The Political Economy of Occupational Licensing Associations*

Nicola Persico
Kellogg School of Management
Northwestern University

November 21, 2011

Abstract

This paper studies the internal politics of a licensing association with regards to expansion of the licensure, and to self-regulation. A theoretical model is presented of a professional association which has the power to restrict entry, and yet a majority of its members may prefer to allow entry, even when doing so reduces the total revenue of its members. This may happen due to a conflict of interests among professional sub-specialties. On the other hand, the model predicts no heterogeneity of interests within the association regarding self-regulation.

1 Introduction

"Occupational licensing directly affects approximately 18 percent of U.S. workers, which is more than either the minimum wage,"

*Thanks to David Baron, Georgy Egorov, Bard Harstad and Pablo Montanes for useful comments.
which has a direct impact on less than 10 percent of workers [...], or unionization, whose membership rates are now less than 15 percent of the labor force.” (Cited from Kleiner 2000, p. 190).

“We propose the general hypothesis: every industry or occupation that has enough political power to utilize the state will seek to control entry.” (Cited from Stigler 1971, p. 5).

The study of occupational licensing has a distinguished tradition in economics. Adam Smith lamented licensing restrictions in the crafts,¹ as did Milton Friedman concerning the professions.² The quotes above capture two views which I take to be almost universal: first, that occupational licensing is important because it covers many workers; and second, that licensing associations will necessarily seek to restrict entry in order to support wages. In this paper I do not question the first point: occupational licensing is truly ubiquitous and important. But I offer new insight on the second point. I present a simple model of a professional association which has the power to restrict entry, and yet may choose to allow entry even when doing so reduces the total revenue of its members. Entry may even be socially excessive. Thus I show that the power to regulate entry can sometimes be used by the association, unexpectedly, to create “excessive” entry. Put differently, merely awarding the power to restrict entry does not necessarily mean that it will be used. I also examine the association’s incentives to adopt quality standards.

¹ In the Wealth of Nations, Smith writes thusly regarding laws requiring a minimum apprenticeship period.

“The patrimony of a poor man lies in the strength and dexterity of his hand. [...] To judge whether he is fit to be employed, may surely be trusted to the discretion of the employers. [...] The affected anxiety of the law-giver lest they should employ an improper person, is evidently as impertinent as it is oppressive.” Cited from Kleiner (2000), p. 189.

² Chapter IX of Capitalism and Freedom (Friedman 1962) is devoted to occupational licensure.
My interest in these questions was motivated by a remarkable fact: between 1990 and 2009 the Italian legal bar almost quadrupled in size, going from 38,000 lawyers to 140,000. This fact alone seems puzzling—why should the Italian bar allow such massive entry? But there is more: the number of notary publics, which like lawyers are licensed, has remained steady at around 4,500. Why the difference? The key difference between the lawyers’ and the notaries’ associations, I argue, lies in their production functions. Lawyers come in two complementary varieties: plaintiff and defense lawyers. These two specialties are complements in the production function because every civil trial needs (at least) one of each. Notaries, on the other hand, are substitutes for each other. In the model, complementarity produces a heterogeneity of interests regarding licensure extension, including possibly a majority of members in favor of expansion. Heterogeneity of interests cannot arise when association members are substitutes, as in the notary publics’ association.

To see how the argument goes, consider an expansion of the lawyers’ licensing association (the bar). An above-average influx of new lawyers is manufactured, in practice, by lowering the bar exam requirements. Since the “extra entrants” who benefit from this shift are the marginals in the bar exam, the extra entrants must be of a different type than the incumbents. Their different types, in our model, lead the extra entrants to be more inclined to choose one of the two specialties, compared to the incumbents. Say, for concreteness, that extra entrants are more likely to become plaintiff lawyers. Now, given the complementarity in the production function, an influx of plaintiff lawyers requires an adjustment: some incumbents need to switch to defense. This occupational switch must be mediated by a relative increase in the defense lawyer’s wages. In fact, defense wages could even go up. In this case expansion is good for incumbent defense lawyers. Since they represent 50% of incumbents, we already have a quasi-majority in favor of

---

3Italy is not alone: In Germany, lawyers almost tripled during the same period.
4And, in Italy, are a very lucrative profession.
expanding licensure. In fact, a little reflection shows that there must be a (possibly small) mass of incumbent types who choose to switch from plaintiff to defense work, and who also benefit from the change. And so in fact a strict majority of incumbents favor expansion.

The incentives for incumbents to expand the licensure are reduced when the extra entrants encroach on the incumbents’ business. Nevertheless I show that theoretically the incentive for expansion could be arbitrarily large and thus overwhelm any encroachment effect. Which effect dominates in practice will depend in part on whether the extra entrants can create their own business. I will return to this issue in Section 2.1.

How broadly does this logic apply? In principle it applies to all licensing associations that represent complementary activities. A medical board, for example, allows access to a set of different specialties (family doctor, specialists of various kinds), many of which are complementary in healing the patient. For example, when a patient first visits a family doctor and is then referred to a surgeon, the two specialists are complements in the production function. According to the mechanism proposed in this paper, if the supply of family doctors should increase then surgeons would benefit, which seems reasonable. Along the same lines, I believe that the mechanism proposed here is applicable to a variety of professional associations. The logic fails, however, if the different specialties are represented by different associations (if, hypothetically, there were two bar exams, one for plaintiff and another for defense lawyers). In this case there are no incentives to expand the association(s). I discuss this case in Section 5. In Section 4.5 I discuss the (lack of) dynamics in licensure extension. In Section 4.6 I discuss the difference between unions and licensing associations as pertains to membership enlargement.

In Section 5 and 6 I explore the forces that generate the “market structure” of licensing associations, in the sense of giving these associations boundaries.
In Section 7 I turn to a different question. I ask whether there is any heterogeneity of interests within the association regarding the adoption of quality standards, or self-regulation. Self-regulation is, arguably, a policy justification for occupational licensing. Our society tolerates the potentially anti-competitive professional associations in part because we believe that associations are able to impose quality standards (codes of professional responsibility) on their members. Self-regulation, as opposed to regulation, is especially prevalent in the case of the professions, arguably because of the expertise required to regulate them. But, will association members choose to self-regulate and, if so, to what extent? The answer to this question depends, again, on the internal political economy of these associations. I show that the model predicts homogeneity of interests, even when the cost of the regulation is unequally distributed across specialties. To maximize the potential for heterogeneity of interests, I consider a rule or regulation the cost of which falls solely on one specialty within the profession, but which benefits all specialties equally. Given this asymmetry, one would guess that there might be a divergence of views within the association, and that the regulation would be supported more strongly by the group which does not bear its cost. However it turns out that, due to a “translation of costs” argument analogous to the analysis of tax incidence, wages adjust so as to fully align the interests of both specialties. As a result, I find that there is perfect unanimity within the professional association with regards to the application of a code of professional ethics. In Section 8 I study the interaction between the two policies: expanding the licensure and self-regulation.

