Monopolistic Group Design with Peer Effects

Simon Board*

This Version: January 14, 2007
First Version: October 2005

Abstract

In a range of settings, private firms manage peer effects by sorting agents into different groups, be they schools, neighbourhoods or teams. This paper considers such a firm, which controls group entry by setting a series of anonymous prices. We show that private provision systematically leads to two distortions relative to the efficient solution: first, agents are segregated too finely; second, too many agents are excluded from all groups. We demonstrate that these distortions are a consequence of anonymous pricing and do not depend upon the nature of the peer effects. This general approach also allows us to assess the way the 'returns to scale' of peer technology and the cost of group formation affect the optimal group structure.

1 Introduction

In an increasingly privatised world, for–profit organisations have come to play an important role in many markets where peer effects are prominent. This paper considers such a market, where a firm posts a series of prices and agents self–select into different groups. The quality of a group, in turn, depends on the characteristics of its members. We show that private provision systematically leads to two distortions in group formation relative to the efficient solution. First, there is too much segregation between different types of agents; that is, groups are excessively homogenous. Second, too many agents are excluded from all available groups.

The model captures the key features of several important markets. First, consider the market for education, where peer effects play an important role in shaping students' goals and learning experience. In such a market, firms can manage peer effects to their advantage by

*Department of Economics, University of Toronto. http://www.economics.utoronto.ca/board. I have received helpful comments from Richard Arnott, Heski Bar–Isaac, Jeremy Bulow, Ed Lazear, John McMillan, Rob McMillan, Andreas Park, Will Strange, Bob Wilson and Jeff Zwiebel. I also thank seminar audiences at Brock, McMaster, Stanford, Toronto, York and ESEM 2006. JEL: D82, H40, L12. Keywords: mechanism design, peer effects, public goods.

charging more for courses which attract above–average students. This type of differentiation is commonplace: providers of higher education and professional training consistently use peer effects to price–discriminate between different institutions and different courses of otherwise similar quality. The growing popularity of vouchers also promises to raise the importance of private primary and secondary schools, which will similarly seek to manipulate peer effects. Epple and Romano (1998) and Caucutt (2002), among others, have investigated the role of selection when schools are highly competitive. This paper analyses the optimal pricing policy for a school with market power.[1]

The second application concerns the market for community formation, where peer effects are a major determinant or consumers' preferences. In recent years, this market has seen a significant expansion in the role of the private sector, with more than forty million Americans currently living in common interest developments (e.g. condominiums, planned unit developments). Proponents argue that these new communities increase welfare by providing safety and comfort for those willing to pay; critics counter that they are discriminatory and isolationist. Our model is consistent with both these arguments, showing how group formation increases welfare, but also that private provision leads to communities that are insufficiently diverse.[2]

Thirdly, peer effects play an important role in the theory of the firm. According to Alchian and Demsetz (1972), facilitating teamwork is a major activity, and perhaps even the defining property, of a firm. While the composition of a team is often taken as exogenous, firms will seek to assemble compatible agents in order to maximise their productivity and minimise their wage bill. This paper thus analyses optimal team formation within a firm, examining how different types of peer effects alter group composition.

One significant problem in analysing peer effects is that the nature of peer technology is likely to differ greatly across environments. In a recent survey on the role of private education, Helen Ladd (2002, p. 14) wrote:

> "This lack of clarity about how peer effects differ among groups rules out any clear predictions about whether a voucher program would be likely to increase or decrease the overall productivity of the education system through the mechanism of peer effects".

Despite this concern, we analyse the distortions induced by private provision while placing very little structure on the nature of peer effects. This general approach enables us to examine how

[1]Due to transportation costs, schools and universities already possess considerable local market power. As private schools become more popular, it is also likely that chains, such as Edison in the U.S. and GEMS in Britain, will become increasingly powerful.

[2]The model applies to many other types of communities: restaurants, golf clubs and luxury good manufacturers all seek to affect the attractiveness of their product through exclusivity. For example, Kaneff owns six golf courses in Ontario, charging a range of fees, separating different types of customers into groups. See Rayo (2005) for other examples.

the degree of segregation depends on the form of peer effects. It also helps us interpret the recent empirical literature quantifying peer effects in different environments.

1.1 Outline of the Paper

The basic structure of the model is as follows. First, a single principal posts a range of anonymous group–entry prices. Agents vary in their willingness to pay for group quality and, after observing these prices, sort themselves into different groups. The quality of a group, in turn, depends upon the types of its members. This quality function is allowed to be very general and subsumes the average quality model (e.g. Rayo (2005)), the Cobb–Douglas quality model (e.g. Epple and Romano (1998)) and the multiplicative quality model (e.g. Lazear (2001)).

Since pricing is anonymous, the principal must rely on agents to self–select into different groups. Self–selection immediately implies that that agents who care more about group quality must be in better quality groups (the monotonicity condition). This result implies that if the agents who generate high quality have a low willingness to pay, then the principal must assign all agents to identical groups. Conversely, if the agents who generate high quality have a high willingness to pay, then the principal can segregate the agents into groups of different standards.

The paper first analyses the principal's problem when group formation is costless, showing that profit–maximisation leads to two distortions relative to the welfare–maximising group structure. The first distortion, the *segregation effect*, states that there are too many groups under profit–maximisation. Intuitively, by splitting a group into two, putting all high types into one group and the low types into another, the principal increases the price the high types are willing to pay in order to avoid the low quality group. Crucially, we do not require any assumptions on the nature of peer effects in order to attain this result: the required restrictions come endogenously from the requirement that agents self–select into groups. This segregation effect implies that the distribution of group qualities under private provision has a lower mean and will tend to be more dispersed than the efficient distribution.

The second distortion, the *exclusion effect*, states that too many agents are excluded from all privately provided groups. The exclusion effect is analogous to the standard result that a monopolist prices above marginal cost. Intuitively, excluding some low types of agents raises the price paid by those agents who are not excluded.

