $\emptyset, \emptyset$ and positive for

$\emptyset, E$, the same sign as the discrete jump in welfare, from zero to the strictly positive amount in the second (operative) row in Table 1, as the equilibrium shifts.[20]

{ $\emptyset, \emptyset$, $I, \emptyset$ }: The reasoning is the same, mutatis mutandis. (The F_f^t other than F_I^{1a} are prohibitively high, expression (8) is the relevant boundary condition, and so forth.)

{ $\emptyset, \emptyset$, I, I }: Here, the analysis is more involved because we need both of firm I's investment decisions to switch together. To ensure that only those are at play, suppose that F_E^{2a} is prohibitively high. The boundary condition is associated with expression (9), firm I's $t = 1a$ investment decision for the case in which inequality (4) holds, that is, if firm I makes the initial investment at $t = 1a$, it will also make the continuation investment at the end of $t = 2a$ (anticipating that firm E will not invest at the beginning of $t = 2a$). We also need inequality (8) to fail, for otherwise the initial equilibrium would have had firm I already investing at $t = 1a$. Moreover, we already know that it fails as a strict inequality, indeed, by the same degree to which inequality (4) holds as a strict inequality, because the sums of the two expressions (summing the investment recovery terms on the left sides and the investment cost terms on the right sides) yields expression (9) which, because we are examining it as the boundary condition, is taken to hold as an equality. At this point, we can proceed as before, mutatis mutandis (noting that marginal changes in variables that flip the equilibrium in the neighborhood of the boundary condition given by expression (9) holding as an equality cannot change the strict inequalities we had assumed for expressions (4) and (8).)

{ $\emptyset, \emptyset$, I, E }: Suppose that F_I^{2a} is prohibitively high. Then we have joint boundary conditions given by expressions (6) (for firm E's investment decision at $t = 2a$) and (8) (for firm I's investment decision at $t = 1a$ when it knows it will not make its continuation investment at $t = 2a$). The remainder of the analysis proceeds as before. (As a practical matter, this is an uninteresting case in the sense that, for a range of possible reduced forms and parameterizations, moving simultaneously across both boundaries has probability zero. A discrete change can cross both boundaries in a more meaningful class of cases.)

{ $\emptyset, \emptyset$, I, IE }: This equilibrium pair is one of the two exceptions listed in Proposition 2.b. One of the necessary conditions for both firms to operate at $t = 2b$ is that inequality (7) holds (firm E invests at the beginning of $t = 2a$ knowing that, if it does, firm I will make its continuation investment at the end of $t = 2a$). But if that is

[20] As mentioned in note 2, formally we can take the distributional derivative (of an expression that includes the pertinent indicator variable) at the boundary point, at which welfare jumps discontinuously. This derivative is given by the Dirac delta function (corresponding to the associated indicator variable going from 0 to 1; if it goes from 1 to 0, one takes the negative of the delta function) times the discrete jump in welfare, which at that point will simply yield the value of that discrete jump.

true (after at most a marginal change in a variable), then inequality (6) must have held previously because firm E's $t = 2b$ monopoly investment recovery π_E^{2b} strictly exceeds its $t = 2b$ competitive investment recovery $\pi_{E,IE}^{2b}$. Hence, $\emptyset, \emptyset$ could not initially have been an equilibrium.

$\{ \emptyset, E \ , \ I, \emptyset \}$: Suppose that F_I^{2a} is prohibitively high. For this pair of equilibria, use of our parameterized benchmarks is insufficient because any marginal change in a variable that, say, causes firm I to invest when it was previously indifferent would further boost firm E's profit from investing (when it knows that firm I will not make its continuation investment at $t = 2a$). But Proposition 3 does not restrict attention to the parameterized benchmarks. Suppose that, from the pertinent boundary conditions (expression (6) for firm E and expression (8) for firm I), we take the pertinent derivatives with respect to κ, which marginally raises π_I^{1b} but marginally lowers π_E^{2b}. (As noted in the text, we could imagine that a marginal increase in κ means that firm E arrives slightly later, which reduces the duration of its monopoly but raises that for firm I.) Such a marginal change could cause the flip in positions postulated in this pair of equilibria. The remainder of the analysis proceeds as before. Note here that, except in the knife-edge case in which $W^{\emptyset,E} = W^{I,\emptyset}$, there will be a discrete jump in welfare at the associated boundaries, one that could go in either direction. Even though each firm's profits equal zero at the boundary, we have $CS_I^{1b} > 0$ and $CS_E^{2b} > 0$ and the level of either could be higher. As noted at the outset of the proof, the possibility that welfare could jump discontinuously in either direction entails that the sign of the jump could be opposite to the sign of the marginal effect on welfare in the neighborhood of the boundary. (As a practical matter, this is another uninteresting case because, for a range of possible reduced forms and parameterizations, moving simultaneously across both boundaries has probability zero, although, again, a discrete change could more readily cross both boundaries.)