Throughout the paper I focus on the welfare of incumbents, and subgroups of them, because I am interested in the positive questions of whether this institution will voluntarily choose to expand, whether there are conflicts about self-regulation, and on the effect of expansion on self-regulation. I do not take a normative position on whether expansion or self-regulation is good or bad. Normative views are sundry and sometimes conflicting: lawyers are too expensive (suggesting that expanding licensure would be good) and yet
there are too many frivolous lawsuits (suggesting the opposite). The normative question, I believe, is beyond the scope of mere theory and hence of this paper. I say this in Section 9.

1.1 Related Literature

A large and distinguished literature focuses on the ill effects of occupational licensing. Adam Smith lamented licensing restrictions in the crafts, as did Milton Friedman concerning the professions. Stigler (1971, 1972) wrote the seminal political economy papers in this area. Kleiner (2000) provides a good survey. From an empirical viewpoint, the challenge in measuring the causal effects of licensing is finding sources of exogenous variation in licensing.

The sociologist Emile Durkheim emphasized the importance of professional associations in providing their members with moral rules. Professional associations would potentially be able to fill in an ethics gap left open by the disorderly, violent, and confused market system. Durkheim also briefly discusses the internal political organization of a theoretical guild.

\[\text{In the Wealth of Nations, Smith writes thusly regarding laws requiring a minimum apprenticeship period.}
\]

\[\text{“The patrimony of a poor man lies in the strength and dexterity of his hand. [...] To judge whether he is fit to be employed, may surely be trusted to the discretion of the employers. [...] The affected anxiety of the law-giver lest they should employ an improper person, is evidently as impertinent as it is oppressive.” Cited from Kleiner (2000), p. 189.}
\]

\[\text{6Chapter IX of Capitalism and Freedom (Friedman 1962) is devoted to occupational licensure.}
\]

\[\text{7Stigler (1972), p. 100 writes: Particular industries and occupations obtain from the state a variety of economic privileges which are injurious to the vast majority of the population. Farm subsidies, oil import quotas, tariffs, and occupational licensing are examples.}
\]

\[\text{8Durkheim (2003), p. 13-17.}
\]

\[\text{9“Let us imagine—spread over the whole country—the various industries grouped in separate categories based on similarity and natural affinity. An administrative council, a kind of miniature parliament, nominated by election, would preside over each group. We}
\]
The model presented here could be recast as a model of extension of political franchise or club membership. There are literatures on franchise extension (see Acemoglu and Robinson 2000, Lizzeri and Persico 2004) and on club membership extension (see e.g. Barbera’ et al. 2001, Acemoglu et al. 2008), but I am not aware of a model in these literatures that is similar.

In our model, a majority of incumbents may favor expansion but the specialty which is directly hit by the influx (plaintiff lawyers) is unambiguously against expansion. This is because new entrants are substitutes for them. This logic is reminiscent of the labor literature on the impact of immigration on the wages of natives (see, e.g., Borjas 2003).

2 Model

We start with an exogenously fixed measure \( N \) of atomistic licensed professionals. For simplicity I will call them lawyers, and the professional association I will call the bar. Each lawyer is indexed by \( i \) and has an occupational choice: she can choose to become either a plaintiff or a defense lawyer, or to be inactive. There is a cost \( c^i_p \) of being an active plaintiff lawyer and \( c^i_d \) of being an active defense lawyer. These costs represent the cost of training in a specialty, keeping an office, professional education (CLE), insurance, customer development, etc. Being inactive costs zero.

Each lawsuit requires exactly one plaintiff and one defense lawyer. Each active lawyer can be involved in at most one lawsuit. Each lawsuit creates a surplus \( V(N) \) which is is split between the plaintiff and defense lawyers through their wages. Wages by specialty (plaintiff and defense) are denoted by \( w_p \) and \( w_d \), respectively, and are assumed to be constant within specialty. By assumption \( w_p + w_d = V(N) \). The per-lawsuit surplus is assumed to be

---

go on to imagine this council or parliament as having the power, on a scale to be fixed, to regulate whatever concerns the business: [...] and there we have the guild restored, but in an entirely novel form.” Durkheim (2003), p.37.
weakly decreasing in \( N \), which implies that enlarging the bar reduces (or at least does not increase) the incumbents’ business.

**Definition 1** Fix the set of lawyers admitted to the bar. An occupational equilibrium is a pair of wages \( w_p^*, w_d^* \) such that given those wages half the active lawyers choose to become plaintiff and the other half choose to become defense lawyers.

By this definition agents are wage-takers: in choosing her occupation an agent does not take into account the effect of her individual decision on market wages. This assumption seems appropriate in light of the small size of each agent. Still, perhaps this equilibrium notion should be called *Walrasian* occupational equilibrium to emphasize its price-taking nature, in contrast to the price-making features of Makowski and Ostroy’s (1995) definition of occupational equilibrium.

Let us arrange lawyers so that \( c_p^i - c_d^i \) is increasing in \( i \). This convention means that lower-\( i \) lawyers have a comparative (not necessarily absolute) advantage in being plaintiff lawyers. The cumulative distribution of the \( i \)'s is continuous and denoted by \( F(i) \).

**Assumption 1** Bar membership is monotonic in \( i \).

This assumption is much stronger than is needed to get the results. All we need is that admission to the bar is regulated by a test the outcome of which is correlated—positively or negatively, strongly or weakly—with \( i \). However, for ease of exposition I stick with the strong version of the assumption and defer further discussion and refinement of it to Section 2.1 and Remark 1.