We further analyse how the optimal group structure depends upon the nature of peer interactions. When a quality function has decreasing returns to scale, in that splitting a group into two subgroups raises the average quality, then welfare and profit are maximised by complete separation. That is, every type is in a group of his own,[3] so agents associate with those just like themselves and ignore everyone else. Conversely, when a quality function has increasing returns

[3]The principal is female, while agents are male.

to scale, in that splitting a group into two subgroups lowers the average quality, then matching will be assortative (i.e. groups will be connected) and there will tend to be some pooling.

The paper also examines the principal's problem when group formation is costly. This setting introduces a new factor, the *appropriability effect*, according to which a welfare–maximising principal may invest more in group formation than a profit–maximising principal. Intuitively, a profit–maximiser cannot appropriate agents' consumer surplus and may opt for larger groups than is optimal. Nevertheless, under increasing returns to scale and the usual monotone hazard rate condition, the segregation effect dominates the appropriability effect and groups are smaller under profit–maximisation.

We also investigate how welfare– and profit–maximising group structures change with relative position. This is motivated by Lazear (2001) who argues that more able students will tend to be in larger classes. In our model, when group formation is costless, we also find that higher types will tend to be in larger groups, albeit for a very different reason. The intuition behind our result is that the ratio between the highest and lowest types in a group declines as everyone's type rises. This means a group split, which helps the high types but hurts the low types, becomes less desirable. This suggests that selective grammar schools were of more use in the 1950s, when education levels were relatively low, than today. In comparison, Lazear's finding derives from the specifics of the multiplicative quality model, under which returns to scale increase in agents' types

The paper proceeds as follows. The remainder of this section provides a literature review, while Section 2 considers a simple two–type example that captures a number of the main effects. Section 3 describes the model. Section 4 assesses the implications of self–selection. Sections 5 and 6 analyse the costless group formation problem, deriving the segregation and exclusion effects. Section 7 examines the costly group formation problem, comparing appropriability and segregation effects. Section 8 derives conditions under which higher types are in larger groups, and Section 9 concludes.

1.2 Literature

It is helpful to break the peer group literature into three branches.

The first branch considers a single principal with perfect information about agents' characteristics. In their classic paper, Arnott and Rowse (1987) analyse the socially optimal way to break students into N groups in the presence of peer effects. A student's utility is a function of his ability, the mean ability of the other students in the class and educational expenditure. Using a Cobb–Douglas quality function, the authors obtain sufficient conditions for assortative matching and computationally solve several examples. Lazear (2001) considers a highly tractable model where each student is disruptive with probability p. If there are m students in the class who act independently of each other then the class is attentive proportion $(1-p)^m$

of the time. Lazear shows that a welfare–maximising school increases class sizes as p increases and, in a two–type model, will segregate students by type.[4]

In the second branch, there is a single principal with imperfect information about agents' characteristics. Helsley and Strange (2000) analyse common interest developments with social interactions. Agents, who vary in their type, choose whether to stay in the public sector or join a single private community, and subsequently choose an action. Agents' utility then depends upon their action, their type and the mean action of those in their community. Helsley and Strange allow the private community to choose both a minimum required action and an entry price. In a numerical example they show fewer people secede from the public sector when the community is profit–maximising, in a similar spirit to our exclusion effect.

The two closest papers to the current one both consider a principal who price discriminates between agents by sorting them into different groups of different qualities. Rayo (2005) considers a one–sided matching problem, similar to ours, where the principal breaks the agents up into groups. Rayo uses the average–quality function and investigates the role of non–monotone marginal revenue functions (see Section 5.4). Damiano and Li (2006) analyse a two–sided matching market where the principal can discriminate between different sides of the market and between different groups. Damiano and Li derive necessary and sufficient conditions for full separation.[5]

The third branch analyses competition between peer groups. Epple and Romano (1998) analyse a model of private school competition, where agents differ in their income and ability, both of which are publicly observable. Epple and Romano show that monopolistic competition between schools with fixed costs leads to stratification of the market where poor talented agents attend the same schools as wealthy untalented agents. Caucutt (2002) introduces educational expenditure and shows that complete segregation may not be desirable, even without fixed costs of setting up schools. Intuitively, a school can keep its quality constant by lowering its expenditure on teachers but recruiting a few talented students.[6]

The discussion of the empirical literature is postponed until Section 3.1.

[4] In a related model, Kremer (1993) considers a groups of agents on a production line who each make a mistake with probability p_i. In the competitive equilibrium, there is assortative matching and higher quality workers work in longer production lines.

[5] Also relevant is Pesendorfer (1995) who supposes that status is driven by a two–sided matching problem, where a durable status good is sold by a monopolist. Pesendorfer argues that the firm will regularly introduce new designs if they cannot commit to a price path or if there is imitation.

[6] One can view these papers of applications of club theory (e.g. Scotchmer (2002)). Related papers include Nechyba (2000) and Benabou (1993). For a model with imperfect information see Damiano and Li (2005).

2 Two–Type Example

There are equal numbers of two types of agents, $\theta_H > \theta_L$, where an agent's type describes his willingness to pay for quality. The utility of type θ_i who is assigned to a group of quality $Q(\theta_i)$ and pays price $y(\theta_i)$ is given by $u(\theta_i) = \theta_i Q(\theta_i) - y(\theta_i)$, for $i \in \{L, H\}$. The quality of a group is determined by the types of its members. A group consisting of θ_H agents has quality Q_H; a group consisting of θ_L agents has quality Q_L; and a group consisting of both types has quality Q_{LH}. An agent's outside option is 0. Finally, we suppose that agents are small, so no individual agent can affect the quality of a group.

The principal posts anonymous group–entry prices and lets agents self–select into the different groups. This means that, in order to stop the high type copying the low type, we must have $Q(\theta_H) \geq Q(\theta_L)$ (the monotonicity condition). Consequently, the principal can separate the agents if and only if $Q_H \geq Q_L$; otherwise a high type would enter the low type's group rather than his own.