$\{ \emptyset, E \ , \ I, I \}$: In the first equilibrium in this pair, we have that inequality (6) holds, meaning that firm E makes strictly positive profits if it invests at $t = 2a$ and operates as a monopolist at $t = 2b$. Assume moreover that inequality (7) fails, specifically, that firm E loses money if it invests at $t = 2a$ but faces competition from firm I at $t = 2b$. Suppose further that inequality (5) holds, so that firm I would indeed make its continuation investment at the end of $t = 2a$ even if firm E had invested at the beginning of $t = 2a$. But, finally, firm I does not initially invest at $t = 1a$, specifically: Inequality (8) does not hold (firm I will not invest at $t = 1a$ if it only expects to operate, as a monopolist, at $t = 1b$), but we are at the boundary condition associated with expression (9) holding as an equality (firm I is indifferent between investing in both periods when it expects to operate as a monopolist in both periods). Note that, at this boundary, if firm I did invest, it would make losses overall at $t = 1$ and profits overall at $t = 2$, and these

two amounts are equal in magnitude. Then any marginal change that raises its profits (reduces its losses) in either period leads it to make both investments rather than neither. And our prior assumptions imply that firm E is thereby induced not to invest at $t = 2a$. Hence, a single boundary condition, expression (9), combined with a number of other strict inequalities, makes it possible for marginal changes in variables to move between the two equilibria in this pair. Again, except in the knife-edge case, welfare jumps discontinuously, and it could do so in either direction.

There is another, more subtle boundary condition associated with this pair of equilibria: expression (6) holding as an equality. (Although demonstrating this is not required to prove the Proposition's claim, the possibility of multiple boundaries is of interest and, more importantly, this more subtle boundary is used in the text in subsection 4.1.2.) Recall that the definition of a boundary condition does not specify whether an investment is made; relatedly, it does not depend on the indifference convention. Instead, it only posits that there are points in any neighborhood associated with each in the pair being the equilibrium. To proceed, assume that, when firm E does not invest at $t = 2a$, inequality (9) holds (firm I makes strictly positive profits if it invests in both of the a subperiods and is a monopolist in the b subperiods). Assume further that inequality (5) fails (if firm E invests at the beginning of $t = 2a$, firm I will not make its continuation investment at the end of $t = 2a$). Finally, assume that inequality (8) fails (if firm I anticipates that firm E will invest at the beginning of $t = 2a$, in which case it would not make its continuation investment at the end of $t = 2a$, its investment recovery from $t = 1b$ alone is insufficient to cover its $t = 1a$ investment cost). If all these conditions hold, any marginal change that raises firm E's profits for this case leads it to invest at $t = 2a$ (and thus for firm I not to come into existence), whereas at the boundary (given our indifference convention) or for any marginal reduction in firm E's profits in this case, firm E does not invest at $t = 2a$ (and thus firm I makes both of its investments).

$\{ \emptyset, E \ , \ I, E \}$: This case is straightforward. Suppose that F_I^{2a} is prohibitively high and that inequality (6) holds (firm E invests at $t = 2a$ if it expects to be a monopolist at $t = 2b$, which it does given the initial assumption). Then expression (8), as an equality, is the relevant boundary condition. Because the earlier assumptions isolate firm I's t=1a investment decision from everything that happens at $t = 2$, no further analysis is required. Note that when firm I does enter, welfare rises: firm I's profits at the boundary equal zero, and CS_I^{1b} is strictly positive.

$\{ \emptyset, E \ , \ I, IE \}$: Assume that inequality (7) holds (firm E invests at $t = 2a$ even if it expects competition from firm I at t=2b) and that inequality (5) holds (firm I makes its continuation investment at $t = 2a$ even if it expects competition from firm E at $t = 2b$). Then, we have the boundary condition associated with expression (10) holding as an equality. For this pair of equilibria, effects on welfare are more complicated. At the

boundary, firm I's profits equal zero, its investment at $t = 1a$ creates CS_I^{1b}, which is strictly positive. But its effects at $t = 2b$ must now be added as well: under standard assumptions, $CS_{IE}^{2b} > CS_E^{2b}$, but we also have $\pi_{E,IE}^{2b} < \pi_E^{2b}$ (the combination of which is ambiguous).[21]

$\{ I, \emptyset \ , \ I, I \}$: This case is straightforward from the above. Here, the boundary condition is expression (4) holding as an equality (firm I, having invested at $t = 1a$ and knowing that firm E will not invest regardless at $t = 2a$, is indifferent to making its continuation investment at $t = 2a$).