Let \( \hat{i} \) be the lowest type admitted to the bar, that is, the admission threshold. By assumption all lawyers with \( i > \hat{i} \) belong to the bar and, therefore the total number of lawyers \( N = 1 - F(\hat{i}) \). Bar membership can be expanded by lowering the admission threshold.
Example 2 (functional form example) Suppose $c_p^i = P/i$ and $c_d^i = D/i$, where $D > P$ are positive numbers. Since lawyers with larger $i$ have a lower cost of entering both specialties, $i$ can be interpreted as an index capturing a general-purpose intelligence or ability trait. Given this interpretation of $i$, it is natural to assume that only those lawyers with high $i$ (the smartest ones) pass the bar exam. Furthermore, $c_p^i - c_d^i = (P - D)/i$ is increasing in $i$, as required by the model.

2.1 Discussion of Modeling Assumptions

Assumption 1 implies that new members who are admitted through enlargement have a comparative advantage in becoming plaintiff lawyers. This important assumption is substantiated factually in some important cases. In the case of lawyers, for example, there is widespread lore (and some evidence) that lawyers with low academic credentials are more likely to become plaintiff lawyers.\(^\text{10}\) Analogously, medical students with lower MCAT scores are more likely to become family doctors.\(^\text{11}\) Thus, if licensure is expanded by lowering the threshold in licensing exams, then the “extra” licensees would be more likely than average to become plaintiff lawyers or family doctors, consistent with the model. Moreover Assumption 1 can be relaxed considerably; this will be shown later in Remark 1.

It is worth remarking that, although we presented the model so that lawyers choose their occupation after gaining admission to the bar, nothing prevents the occupational choice from being made contemporaneously or even before bar admission.

We assume that $V(N)$ is (weakly) decreasing in $N$. This property cap-


tures the possibility that new association members may encroach on the
incumbents’ business. As we will see, licensure is more likely to be expanded
if \( V(N) \) does not decrease too sharply with \( N \), that is, if the encroachment
effect is not too strong. In practice, this means that new entrants must
be able to generate at least some new business. Whether new entrants in
a profession actually generate new business will depend on the situation.
Empirical studies attempting to estimate the elasticity \( \partial V(N)/\partial N \) have to
deal with the serious concern that \( N \) is endogenous. I am aware of only
very few studies that make use of plausibly exogenous variation in \( N \). For
Italian lawyers, Buonanno and Galizzi (2010) use geographic variation in the
location of law schools as an instrument for lawyer density; they estimate a
2-6\% increase in lawsuits for every 10\% increase in lawyers,\(^{12}\) suggesting that
extra entrants are in fact capable of generating a significant amount of new
business. For German physicians, Jurges (2007) uses a similar instrument
and finds evidence of physician-induced demand of magnitude comparable
to that of Italian lawyers.

One can provide a microfoundation for \( V(N) \) as follows. Suppose there
is a large number of cases, varying in value. Of course, the highest-valued
cases will be picked first. Then, when \( N \) increases more of the lower-valued
cases must also be chosen, although it is likely that incumbents retain at least
some advantage in securing high-value cases. In this formulation new entrants
may not encroach on the incumbents’ business very much, and the decline
in average case-value for bar members overestimates the encroachment effect
as felt by the incumbents.

We assume that all lawyers can switch occupation in response to extra
entry. In reality, it is possible that incumbent lawyers are less adept at switching
occupations, compared to new entrants. To the extent that incumbent
cannot (or find it hard to) switch occupation, the size of the supermajority
in favor of extension will be smaller. But it will remain a supermajority

\(^{12}\)See also Carmignani and Giacomelli (2009), who obtain similar estimates.
provided that a fraction, albeit small, of incumbents switches occupations in equilibrium.

3 Occupational Equilibrium for Fixed License

In this section I work out the occupational equilibrium prices. To avoid keeping track of voluntary unemployment, I will restrict attention to occupational equilibria in which every bar member is active. Sufficient conditions are provided for all lawyers to be active in equilibrium. Throughout this section \( N \), the size of the bar, is kept fixed.

Fix \( i \), and hence the set of lawyers admitted to the bar. An active bar member \( i \) chooses to be a plaintiﬁ lawyer if

\[
 w_p - w_d \geq c_p^i - c_d^i \equiv c^i, \tag{1}
\]

where \( c^i \) represents the comparative cost of being a plaintiﬁ lawyer. By construction \( c^i \) is increasing in \( i \). Let \( i^m \) be the median lawyer among those admitted to the bar and \( c^m \equiv c^{i^m} \) denote the comparative cost of that median lawyer. Equation (1) must hold with equality at \( c^m \). Indeed, if this is so then all lawyers with \( i \) smaller than \( i^m \) strictly prefer to become plaintiﬁ lawyers, and the others will strictly prefer become defense lawyers; and both groups number exactly \( N/2 \), as they must in any equilibrium in which all lawyers are active. Figure 1 represents graphically how the different types split across occupations.

Substituting \( c^m \) and \( w_d = V(N) - w_p \) in (1) we can solve for the equilibrium plaintiﬁ wages:

\[
 w_p^* = \frac{V(N) + c^m}{2}. \tag{2}
\]

We see that the share of the surplus appropriated by the plaintiﬁ lawyer, is increasing in \( c^m \). This is because an increase in \( c^m \) means that the marginal
Figure 1: Occupational choices by different types.

lawyer has a higher opportunity cost of being a plaintiff lawyer. Since the marginal lawyer must be indifferent between the two specialties, equilibrium wages must go up for plaintiff work.

The wage of defense lawyers is

\[ w_d^* = V(N) - w_p^* = \frac{V(N) - c^m}{2}. \]  

(3)

As \(c^m\) shifts, we see from (2) and (3) that defense and plaintiff wages respond by moving in opposite directions. This effect causes heterogeneity of views among organization members. We call this the “wage-pivot” effect.

Let us now give conditions under which all lawyers choose to be active in equilibrium. For plaintiff lawyers this means ensuring that \(w_p^* - c_p^i \geq 0\) and
for defense lawyers that $w_d^* - c_d^i \geq 0$. Using (2) and (3) these participation constraints read, respectively,

$$\frac{V(N) + c^m}{2} \geq c_p^i \text{ for all } i \leq i^m, \quad (4)$$

$$\frac{V(N) - c^m}{2} \geq c_d^i \text{ for all } i \geq i^m. \quad (5)$$

Conditions (4) and (5) are more likely to be verified if the value of a lawsuit $V$ is large relative to the costs.