2.1 Segregation Effect

Let us first consider the principal's incentive to separate the two types of agents. For simplicity, assume that $2\theta_L \geq \theta_H$ and that the principal does not exclude either type. Utility is quasi–linear, so welfare equals $\theta_L Q(\theta_L) + \theta_H Q(\theta_H)$. A welfare–maximising principal would therefore like to separate the agents when

$$\theta_H Q_H + \theta_L Q_L \geq \theta_H Q_{LH} + \theta_L Q_{LH} \tag{2.1}$$

Define Q_{LH}^W as the pooling quality that equates both sides of (2.1). If $Q_H < Q_L$, the principal can only pool the agents. If $Q_H \geq Q_L$, then the principal will separate the agents when $Q_{LH} \leq Q_{LH}^W$. Since $Q_{LH}^W \geq (Q_H + Q_L)/2$, separation is optimal if the quality function is increasing, $Q_H \geq Q_L$, and satisfies *decreasing returns to scale*, in that $Q_{LH} \leq (Q_H + Q_L)/2$.

A profit–maximising principal maximises total payments, $y(\theta_L) + y(\theta_H)$. If the principal pools both types, she will charge $y(\theta_L) = y(\theta_H) = \theta_L Q_{LH}$ in order to fully extract from the low type, θ_L. On the other hand, if the principal separates both types, she will charge $y(\theta_L) = \theta_L Q_L$ to the low group and $y(\theta_H) = \theta_L Q_H - (\theta_H - \theta_L) Q_L$ to the high group. Under these prices, the low type is just willing to join the low group, while the high type is indifferent between joining the high and low groups. Putting this together, the profit–maximising principal would like to separate the agents when

$$\theta_H Q_H + (2\theta_L - \theta_H) Q_L \geq 2\theta_L Q_{LH} \tag{2.2}$$

Define Q_{LH}^Π as the pooling quality that equates both sides of (2.2). If $Q_H < Q_L$, then the principal can only pool the agents. If $Q_H \geq Q_L$, then $Q_{LH}^\Pi \geq Q_{LH}^W$, so a profit–maximising

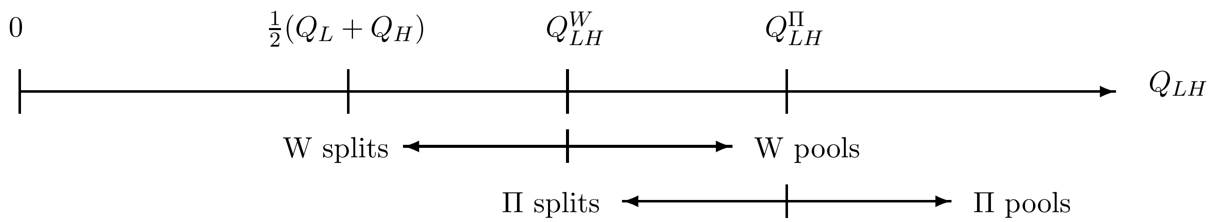

Figure 1: Two–Type Model with $Q_H \geq Q_L$.

principal is more willing to separate the agents than a welfare–maximising principal (see Figure 1). Intuitively, by separating high and low types, the good agents become very keen to avoid the bad agents and can be forced to pay higher prices. Notice that this *segregation effect* requires no assumptions about the structure of qualities (Q_L, Q_H, Q_{LH}): the fact that $Q(\theta_H) \geq Q(\theta_L)$ follows from the endogenous self–selection constraint.

2.2 Exclusion Effect

If $2\theta_L < \theta_H$, then the profit–maximising principal may wish to exclude the low types in order to increase revenue. To see this, consider the case where $Q_H \geq Q_L$.[7] The welfare–maximising principal never excludes any type of agent, and separates the two types if (2.1) holds. In contrast, the profit–maximising principal may wish to exclude the low type, enabling her to charge $y(\theta_H) = \theta_H Q_H$ to the remaining high types. She therefore wishes to separate the two types if

$$\theta_H Q_H + \max\{2\theta_L - \theta_H, 0\} Q_L \geq 2\theta_L Q_{LH} \tag{2.3}$$

As above, (2.1) implies (2.3). This shows that the segregation effect extends to the case where we allow exclusion. Moreover, a profit–maximising principal is more willing to exclude agents than a welfare–maximising principal. This *exclusion effect* is analogous to the standard monopoly distortion: by cutting out low types the principal increases the price she can charge the high types.

2.3 Appropriability Effect

So far we have assumed that splitting the agents into two groups is free of charge. Costly group formation introduces a third effect. To illustrate, let us assume that $Q_H \geq Q_L$. Using equation (2.1), the benefit of separation for a welfare–maximising principal is

$$\theta_H(Q_H - Q_{LH}) + \theta_L(Q_L - Q_{LH}) \tag{2.4}$$

[7]This assumption is not necessary. If $Q_H < Q_L$ then the principal may wish to exclude the low type in order to 'monotonise' the quality function (see Section 6). However, it is straightforward to show that both the segregation and exclusion effects continue to apply.

If $2\theta_L \geq \theta_H$, equation (2.2) implies that the benefit of separation for a profit–maximising principal is

$$\theta_H(Q_H - Q_{LH}) + (2\theta_L - \theta_H)(Q_L - Q_{LH}) \tag{2.5}$$

Hence, if group formation is costly, and there are very strong decreasing returns to scale, $Q_L \geq Q_{LH}$, then the welfare–maximising principal is willing to pay more to separate the agents than the profit–maximising principal. This *appropriability effect* is caused by the profit–maximising principal's inability to appropriate the agents' consumer surplus. However, when there are sufficient returns to scale, then the segregation effect outweighs the appropriability effect and the profit–maximising principal is more likely to separate the groups.

3 Basic Model

Agents' Preferences. A single principal faces a continuum of agents with privately known willingness to pay $\theta \in [\underline{\theta}, \overline{\theta}] \subset \mathbb{R}_+$. Types are distributed according to positive density $f(\theta)$ with distribution function $F(\theta)$. Agents then choose to join one of the available groups, or choose not to participate. If agent θ joins a group of quality Q and pays price y, he obtains utility

$$u = \theta Q - y.$$

If an agent chooses not to participate, he obtains zero utility.