$\{ I, \emptyset \ , \ I, E \}$: Suppose that F_I^{2a} is prohibitively high. Then the boundary condition is expression (6) holding as an equality (firm E is indifferent to investing at $t = 2a$ when it knows that firm I will not make its continuation investment at $t = 2a$ regardless).

$\{ I, \emptyset \ , \ I, IE \}$: This equilibrium pair is the other of the two exceptions listed in Proposition 2.b, and the analysis is similar to that for $\{ \emptyset, \emptyset \ , \ I, IE \}$. Here, if we start with the first equilibrium in the pair, to move to the second requires (among other things) that inequality (7) holds after some marginal change (firm E invests at $t = 2a$ even when it knows that firm I is expected to make its continuation investment at $t = 2a$), which is inconsistent with inequality (6) failing to hold initially (firm E does not invest at $t = 2a$ even when it knows that it will be a monopolist at $t = 2b$). This follows (as before) because $\pi_E^{2b} > \pi_{E,IE}^{2b}$.

$\{ I, I \ , \ I, E \}$: This case is straightforward from foregoing analysis. Assume that expressions (8) and (9) hold, which assures that firm I invests at $t = 1a$ in the associated equilibria. The boundary condition is expression (6) holding as an equality (firm E is indifferent to investing at $t = 2a$ when it knows that it will be a monopolist at $t = 2b$). This case also requires two additional assumptions regarding firm I's continuation investment at the end of $t = 2a$: that it is profitable when firm E does not invest at the beginning of $t = 2a$ (inequality (4) holds) but unprofitable when firm E does invest at the beginning of $t = 2a$ (inequality (5) does not hold); these are, of course, mutually consistent for a range of reduced forms and parameter values. Note that, as the equilibrium shifts in the stated direction, firm E preempts firm I's continuation investment, welfare changes discretely (except in the knife-edge case), and that this change can be of either sign. For example, suppose that a marginal reduction in ϕ marginally lowers all of the investment cost terms by the same amount. That raises

[21] Although the rise in consumer surplus generally exceeds the fall in investment recoveries from competition being more intense, here we are comparing the former to firm E's loss in its investment recovery at $t = 2b$, much of which goes not to consumers but to firm I (the business-stealing externality), which is not profit to firm I but investment recovery that is fully offset (along with firm I's $t = 1b$ investment recovery) by its total investment costs.

welfare in a given equilibrium in which any investment is made, covering both of the two equilibria here, with the increases being the same in each. In addition, firm E replaces firm I at $t = 2b$, where the discrete effect on welfare could be negative.

$\{ I, I \ , \ I, IE \}$: Assume that expression (10) holds (which also implies that expression (9) holds), which assures that firm I invests at $t = 1a$ in the associated equilibria. Suppose further that inequality (5) holds (firm I invests to continue at the end of $t = 2a$ regardless of whether firm E invests at the beginning of $t = 2a$). Then the boundary condition is expression (7) holding as an equality (firm E is indifferent to investing at the beginning of $t = 2a$ when it knows that firm I will make its continuation investment at the end of $t = 2a$). Note that a marginal change in a variable that induces firm E to enter may raise welfare conditional on being in a given equilibrium even though the discrete (except for the knife-edge case) change in welfare from this shift between the two equilibria could be negative. The latter is clear if one compares the corresponding rows in Table 1. The intuition is that firm E's investment recovery, $\pi^{2b}_{E,IE}$, which just covers its investment cost, F^{2a}_{E}, comes at the expense of firm I's $t = 2b$ investment recovery (a business-stealing externality), which welfare reduction could exceed the rise in consumer surplus due to moving from a firm I monopoly to competition at $t = 2b$ (which is more likely the less intense that competition would be).

$\{ I, E \ , \ I, IE \}$: This case is straightforward from the foregoing analysis. Here, we assume that expression (7) holds (firm E invests at the beginning of $t = 2a$ regardless of whether it expects firm I to make its continuation investment at the end of $t = 2a$). The boundary condition is expression (5) holding as an equality (firm I is indifferent to making its continuation investment at the end of $t = 2a$ when it knows that firm E invests at the beginning of $t = 2a$). The welfare implications are qualitatively the same as in the preceding case, for the same reasons. (Here, we have firm I, in making its continuation investment, stealing business from firm E at $t = 2b$.) ■