**Proposition 3** Fix the bar admission threshold $\hat{i}$ (which pins down the membership size $N$, the median type $i^m$, and that type’s comparative cost $c^m$). If conditions (4) and (5) are verified then there exists a unique occupational equilibrium and in this equilibrium all bar members are active. Equilibrium wages are given by (2) and (3).

From now on I will implicitly assume that conditions (4) and (5) are verified for all bar admission thresholds we consider.

4 The Political Economy of Licensure Expansion

4.1 Conditions for a Majority of Incumbents To Favor a Small Licensure Expansion

Let us consider the effect of a small expansion of the licensure on the welfare of incumbents. Consider a small expansion of the licensure, which by assumption is obtained by lowering the admission threshold to some $\hat{i}' < \hat{i}$. Using equation (3), and denoting by $w_d^*(\hat{i}')$ the equilibrium wage for defense
Lawyers with admission threshold \( i \), we can write

\[
\frac{\partial w^*_d(i)}{\partial i} = \frac{1}{2} \left( \frac{\partial V(N) \partial N}{\partial N \partial i} - \frac{\partial c^m}{\partial i} \right)
\]

(6)

We see that the sign of the wage variation depends on the sum of two components. The first component, \( \frac{\partial V(N) \partial N}{\partial N \partial i} \), captures the encroachment effect as the profession grows in response to a decrease in the threshold. This component is positive but bounded above. Indeed, even in the case of full encroachment where \( V(N) = V/N \) this effect cannot be larger than \( \frac{V}{N^2} \left( -\frac{\partial N}{\partial i} \right) \). The second component, \( -\frac{\partial c^m}{\partial i} \), captures the wage-pivot effect. This effect is negative because increasing the threshold \( i \) increases the median \( c^m \). The size of the wage-pivot effect can be as large as one wants, depending on the shape of the distribution \( F \) around its median \( c^m \). If the density of \( F \) is very low around the median then the marginal effect of \( i \) on \( c^m \) is very large. In this sense we can say that even under full encroachment, the wage-pivot effect can dominate under appropriate distributional assumptions.

Whenever the wage-pivot effect dominates, a majority of bar members favors expansion. This is shown in the next proposition.

**Proposition 4** Suppose the wage-pivot effect dominates the encroachment effect, i.e., \( \frac{\partial V(N) \partial N}{\partial N \partial i} < \frac{\partial c^m}{\partial i} \). Then a majority of incumbent bar members prefers a small expansion of bar membership to the status quo.

**Proof.** Lowering \( i \) by a small amount will lower \( c^m \) to some \( c^{m'} \) and hence, through (2), it will lower the plaintiff lawyer wage to some \( w^*_p < w^*_p \). Conversely, the defense lawyer’s wage goes up if and only if expression (6) is positive. This means that all incumbent lawyers of a type above the previous median lawyer \( i^m \), who used to receive a payoff of \( w^*_d - c^m_d \) can now (and will choose to) keep being defense lawyers at a greater payoff. So at least 50 percent of incumbents strictly favors expansion. And in fact, more than 50
percent of incumbents are strictly in favor of expansion. Indeed, since the new median $c^m$ is discretely below the old median $c^m$, there will be some incumbents with type slightly below the old median $i^m$ who at the old wages were almost indifferent between being plaintiff or defense, and now with the new equilibrium wage structure switch to defense lawyer and are made better off. ■

The proposition says that a small expansion is attractive to a majority of the members if the comparative cost $c^i$ for the median voter is highly responsive to changes in the type, and if the encroachment effect $\frac{\partial V(N)}{\partial N}$ is small. Under these same circumstances, also, a majority among the current incumbents strictly prefers not to restrict licensure. To see this compare payoffs in an economy with $N + \Delta$ lawyers are we contemplate cutting $\Delta$ lawyers from among those with the lowest $c^i$. Lawyers who get cut do not like the idea of restricting themselves out of their profession. But even among those who survive the cut, the majority is against restricting the licensure. To see this observe that, when $\Delta$ is small, those who gain from restricting to $N$ are the same people who lose from expanding from $N$ to $N + \Delta$; and we have shown already that these people are fewer than $N/2$.

We end this section with a remark about the role played by Assumption 1.

**Remark 1** Assumption 1 states that aptitude in the bar admission test is positively and perfectly correlated with $i$. There is obviously no difficulty in relaxing the perfect correlation, provided new admittees are more likely to become plaintiff lawyers. Also, the direction of the assumption can be flipped while still preserving Proposition 4: if we assume that new entrants are more likely to become defense lawyers then the model predicts a majority of existing lawyers (including all plaintiff lawyers) being in favor of expansion. What really matters then for the results to go through, is that the influx of new lawyers should move the median $c^m$ in some direction, and through it the wages. This is a mild assumption.
4.2 Tyranny of the Majority: Politico-Occupational Equilibrium Extension and Its Effect on Incumbents Welfare

In this section we show that when the size of the expansion is left to majoritarian politics, there can be expansion even if expansion reduces the payoff (welfare) of incumbents.

To see this, observe that the total payoff of the incumbents is given by

$$\int_i^\infty \frac{V(N)}{2} dF(i) - \int_{i}^{i_m} c_p(i) dF(i) - \int_{i}^{\infty} c_d(i) dF(i).$$

Let us focus on small changes in the admission threshold. As the threshold $i$ is moved, the change in incumbent welfare is given by

$$\frac{1}{2} \frac{\partial V(N)}{\partial N} \frac{\partial N}{\partial i} [1 - F(i)] + \frac{\partial i_m}{\partial i} \left[ c_d(i_m) - c_p(i_m) \right] f(i_m)$$

(7)

The first addend represents the encroachment effect. The sign of this term is positive which means that this effect reduces the incumbents’ welfare as the threshold is lowered. The second term reflects the gain or loss in allocative efficiency as the incumbents of median type switch from plaintiff to defense work. This switch might be welfare-improving, but if $c_d(i_m) > c_p(i_m)$ it is not. In this case incumbent welfare is unambiguously hurt by expanding licensure. A small licensure extension is welfare-improving for the incumbents if and only if (7) is negative, which after rearranging means

$$\frac{1}{2} \frac{\partial V(N)}{\partial N} \frac{\partial N}{\partial i_m} < c_p(i_m) f(i_m) \frac{1 - F(i)}{1 - F(i_m)}.$$  

By contrast, a small expansion is favored by a majority of incumbents if and only if (6) is negative, that is, if

$$\frac{\partial V(N)}{\partial N} \frac{\partial N}{\partial i_m} < \frac{\partial c}{\partial i} \bigg|_{i=i_m}.$$  

16
The point is that the two conditions are different. Whereas expansion increases the incumbents’ welfare if the level of $c^m$ is large, the political likelihood of expansion depends on $c^i$ being very responsive to changes in $i$ around the median lawyer. Thus if the majority of incumbents is in charge, expansion can happen even if it is welfare-reducing for the incumbents.