Principal's Problem. The principal first chooses a series of group–entry prices; agents subsequently self–select into groups $G \subset [\underline{\theta}, \overline{\theta}]$. Applying the revelation principle, we analyse the direct revelation mechanism $\langle \mathcal{G}, y \rangle$ whereby agents announce their types, and the principal assigns them to a group $G \in \mathcal{G}$ and charges a fee y. Given any equilibrium in the price–setting game, then there exists a corresponding direct revelation mechanism such that all agents accept the mechanism (individual rationality) and all agents announce their types truthfully (incentive compatibility). The *principal's problem* is thus to choose the mechanism $\langle \mathcal{G}, y \rangle$ to maximise welfare/profits subject to individual rationality and incentive compatibility.

Groups. The principal breaks the agents into groups $\mathcal{G}$. For technical reasons, we restrict how the principal can break up the agents. Let the collection of sets $\mathcal{P}$ be a finite partition of the type space; that is, a collection of nonintersecting connected sets whose union equals $[\underline{\theta}, \overline{\theta}]$. A *group* G is then the union of sets lying in $\mathcal{P}$. A *group structure* $\mathcal{G}$ is a collection of nonintersecting groups whose union equals $[\underline{\theta}, \overline{\theta}]$. Taking two group structures, $\mathcal{G}_L$ and $\mathcal{G}_H$, we write $\mathcal{G}_L \subset \mathcal{G}_H$ if $\mathcal{G}_H$ is finer than $\mathcal{G}_L$. Two groups, G and G', *overlap* if there exists $\theta_H > \theta_M > \theta_L$, such that $\theta_H, \theta_L \in G$ and $\theta_M \in G'$, or $\theta_H, \theta_L \in G'$ and $\theta_M \in G$. Denote the sigma–algebra of a group structure by $\sigma(\mathcal{G})$; since $\mathcal{P}$ is finite, this equals the collection of unions of sets in $\mathcal{G}$.

Peer Technology. Each group G is associated with a quality $Q(G) > 0$. Let $Q(\theta, \mathcal{G})$ denote the quality of type θ's group under group structure $\mathcal{G}$. A quality function $Q(G)$ is said to be *weakly increasing* in G if $Q(G_H) \geq Q(G_L)$ whenever G_H is larger than G_L in the sense that $\theta \geq \theta'$ for all $\theta \in G_H$ and $\theta' \in G_L$.

Some remarks are in order. First, we do not insist that groups be connected. This is important because the optimal group structure may place agents with a wide range of abilities in the same group, as suggested by the empirical work of Mas and Moretti (2006) and experimental study of Falk and Ichino (2006).

Second, the model restricts groups to be unions of sets in some underlying finite partition, $\mathcal{P}$. This assumption is for technical simplicity. It enables us to avoid measure–theoretic problems such as the creation of non–measurable sets. It also guarantees the problem has an optimal solution. This restriction is analogous to having a finite number of types, although the continuous type representation remains useful.[8,9]

Third, in the model above, we assume that the principal places each agent into a group. That is, we assume it is not optimal for the principal to exclude any types of agents. This assumption is for simplicity: we extend the analysis in Section 6.[10]

Finally, we say a function $\phi : \mathbb{R} \to \mathbb{R}$ is *quasi–increasing* if $\phi(x_L) \geq 0$ implies $\phi(x_H) \geq 0$ for $x_H > x_L$, and *weakly quasi–increasing* if $\phi(x_L) > 0$ implies $\phi(x_H) \geq 0$.

3.1 Group Quality Functions

The paper allows for a large range of quality functions, $Q : G \to \mathbb{R}_{++}$, subsuming those used in a number of previous papers. This level of generality is particularly important since the peer technology depends on the specific environment and is hard to quantify in any given application. Some examples of quality functions are as follows:

- Average–quality: $Q(G) = E[\theta | \theta \in G]$. This states that the quality of a group is given by the average type of its members. This is used by Rayo (2005) and matching papers such as Damiano and Li (2006).

[8]In Examples 10–11 we drop the finiteness restriction on the principal's choice set and let $\sigma(\mathcal{P})$ equal the Borel sets, enabling the use of calculus.

[9]We also assume that different agents of a single type are assigned to the same group. However, since groups may be disconnected, any mixed strategy can be approximated by such a pure strategy. For example, if $\theta \sim U[0,1]$, then a group with measure 1/2 on [0,1/2] can be approximated by a group with measure 1 on $[0, 1/8] \cup [3/8, 4/8]$. This restriction is therefore minor if the quality functional, which maps measures on the type space to the real line, is continuous in, say, the Prohorov metric.

[10] The assumption of no exclusion is without loss of generality if the principal's objective, $MR(\theta)$, is positive $(\forall\theta)$. In this case, one can define the quality function so that $Q(G) = 0$ if $\underline{\theta} \in G$. Pooling agent θ with type $\underline{\theta}$ is then equivalent to excluding θ. This 'evil type' model does not work if $MR(\theta)$ is negative since the objective fails to be log–supermodular. Consequently, when we allow $MR(\theta)$ to be negative in Section 6, we use a different approach.

- Average–quality with crowding out: $Q(G) = E[\theta|\theta \in G] - mE[\mathbf{1}_G]$, where $m > 0$ represents the importance of crowding out.

- Generalised average–quality: $Q(G) = \phi_1(E[\phi_2(\theta)|\theta \in G])$. As a special case, this includes the Cobb–Douglas quality function, $Q(G) = E[\theta^{1/\alpha}|\theta \in G]^\beta$, which is used by Epple and Romano (1998), Nechyba (2000), Caucutt (2002) and the latter parts of Arnott and Rowse (1987).

- Linear–quality: $Q(G) = \alpha \sup(G) + \beta \inf(G)$. One special case of this is min–quality, $Q(G) = \inf(G)$, where the group is only as good as its worst member. Another special case is max–quality, $Q(G) = \sup(G)$, where the best agent becomes the "leader" of the group.

- Multiplicative–quality: $Q(G) = \exp\left(-m \int_G (1-\theta) dF(\theta)\right)$, where $m > 0$. As shown in Appendix A.1, this is a continuous analogue of the production functions in Kremer (1993) and Lazear (2001).