**Proposition 5** An expansion which is preferred by the majority of incumbents need not improve incumbent welfare: and it will decrease welfare if the incumbent median voter has a comparative advantage in being a plaintiff lawyer (i.e., if $c^m < 0$). Conversely, if $c^m > 0$ an expansion may be welfare improving for the majority of incumbents and yet it it is possible that the majority prefers the status quo.

This proposition highlights the misalignment between the interests of the majority and those of the universe of incumbents. Ceteris paribus, the majority tends to favor those extensions which cause large wage swings. Incumbent welfare, in contrast, improves when the incumbents who switch specialty realize a large comparative cost (irrespective of the size of the wage swing).

### 4.3 Different Groups and Their Favored Extension Size(s)

The previous sections focused on small extension of the licensure. Let us now expand our focus to extensions of any size. When there are several possible extension sizes, which size of extension is supported by the largest plurality? And, more generally, which size of extension do different types of incumbents prefer? These questions are interesting from a governance viewpoint, because if the majority has heterogeneous preferences over the ideal extension size then it may be more difficult for them to agree on a common agenda; and it may be easier for their opponents to “divide and conquer” them.

Consider reducing $i$ to some $i' = i - \Delta$. The influx of low-$i$ lawyers causes
c^{m} to go down. Therefore the 50% of incumbents who before the extension do defense work keep doing it after the extension; such a lawyer’s utility after an expansion of size $\Delta$ is given by

$$w^{*}_{d}(i - \Delta) - c_{d}^{i}.$$  \hfill (8)

This expression is not necessarily monotonic in $\Delta$.

Consider now an incumbent $i$ who was a plaintiff lawyer before the extension. After the extension he compares his payoff in the two occupations. If $i$ chooses to be a defense lawyer his utility will be given by (8). If $i$ chooses to remain plaintiff then his payoff as a function of $\Delta$ will be given by

$$w^{*}_{p}(i - \Delta) - c_{p}^{i}.$$  \hfill (9)

This expression is monotonically decreasing in $\Delta$ because for any $\hat{i}$

$$\frac{\partial w^{*}_{p}(\hat{i})}{\partial \hat{i}} = \frac{1}{2} \left( \frac{\partial V(N)}{\partial N} \frac{\partial N}{\partial \hat{i}} + \frac{\partial c^{m}}{\partial \hat{i}} \right) > 0.$$  

Lawyer $i$’s equilibrium payoff as a function of $\Delta$ is given by the upper envelope of the two functions (8) and (9); refer to Figure 2.

Lawyer $i$ prefers the $\Delta$ which maximizes the upper envelope. Since $w^{*}_{p}(i - \Delta)$ is monotonically decreasing in $\Delta$, this maximizer can be either 0 or

$$\Delta^{*} \equiv \arg \max_{\Delta} w^{*}_{d}(i - \Delta).$$

Lawyers with low type $i$ have low comparative cost of being a plaintiff $c^{i}$ and so will receive a higher payoff from 0 than from $\Delta^{*}$; conversely types with high $i$ prefer $\Delta^{*}$ to 0. But regardless, all incumbent types $i$, whatever their occupation in the status quo, have ideal points within the same very limited set: each type either prefers no extension or, if she prefers extension, the preferred extension size is the same $\Delta^{*}$ for all $i$. This extension is also the one that is preferred by the 50% of defense incumbents. Therefore, $\Delta^{*}$ is the size of extension that, when offered as a choice against the status
Figure 2: Wages available to incumbent of type $i$ for each size of expansion. Incumbent’s payoff is upper envelope of the two wage functions.

quo, maximizes the number of lawyers who support extension. Moreover, $\Delta^*$ would also be the choice of the 50% incumbent defendant lawyers, if for some reason they dominated the decision-making process. Thus $\Delta^*$ is a focal alternative to challenge the status quo; if an extension happens, we should expect it to be of size $\Delta^*$.\footnote{It is possible that $\Delta^* = 0$, which means that the demand is so inelastic that no-one prefers extension.} These findings are collected in the following proposition.

**Proposition 6** Incumbents are split into at most two groups regarding the ideal size of expansion. A minority composed of low types has zero as its ideal expansion; the rest, a majority composed of high types, all have the same $\Delta^*$ (which could equal zero) as their ideal expansion. Moreover, the
size of the majority in favor of expansion is maximized at expansion $\Delta^*$. Thus an expansion of size $\Delta^*$ will be implemented in any governance system in which a majority chooses policy.

An implication of our analysis is that simply increasing the size of the extension will not necessarily increase the support for it. This is because support for an extension depends on whether it generates higher wages. These higher wages cannot be plaintiff wages, for these are monotonically decreasing in the extension; and so they must be defense wages. But these are maximized at $\Delta^*$. Extending beyond $\Delta^*$ will worsen both wages. In other words, the level of support for extension is not monotonic in the size of the extension.

### 4.4 Supermajoritarian Politics and the Favorite Extension Size

In some cases more than 50% of a given incumbent population needs to be in favor of expansion in order for it to take place. Sometimes this might be because of explicit statutory requirements (supermajority voting rules). In other cases, it may be due to more subtle institutional features. Consider the presence of criminal lawyers, for example, which we have ignored until now. Criminal lawyers are unanimous in opposing extension, because new entrants cannot possibly increase the ranks of their complementary specialty.14 Now if, say, 30% of incumbent lawyers are criminal lawyers and the requirement for expansion is simple majority among all lawyers, then for expansion to happen it must be favored by at least 5/7ths of the incumbent civil lawyers.

When expansion is subject to supermajority requirements it is less likely to happen. However, support for expansion cannot be increased by distorting

---

14 This is because the complementary “specialty” to the criminal lawyer is the district attorney, a profession which is not controlled by the bar association.
the expansion size away from $\Delta^*$ (refer to Proposition 6). Hence the following testable implication follows.

**Proposition 7 (Testable Implication)** The probability of implementing an extension decreases with the size of the supermajority required for extension. The size of the extension, conditional on it being implemented, is independent of the supermajority requirement.

### 4.5 No Progressive Extensions and No Coalition Build-Up Time

In some models of club membership, enlarging the club membership creates a new constituency for enlargement. In these models club membership can grow progressively under the impulse of each wave of new members. This is not the case in our model. The reason is that the newly admitted members, by assumption those with the lowest $c$, are the ones most set against further expansion. So the drivers of expansion must be the incumbent defense lawyers. But these could have engineered a larger expansion in the first go, and chose not to. This reasoning shows that expansion will be one shot, or to put it differently, expansion takes place in reaction to changes in the environment but not to changes in membership.

Note also that, if there is any support for expansion among incumbents, then in our model that support exceeds 50% of the incumbents. This follows from the fact that 50% of incumbents (the defense lawyers, who do not switch) have an objective function that is identical up to an additive constant (their cost $c^*_d$). Moreover, defense incumbents are the sub-group most inclined to favor expansion. This observation implies that change is expected to be sudden: when pluralities coalesce in favor of expansion, perhaps in reaction to a change in the environment, we should expect these pluralities to be large (in the model, majorities) and to coalesce quickly.
These observations are collected in the following proposition.

**Proposition 8** As long as primitives do not change, there is no progressive expansion in our model. Expansion, when it happens, is one-shot. Moreover, when a majority coalesces in support of an extension (perhaps in reaction to a change in primitives) the majority coalesces suddenly, not gradually.

### 4.6 Difference with Unions

Unions are different from licensing organizations in many dimensions, of course. One that is critical for my argument is that unions, at least in the US, do not by law have a monopoly on an occupation. Unions may have a *de facto* monopoly over a specific workplace, but not over an occupational choice at large. Thus the issue of enlargement is different for a union, compared to a licensing organization. For a union, enlargement means increasing its coverage of workers in a given occupation. A key objective (aside from raising membership dues) is to monopsonize a segment of the labor market. The advantage of monopsony is greater bargaining power vis a vis the employer. In our notation, we can capture this effect in reduced form as $\frac{\partial V}{\partial N} / \partial N > 0$. As membership in the union increases, its bargaining power increases and so does its ability to extract surplus ($V$) from employers. In our language, it is as if we have a “reverse encroachment” effect. Recall that, in our previous analysis, the encroachment effect is what limited the politically desirable licensure size. In the case of the union the reverse encroachment effect militates is a force pushing towards enlargement. According to this analysis, therefore, a union will always want to expand regardless of whether it covers complementary or substitute trades.
5 No Complementarities, Segregated Complementarities, and Licensure Extension

I have assumed up to now that two complementary specialties (plaintiff and defense lawyers, in our case) are part of the same association. Two natural questions arise. First, what happens if the two specialties are not complementary or, which is the same, if only one specialty is represented by the association. This is the case of the notary publics which I mentioned in the introduction. This case can be modeled by assuming that the compensation of each member is $V(N)/2$. This assumption reflects the absence of a need to split the surplus with a different professional figure, which in this case is not required for production. Extending the licensure in this case increases $N$ and decreases wages. So in these associations there is never any support for extending licensure.

A more subtle analysis is required when there are two complementary specialties, but they are segregated into separate associations, one for each specialty. We now sketch out a model of a market structure with “segregated complementarities” and show that in this case the incentives to expand disappear.

Suppose there are two licensing associations, one for plaintiff and the other for defense lawyers. These associations choose their minimum thresholds $t_p$ and $t_d$ simultaneously and independently. After each association has set its minimum threshold, the following subgame is played among the individual lawyers. Every lawyer $i$ chooses which association to apply to based on wages. Any type $i$ who is above the threshold and applies is admitted. Wages are determined by the occupational equilibrium.

We start by describing the equilibrium in the subgame.

**Case A:** Suppose first that the two associations have the same number of open slots: $2(1 - F(t_d)) = (1 - F(t_p))$. Suppose that wages are as in
the occupational equilibrium of Section 3. At these wages all lawyers with \( i < t_d \) choose to become plaintiff, the rest opt to become defense lawyers. So for this choice of \( t_p \) and \( t_d \) the occupational equilibrium wages give rise to an equilibrium in the subgame where lawyers self-select into associations exactly as in Section 3.

**Case B:** Suppose now that we decrease \( t_d \) to \( t'_d \) without changing \( t_p \), so that \( 2 \left(1 - F(t'_d)\right) > \left(1 - F(t_p)\right) \). This means that, in principle, there could be more licensed defense than plaintiff lawyers. Indeed, if \( t'_d \) is below \( t_p \) then the threat of having more defense than plaintiff lawyers, which would be incompatible with an occupational equilibrium, pushes equilibrium wages down for defense lawyers to the point at which entry is fully deterred. If instead \( t'_d \) is above \( t_p \) then we now show that the total number of active lawyers does not change and, in fact, the occupational equilibrium wages of Section 3 still represent an equilibrium. To see this, recall that in the occupational equilibrium of Section 3 the lawyer with type just below the median type strictly preferred becoming a plaintiff lawyer. Lowering \( t_d \) a little bit does not constrain how this or any other lawyer self-selects into an occupation. Thus the wages, and behaviors of Section 3 remain an equilibrium. In this equilibrium the defendants association will be “undersubscribed.”

**Case C:** Suppose instead that we decrease \( t_p \) to \( t'_p \) without changing \( t_d \), so that \( 2 \left(1 - F(t_d)\right) < \left(1 - F(t'_p)\right) \). This means that, in principle, there could be more licensed plaintiff than defense lawyers. In this case the equilibrium cannot be the same as in Section 3. These wages are no longer an occupational equilibrium because now more lawyers join the plaintiff bar than there are defense lawyer. Now, the plaintiffs’ wages have to dissuade some types from joining the plaintiff association. This shows that the new equilibrium wage \( w'_{p} \) must be lower than \( w^*_p \).

Now let us move back to the previous stage and consider the incentives for each association to expand its licensure. There is no longer an incentive to expand. Indeed, the defense lawyers association is at best indifferent between
expanding and not (note the difference with Proposition 4). This is because now the defense bar only controls its own membership, but it cannot change the composition of the other specialty. And, as before, plaintiff lawyers are strictly against expansion.

I collect these findings below.