There is a large empirical literature which seeks to estimate peer technology. While it is hard to generalise, the magnitude of these peer effects can be substantial. In the classroom, Henderson, Mieszkowski, and Sauvageau (1978) find that moving a student from a weak class to a strong class can increase their overall rank from the 50^{th} percentile to the 20^{th} percentile. In the workplace, Mas and Moretti (2006) and Falk and Ichino (2006) find a 10% increase in one's colleagues productivity raises a given worker's productivity by around 1.5%. This literature has analysed three major aspects of the production function.

First, nonlinearities in the peer technology. Looking at college roommates, Zimmerman (2003) finds that bad students have a bigger effect on their roommates than good students. However, in a similar study, Sacerdote (2001) finds the converse: bad students seem to have a smaller effect on their roommates than good students. We will see that the former is an example of decreasing returns to scale, implying that bad students should be segregated; while the latter is an example of increasing returns to scale, implying that some mixing of abilities is optimal (see Proposition 2). In a similar spirit, Henderson, Mieszkowski, and Sauvageau (1978) find that test scores are a concave function of mean class ability. These results are consistent with a generalised average–quality function where $\phi_1(\cdot)$ is concave, implying increasing returns to scale.

Second, interaction effects. In their workplace studies, Mas and Moretti (2006) and Falk and Ichino (2006) find that having good peers have a stronger effect on poor workers. In contrast, with college roommates, Zimmerman (2003) finds that peer effects have the biggest impact on students of middling ability. Looking at the classroom, Henderson, Mieszkowski, and Sauvageau (1978) and Hanushek et al. (2003) report that there are few cross effects. This latter result implies that, if all students care equally about their test scores, then it is impossible to separate different types of agents (see Lemma 2). However, if high ability students care more about their test scores than low ability students, then separation can be sustained.

Third, scale effects. In the education literature, there has been a long standing debate about the impact of reductions in class size. While the desirability of small classes may seem obvious, the evidence seems to find beneficial effects only in certain environments (Hanushek (1999)). In the workplace, Falk and Ichino (2006) find agents are more productive at stuffing envelopes when they work in the presence of others, although this result clearly depends on the task at hand.

Looking across these studies, it seems that the peer technology can vary greatly with the environment. This observation has two important implications. First, it is important to derive results that do not depend on the exact nature of the peer effects. To illustrate, both the multiplicative quality and Cobb–Douglas quality are widely used models. However, while the multiplicative model predicts that more able agents should be in larger groups (Lazear (2001)), the Cobb–Douglas model predicts the opposite (Figure 4). The second implication is that theory should identify which aspects of the peer technology are critical for a given result, rather than working with a single functional form, which contains many hidden assumptions. This approach both helps us categorise different classes of peer technologies, and helps us understand what to look for in the data.

4 Agents' Problem

The principal runs a direct revelation mechanism $\langle \mathcal{G}, y \rangle$, in which an agent of type θ declares that they are type $\hat{\theta}$, receives quality $Q(\hat{\theta}, \mathcal{G})$ and makes payment $y(\hat{\theta})$. Since there are a continuum of agents, the quality of an agent's group depends on his declaration but not his type. Utility is then

$$u(\theta, \hat{\theta}) = \theta Q(\hat{\theta}, \mathcal{G}) - y(\hat{\theta}) \tag{4.1}$$

Define equilibrium utility to be $U(\theta) = u(\theta, \theta)$.

Lemma 1. *A mechanism $\langle \mathcal{G}, y \rangle$ is incentive compatible and individually rational if and only if:*
(a) Utility is given by

$$U(\theta) = \int_{\underline{\theta}}^{\theta} Q(s, \mathcal{G})\, ds + U(\underline{\theta}) \tag{4.2}$$

(b) The lowest type obtains $U(\underline{\theta}) \geq 0$; and
(c) The monotonicity condition holds. That is, $Q(\theta, \mathcal{G})$ is increasing in θ.

Proof. Since $Q(\theta, \mathcal{G})$ is integrable, Milgrom and Segal (2002, Corollary 1) shows that incentive compatibility implies (4.2). The rest of the proof is the same as Mas-Colell, Whinston, and

Green (1995, Proposition 23.D.2). ☐

Lemma 2. *In any incentive compatible group structure:*
(a) Any overlapping groups have the same quality.
(b) If $Q(G)$ is weakly decreasing then every agent will be in a group of the same quality.

Proof. Follows from the monotonicity condition (Lemma 1(c)). ☐

Lemma 2(a) says that while groups do not have to be connected, any overlapping groups must have the same quality. Lemma 2(b) says that the principal cannot separate different types when the agents who generate high quality have a low willingness to pay. This may be the case in the workplace if good workers most improve the performance of poor workers (e.g. Mas and Moretti (2006), Falk and Ichino (2006)). Separation may also be difficult with some conspicuous goods, where agents seek to signal a certain image. For example, the consumers who generate Harley–Davidson's reputation are unlikely to have the highest incomes. Similarly, the supporters with the highest willingness to pay for football tickets may not create the best atmosphere.[11]

5 The Segregation Effect

5.1 Principal's Problem

Welfare equals the sum of utilities plus transfers,

$$W = E[\theta\, Q(\theta, \mathcal{G})] \tag{5.1}$$

Integrating utility (4.2) by parts, consumer surplus is

$$E[U(\theta)] = E\left[\frac{1 - F(\theta)}{f(\theta)}\, Q(\theta, \mathcal{G})\right] + U(\underline{\theta}) \tag{5.2}$$

Profit equals welfare (5.1) minus consumer surplus (5.2). The profit–maximising principal will set prices so that the lowest type's individual rationality constraint binds, $U(\underline{\theta}) = 0$. Profit is then given by

$$\Pi = E\left[MR(\theta)\, Q(\theta, \mathcal{G})\right] \tag{5.3}$$

where marginal revenue is defined by

$$MR(\theta) := \theta - \frac{1 - F(\theta)}{f(\theta)}$$

[11]In both these examples the firms use non–price mechanisms to maintain quality. Harley–Davidson uses waiting lists, while football clubs force supporters to buy season tickets.

Welfare and profit can be thus combined into a single objective:

$$H = E\left[h(\theta)\, Q(\theta, \mathcal{G})\right] \tag{5.4}$$

where $h(\theta) \in \{\theta, MR(\theta)\}$. Let Γ be the set of group structures that satisfy the monotonicity condition (Lemma 1(c)). The *principal's problem* is then to choose $\mathcal{G} \in \Gamma$ to maximise (5.4).

The choice set Γ is finite, so a solution to the principal's problem exists. Nevertheless, there are two difficulties with this maximisation problem. First, Γ is not generally a lattice. Second, $Q(\theta, \mathcal{G})$ is unlikely to be quasi–supermodular in $\mathcal{G}$. Intuitively, two different ways of splitting a group are likely to be substitutes rather than complements. Consequently, the optimal set of groups structures may not be a lattice.

5.2 Welfare– and Profit–Maximisation

For a fixed group structure $\mathcal{G}$, let $I_{\mathcal{G}}(G)$ be the smallest interval containing G that is made up of elements of $\mathcal{G}$.[12] By Lemma 2(a), quality must be constant over all groups in $I_{\mathcal{G}}(G)$. Let $\mathcal{I}(\mathcal{G})$ be the partition formed by collecting the intervals $\{I_{\mathcal{G}}(G)\}_{G \in \mathcal{G}}$. Equivalently, let $\mathcal{I}(\mathcal{G})$ be the partition induced by merging all overlapping groups in $\mathcal{G}$.

Lemma 3. *$\mathcal{G}_L \subset \mathcal{G}_H$ implies $\mathcal{I}(\mathcal{G}_L) \subset \mathcal{I}(\mathcal{G}_H)$.*

Proof. See Appendix A.2. □

Assumption (MON). $[1 - F(\theta)]/\theta f(\theta)$ is decreasing in θ.

This assumption implies that $MR(\theta)$ is quasi–increasing. It is weaker than the usual hazard rate condition (see Section 7).

Proposition 1 (Segregation Effect)**.** *Suppose (MON) holds and $MR(\underline{\theta}) \geq 0$. For any welfare–maximising solution, $\mathcal{G}^W$, $\Pi(\mathcal{G}^W) \geq \Pi(\mathcal{G})$ on $\{\mathcal{G} \in \Gamma : \mathcal{G} \subset \mathcal{G}^W\}$. Hence if any optimal solutions, $\mathcal{G}^W$ and $\mathcal{G}^\Pi$, are ordered in terms of set inclusion, then there exists a profit–maximising solution, $\mathcal{G}^{\Pi*}$, such that $\mathcal{G}^W \subset \mathcal{G}^{\Pi*}$.*

Proof. Suppose $\mathcal{G}^W$ maximises welfare and fix $\mathcal{G} \in \Gamma$ such that $\mathcal{G} \subset \mathcal{G}^W$. Since $\mathcal{G}^W$ is welfare–maximising, $E[\theta \Delta Q(\theta)] \geq 0$, where $\Delta Q(\theta) := Q(\theta, \mathcal{G}^W) - Q(\theta, \mathcal{G})$. Define $\mathcal{I}^*$ to be the coarsest partition on which $\Delta Q(\theta)$ is quasi–increasing. Applying Lemma 3, $\mathcal{I}(\mathcal{G}) \subset \mathcal{I}(\mathcal{G}^W)$. Monotonicity thus implies that $\Delta Q(\theta)$ is increasing on each $I \in \mathcal{I}(\mathcal{G})$, so $\mathcal{I}^* \subset \mathcal{I}(\mathcal{G})$. See Figure 2 for an illustration. The proof is now based on two steps.

[12] Formally, $I_{\mathcal{G}}(G)$ is the smallest interval in $\sigma(\mathcal{G})$ containing G. This is uniquely defined.

For the first step, we claim that $E[\theta\Delta Q(\theta)|\mathcal{I}^*] \geq 0$.[13] To see this suppose, by contradiction, that $E[\theta\Delta Q(\theta)|\mathcal{I}^*] < 0$ on some set $A \in \sigma(\mathcal{I}^*)$. Then define a new group structure, $\mathcal{G}'$, equal to $\mathcal{G}$ on A and $\mathcal{G}^W$ elsewhere. This new structure has two properties. First, $\mathcal{G}'$ has higher welfare than $\mathcal{G}^W$, $E[\theta Q(\theta,\mathcal{G}')] > E[\theta Q(\theta,\mathcal{G}^W)]$. Second, $\mathcal{G}' \in \Gamma$, which we verify below. Together, these contradict the welfare–optimality of $\mathcal{G}^W$.

Let us now verify that $\mathcal{G}' \in \Gamma$. The partition $\mathcal{I}^*$ has the key property that $\Delta Q(\theta)$ goes from negative to positive on each in each $I^* \in \mathcal{I}^*$. Formally, for a sufficiently small ϵ, $\Delta Q(\inf I^* + \epsilon) < 0$ for all $I^* \in \mathcal{I}^*$, except possibly for the lowest interval. Similarly, $\Delta Q(\sup I^* - \epsilon) \geq 0$ for all $I^* \in \mathcal{I}^*$, except possibly for the highest interval. To show $Q(\theta,\mathcal{G}')$ is increasing, pick $\theta_H > \theta_L$ and denote the respective partitions $I_H, I_L \in \mathcal{I}^*$. If $I_H = I_L$, then $Q(\theta_H,\mathcal{G}') \geq Q(\theta_L,\mathcal{G}')$ follows from the monotonicity of $Q(\theta,\mathcal{G})$ and $Q(\theta,\mathcal{G}^W)$. If $I_H \neq I_L$, then

$$\begin{aligned} Q(\theta_H,\mathcal{G}') &\geq Q(\inf I_H - \epsilon,\mathcal{G}') \geq Q(\inf I_H - \epsilon,\mathcal{G}^W) \\ &\geq Q(\sup I_L + \epsilon,\mathcal{G}^W) \geq Q(\sup I_L + \epsilon,\mathcal{G}') \geq Q(\theta_L,\mathcal{G}') \end{aligned}$$

The first, third and fifth inequalities come from monotonicity. The second and forth inequalities then follow from the above properties of $\mathcal{I}^*$. Hence $\mathcal{G}' \in \Gamma$, as required.