**Proposition 9** *When there are no complementarities (notary publics) then expanding is never preferred by association members. When complementary specialties are segregated in separate associations neither association ever gains from expansion regardless of whether* \( \frac{\partial V(N)}{\partial N} \) *is smaller than* \( \frac{\partial c_m}{\partial N} \).

This proposition specifies scenarios in which the conventional view is correct: professional associations want to restrict membership. When these scenarios are nested within our framework, they lead to the following hypothesis: professional associations whose boundaries do not extend to include complementary occupations, are more likely to lobby for restricted access compared to associations which cover complementary occupations.

Another observation follows from the analysis of the model. When it comes to enlargement, whatever defense lawyers prefer they can achieve. So within the model we can think of defense lawyers as running the association. Plaintiff lawyers’ preferences regarding expansion are not respected, in that sometimes there is excessive expansion from their perspective. Note that this expansion would be prevented if the two specialties had separate licensing associations (cf. Proposition 9). Therefore the next proposition follows.

**Proposition 10** *The specialty with high qualifications (defense lawyers) prefers a joint licensing association with the low-qualification specialty (plaintiff lawyers). The low-qualification specialty prefers to have its own separate licensing association.*

The proposition will acquire some relevance in the next section.
6 Division of Labor and the Boundaries of Professional Associations

With technological progress, tasks that become standardized are often parceled out to other, more technical workers. In this way a new professional figure emerges. I refer to this process as “division of labor.” In the present context, division of labor is interesting because it creates two complementary activities where previously there was only one. This raises the question of what “professional association structure” is expected to emerge, that is, whether the newly emerged technicians will be regulated by the old professional association which spawned them, or whether they will create their own independent association. Proposition 10 predicts a conflict of interest between the two specialties, with the specialty with the highest qualification wanting to control the technical one in order to potentially expand it.

An interesting historical example is provided by the emergence of pathology technicians in the 1930’s. I quote from Starr (1982), p. 221.

The pathologist’s control of the laboratory business naturally gave them power over other laboratory workers. In 1929 the recently formed American Society of Clinical Pathologists, made up exclusively of physicians, began operating a system for certifying laboratory personnel. [...] The pathologists opposed any government licensing of technologists, which would have reduced their flexibility in the use of personnel.

This example, which appears to be somewhat typical in the medical field, illustrates the conflict of interest between two emerging complementary specialties: the physician-pathologist and the technician-pathologist. According to Starr, the licensing association of the discipline with the highest qualifications (physician-pathologists) wants to control the licensure of other (technician-pathologists), apparently to ensure a plentiful supply of
the latter’s services. (We are not told what the technicians felt about this, but I speculate that they may have preferred to have control over their own licensing standards.) In the language of the model I interpret the control of one association over the licensure of the other as them being the same association. Seen in this light, the evidence provided by Starr appears to be consistent with Proposition 10.

I am not convinced, however, that merely by controlling the laboratories the physician-pathologists would “naturally have power over” the technician-pathologist, as Starr puts it.\(^\text{15}\) The theory in this paper only points out the conflict between the two disciplines regarding associational boundaries; it does not tell us how this conflict might be composed, that is, whether a new association of technician-pathologists might spring up. The theory in this paper does suggest that the low-skilled specialty has no gain from controlling the high skilled licensing organization; a fact which is interesting I think, but stops short of necessarily predicting market structure.

7 Self-Regulation

A major rationale for our societal (and legal) tolerance of the anti-competitive features inherent in professional associations, is that these associations are able to impose quality standards on its members. Meeting a quality standard is privately costly for the individual member but, the argument goes, the association is willing to impose this cost on its members because the code benefits the profession as a whole. In other words, the professional association is able to self-regulate in a way that its individual members are not. This is probably true. The question I address here is whether the association chooses to self regulate. This, in principle, depends on the governance of the

\(^{15}\)It seems to me that, by the same logic, Detroit’s auto makers legal control over the factories would “naturally give them power” over the UAW, which was not the case for a long time.
In this section, as in the rest of the paper, I do not take for granted that the association behaves as a monolith. I ask instead whether there is heterogeneity of interests within the profession regarding the application of an ethics code. I consider, in particular, a rule or regulation the cost of which falls principally (in the model, solely) on one specialty within the profession, but the benefits of which are the same for all specialties. Given this asymmetry, one would guess that there might be a divergence of views within the association and that the regulation would be supported more strongly by the group which does not bear its cost. However, this is not the case in our model. Due to a “translation of costs” argument analogous to the analysis of tax incidence, it turns out that wages adjust to fully align the interests of both specialties. As a result, the model predicts perfect unanimity within the professional association with regards to the application of a code of professional ethics.

To focus on self-regulation, in this section I fix the size of the licensure $N$ and omit it from the notation. The extent of costly regulation is modeled as a scalar $r$ which raises the plaintiff’s cost which now are given by $c^i_p (r) \equiv c^i_p + r$. Increasing $r$ also increases the reputation of the profession as a whole, so that $V (r)$ is an increasing function of $r$. The additive scalar $r$ might capture the cost of increasing the professional educational requirement (more CLE courses for lawyers), or the professional liability insurance, or the reporting requirements, etc. The stark discrepancy between scopes of the regulation costs (only borne by plaintiff) and its benefits (enjoyed by all) is set up deliberately to highlight the “irrelevance of incidence” result.

For given $r$, we get the payoffs from expressions (2) and (3):

$$w^*_p (r) - c^i_p (r) \equiv \frac{V (r) + c^m + r}{2} - c^i_p - r,$$

$$w^*_d (r) - c^i_d \equiv \frac{V (r) - c^m - r}{2} - c^i_d.$$

Even though the direct costs of the regulation are borne by plaintiff lawyers
only, we see that the payoffs of plaintiff and defense lawyers are impacted in exactly the same way by a change in \( r \) (they vary at the same rate of \([V(r) - r]/2\). Therefore there is unanimous agreement within the whole profession about the net benefits from self regulation. This means that one cannot hope to leverage one side of the profession against the other, and that the form of governance does not matter with respect to self regulation. Let us denote the degree of self-regulation unanimously preferred by the association members by

\[
    r^* \equiv \arg \max V(r) - r.
\]

The degree of self-regulation unanimously preferred by the association members also maximizes the welfare of the association members. Indeed, the welfare function is

\[
    \int_{\frac{1}{2}}^{\infty} \frac{V(r)}{2} dF(i) - \left[ \int_{\frac{1}{2}}^{i_m} c_p^r(r) dF(i) + \int_{i_m}^{\infty} c_d^r dF(i) \right] = \frac{N}{2} V(r) - \frac{N}{2} r - \left[ \int_{\frac{1}{2}}^{i_m} c_p^r dF(i) + \int_{i_m}^{\infty} c_d^r dF(i) \right],
\]

which is a monotone transformation of \( V(r) - r \). These findings are collected in the following proposition.

**Proposition 11** Regarding how the costs of professional regulation are distributed between specialties, the entire profession will unanimously agree on the ideal amount of regulation. This amount maximizes the welfare of association members.

The stark result of zero heterogeneity of interests depends, in part, on the assumption that \( r \) enters additively as a cost. If \( r \) was not additive the message would be less stark. The nuanced interpretation of Proposition 11, then, is that whatever heterogeneity of interests there might be with respect to self-regulation, the heterogeneity stems from functional form assumptions and not from a deeper political economy reason. In particular, there is no
reason to believe that the association will be systematically biased in a particular direction concerning self-regulation.

8 Licensure Extension with Endogenous Self-Regulation

Proposition 11 implies that \( r^* \), the amount of self-regulation preferred by all association members, is a function of the extent of the licensure (summarized by \( N \)). Therefore, if \( N \) varies then so does \( r^* \). In this section we return to the problem of licensure extension analyzed in Section 4, but this time taking into account the endogenous adjustment in professional regulations.

Let us define \( V (N, r) \) as the value of a case which now depends negatively on \( N \) and positively on \( r \). Denote

\[
\begin{align*}
  r^* (N) &\equiv \text{arg max } V (N, r) - r. \\
\end{align*}
\]

The change in the incumbent defense wages as a function of a change in the admission threshold is now given by (cf. expression 6):

\[
\begin{align*}
  &\frac{1}{2} \left( \frac{\partial V (N, r)}{\partial N} \frac{\partial N}{\partial r} + \left( \frac{\partial V (N, r)}{\partial r} - 1 \right) \frac{\partial r^* (N)}{\partial N} \frac{\partial N}{\partial r^* (N)} - \frac{\partial \epsilon}{\partial i} \right) \bigg|_{r = r^* (N)} .
\end{align*}
\]

Since

\[
\frac{\partial V (N, r)}{\partial r} - 1 = 0 \text{ at } r = r^* (N),
\]

it follows that expression (10) coincides with expression (6). The implications are collected in the following proposition.

**Proposition 12** The majority is equally likely to approve of a small extension in the licensure when the changes in self-regulation are anticipated, as when regulation is exogenously fixed.
In other words, taking into account future changes self-regulation does not change the attitude of association members towards licensure expansion.

9 Social Welfare

Like in many political economy models, here too the connection is tenuous between the policies favored by (a majority of) the selectorate and those policies that benefit society as a whole. A major source of “social welfare ambiguity” in the model is the quantity $V$, which captures the value of a lawsuit to the lawyers. $V$ may be smaller than the social value of the lawsuit (if, for example, lawyers are only able to capture a minute amount of the value of the lawsuit to their clients and to society as a whole);\footnote{A lawsuit may have value to society as a whole through its precedential value.} or it may be larger, which could happen in the case of frivolous lawsuits the outcome of which entails large transfers from defendant to plaintiff, and thus potentially a large $V$, but whose social value is minimal. In other professions the situation may be different. Among doctors, perhaps, it could be argued that $V$ tracks social welfare more closely. In light of this ambiguity, the normative question of welfare analysis can, in my view, only be settled by empirical work. I believe, however, that the positive model I presented can help structure the empirical analysis of the normative question.

10 Conclusion

Professional licensing covers a large fraction of workers throughout the world. Since the free entry model does not describe these labor markets, it important to know how access to these market is governed. It is generally assumed that licensing associations will inevitably want to restrict entry.

This paper challenges the inevitability of this logic. I looked closely at
the internal incentives of a licensing association to expand. When the association comprises complementary specialties, there may be heterogeneity of interests within the association about the benefits of expansion, including a majority of members favoring expansion. Expansion may take place even beyond the level that maximizes the incumbents’ rents, and possibly even beyond the socially optimal level. This happens because expanding the licensure entails redistribution among sub-specialties. In principle, then, the power to license may be used to expand the association excessively. I think this is an important point because in the policy debate the entry-restricting behavior of licensing associations is usually implicitly assumed, as if requiring no demonstration. This, I believe, is because of the absence of an alternative paradigm. In this paper I have provided such a paradigm. I hope that this alternative can lead to a more careful, evidence-based assessment of the use of licensure. For example, licensing associations may behave quite differently from labor unions, with which licensing associations are often analogized.

The analysis also points to a taxonomy of licensing organizations, depending on whether (or how much) they comprise complementary specialties. According to the analysis in this paper, broader-scoped organizations are more likely to have conflict of interest regarding expansion, and may be prone to over-expansion. To the extent that governance mediates conflicts of interest, we should expect governance to matter more to members of broader-scoped organizations.

The model also predicts that, as division of labor creates spins off relatively low-skilled occupations (lab technicians, e.g.) from high skilled specialties, (physicians), the high-skilled licensing association would want to prevent the creation of a separate low-skilled licensing association.

This paper also dealt with the internal incentives of a licensing association to self-regulate with regards to quality standards. The paper predicts no conflict of interest within the organization regarding such self-regulation. If this is true, then we should observe the governance of licensing associations
to be less concerned with self-regulation, compared with licensure extension.

The paper’s singular focus on the majority of incumbents was motivated by the special role that the majority plays in many forms of governance; and it makes for sharp results. However “majority” should not be taken literally; after all, most organizations do not formally vote on expanding the licensure. The same goes for the Leontieff technology (exactly one defense for each plaintiff laywer). Reality is more nuanced. The nuanced message of this paper is that licensing organizations need not all be against expansion at all times. Simple as it is, this message is new as far as I know. Of course, the empirical relevance of this message depends in part on the encroachment effect, as I called it, not being too strong. In Section 2.1 I presented some evidence that, in some circumstances, new entrants can bring with them considerable new business, and so the encroachment effect may not be very strong.

Finally, I emphasize that this paper only looks at the internal politics of licensing organizations. Their external politics, and the politico-administrative ecology in which these organizations live, are equally fascinating but are not the object of this study.
References