For the second step, index the objective function $h(\theta,t)$ so that $h(\theta,1) = MR(\theta) \geq 0$ and $h(\theta,0) = \theta$. Under (MON), the function $h(\theta,t) \geq 0$ is log–supermodular. Since $\Delta Q(\theta)$ is quasi–increasing on each $I^* \in \mathcal{I}^*$, Karlin and Rubin (1956, Lemma 1) states that $E[h(\theta,t)\Delta Q(\theta)|\mathcal{I}^*]$ is quasi–increasing in t.[14] Thus $E[\theta\Delta Q(\theta)|\mathcal{I}^*] \geq 0$ implies that $E[MR(\theta)\Delta Q(\theta)|\mathcal{I}^*] \geq 0$. Integrating over θ, we have $E[MR(\theta)\Delta Q(\theta)] \geq 0$. That is, $\Pi(\mathcal{G}^W) \geq \Pi(\mathcal{G})$. □

Corollary 1. *If $\mathcal{G}^W \subset \mathcal{G}^\Pi$ then $E[\phi \circ Q(\theta,\mathcal{G}^W)] \geq E[\phi \circ Q(\theta,\mathcal{G}^\Pi)]$ for all increasing, concave functions $\phi : I\!R \to I\!R$.*

Proof. See Appendix A.3. □

Proposition 1 says that groups will tend to be finer under profit–maximisation than welfare–maximisation. As shown by Corollary 1, this means that profit–maximisation induces a distribution of quality levels that has lower a mean and will tend to be more dispersed. In the school example, if one interprets $Q(\theta,\mathcal{G})$ as the exam scores of agent θ, then Corollary 1 yields testable implications of the theory.

The idea behind the segregation effect is that, under (MON), $MR(\theta)$ is steeper than θ, so a profit–maximising firm puts relatively more weight on the preferences of high types than the

[13] Notation: the function $E[\theta\Delta Q(\theta)|\mathcal{I}^*] : [\underline{\theta},\overline{\theta}] \to I\!R$ maps each type into its conditional expectation.

[14] Karlin and Rubin (1956, Lemma 1) actually shows that the objective function is weakly quasi–increasing. Lemma 12 in Appendix A.4 extends the result, showing the objective is quasi–increasing. The proof is essentially identical.

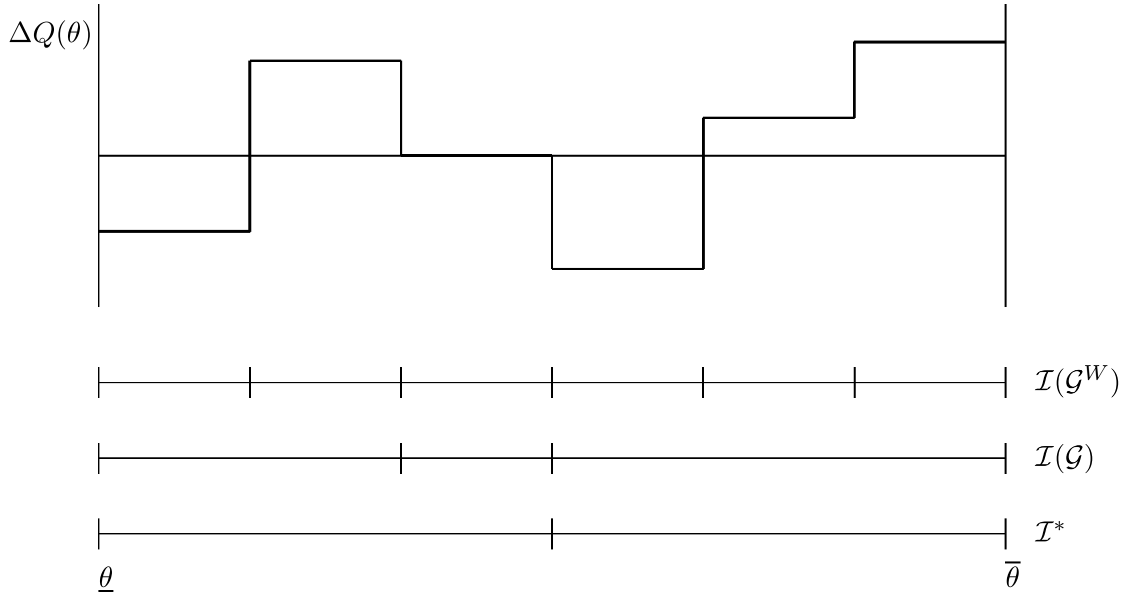

Figure 2: Sets in Proof of Proposition 1

social planner. This means a profit–maximising firm is more likely to split up a group, which helps the high types and hurts the low types. Intuitively, by introducing extra segregation the principal raises the cost of pretending to be a lower type and reduces consumer surplus. That is, by separating good and bad agents, the good agents become very keen to avoid the bad groups and can be forced to pay higher prices.

As stated in the Introduction, Proposition 1 makes no assumption about the nature of the peer effects. This is important because peer technology differs greatly across environments. Instead, Proposition 1 only uses the monotonicity condition that comes endogenously from the agents' self selection constraints.

Proposition 1 does have one limitation in that the welfare– and profit–maximising groups may not be ordered in terms of set inclusion. One should therefore view the result as saying that, if we start from the welfare–maximising group structure, then separating groups may increase profit, but merging groups will not. Moreover, Figure 3 shows that the spirit of the result may remain true even if the optimal solutions are not ordered.[15,16]

Example 1 (Pareto Distribution). Suppose $\theta \sim \text{Par}(\alpha, \beta)$, so that $f(\theta) = \alpha\beta^\alpha\theta^{-(\alpha+1)}$. In this case, (MON) holds with equality and profit is $(1-\alpha^{-1})E[\theta Q(\theta, \mathcal{G})]$. Consequently, the welfare– and profit–maximising group choices coincide. $\triangle$

[15] Figure 3 shows the profit and welfare–maximising group structures where $Q(G) = 0.55\sup(G) + 0.45\inf(G)$ and $\theta \sim U[2,3]$. In this example, $\mathcal{P}$ is a grid with increments of 1/1000. See Examples 5 and 11 for more details.

[16] Proposition 1 shows that a profit–maximising principal introduces more segregation than the welfare–maximising principal. Similarly, one can show that a consumer–surplus–maximising principal introduces less segregation than the welfare–maximising principal.

Figure 3: Optimal Group Formation: Linear Quality

5.3 Group Structure and Returns to Scale

This section analyses how different types of peer technology affect the optimal group structure.

Definition 1. *Consider any* $\mathcal{G}_H, \mathcal{G}_L \in \Gamma$ *such that* $\mathcal{G}_L \subset \mathcal{G}_H$*.*
(a) $Q(\theta, \mathcal{G})$ *has decreasing returns to scale (DRS) if* $E[Q(\theta, \mathcal{G}_H)] \geq E[Q(\theta, \mathcal{G}_L)]$*.*
(b) $Q(\theta, \mathcal{G})$ *has increasing returns to scale (IRS) if* $E[Q(\theta, \mathcal{G}_H)] \leq E[Q(\theta, \mathcal{G}_L)]$*.*

Under DRS, splitting a group raises the average quality. Under IRS, splitting a group lowers the average quality. Which case is appropriate depends upon the application and the interpretation of a group. To illustrate, consider the school example. If one interprets a group as a class, then dividing one class into two is likely to improve all students' performance. This suggests that the quality function will satisfy DRS. On the other hand, if one fixes the class size and interprets a group as an entire school, then the good students may help the poor students more than the poor students harm the good students (Henderson, Mieszkowski, and Sauvageau (1978)). In this case, the quality function will satisfy IRS.

Proposition 2. *Assume* $h(\theta)$ *is positive and increasing, and* $Q(G)$ *is weakly increasing in* G*.*
(a) Under decreasing returns to scale, the optimum is attained under full separation (i.e. $\mathcal{G}=\mathcal{P}$*).*
(b) Under increasing returns to scale, the optimum is attained when groups are connected.

Proof. (a) We prove a more general result: Suppose $h(\theta)$ is positive and increasing, and that DRS holds. Then, for any $\mathcal{G}_L, \mathcal{G}_H \in \Gamma$ such that $\mathcal{G}_L \subset \mathcal{G}_H$, the principal prefers $\mathcal{G}_H$ to $\mathcal{G}_L$. If $Q(G)$ is weakly increasing then $\mathcal{P} \in \Gamma$, so the optimum is attained when $\mathcal{G} = \mathcal{P}$.

Pick $\mathcal{G}_L, \mathcal{G}_H \in \Gamma$ such that $\mathcal{G}_L \subset \mathcal{G}_H$, and denote $\Delta Q(\theta) := Q(\theta, \mathcal{G}_H) - Q(\theta, \mathcal{G}_L)$. Let $\mathcal{I}^*$ be the coarsest partition on which $\Delta Q(\theta)$ is quasi–increasing. By Lemma 3, $\mathcal{I}^* \subset \mathcal{I}(\mathcal{G}_L) \subset \mathcal{I}(\mathcal{G}_H)$. We claim that DRS implies

$$E[\Delta Q(\theta)|\mathcal{I}^*] \geq 0 \tag{5.5}$$

To see this pick $I^* \in \mathcal{I}^*$ and let $\mathcal{G}'$ equal $\mathcal{G}_H$ on I^* and equal $\mathcal{G}_L$ elsewhere. First, $\mathcal{G}' \in \Gamma$, as in the proof of Proposition 1. Second, since $\mathcal{G}_L \subset \mathcal{G}'$, DRS implies

$$E[\Delta Q(\theta)|I^*] = E[Q(\theta, \mathcal{G}')] - E[Q(\theta, \mathcal{G}_L)] \geq 0$$

as required. This result implies that,

$$E[h(\theta)\Delta Q(\theta)|\mathcal{I}^*] \geq E[h(\theta)|\mathcal{I}^*]E[\Delta Q(\theta)|\mathcal{I}^*] \geq 0$$

where the first inequality comes from the fact that an increasing function and a quasi–increasing function have positive covariance (e.g. Persico (2000, Lemma 1)), and the second from (5.5). Integrating over θ, $\mathcal{G}_H$ yields a higher payoff that $\mathcal{G}_L$.

(b) Suppose there is IRS. Consider an arbitrary group structure, $\mathcal{G} \in \Gamma$. Form $\mathcal{I}(\mathcal{G})$ by merging overlapping groups in $\mathcal{G}$. Since $Q(G)$ is weakly increasing, we have $\mathcal{I}(\mathcal{G}) \in \Gamma$. Moreover, IRS implies that merging increases group quality so that $E[h(\theta)Q(\theta, \mathcal{I}(\mathcal{G}))] \geq E[h(\theta)Q(\theta, \mathcal{G})]$. □

Proposition 2 says that when $h(\theta)$ is increasing, there is full separation under DRS and may be pooling under IRS. This result applies to a welfare–maximising principal and, when $MR(\theta)$ is increasing, to a profit–maximising principal.[17]

Example 2 (Exponential Distribution). Suppose $f(\theta) = (1/\lambda)\exp(-(\theta - \underline{\theta})/\lambda)$, where $\underline{\theta} \geq \lambda$. Then profit is $E[\theta Q(\theta, \mathcal{G})] - \lambda E[Q(\theta, \mathcal{G})]$, and only differs from welfare in the second expression. Under DRS, full separation is optimal under both welfare– and profit–maximisation. Under IRS the second term decreases as $\mathcal{G}$ becomes finer. Consequently, profit is increased by splitting a group only if welfare is increased by splitting a group, illustrating the segregation effect. △

5.4 Group Quality Functions

This section considers a number of examples which have occurred in the literature. Examples 3–6 satisfy increasing or decreasing returns to scale; Example 7 shows that the optimal group structure may be more complex.

Example 3 (Average Quality). The average quality function, $Q(G) = E[\theta|\theta \in G]$, satisfies both increasing and decreasing returns to scale. As shown by Rayo (2005), one can then handle objective functions that are non–monotone. In particular, when the ironed $E[h(\theta)|\mathcal{P}]$ is increasing, the principal chooses full separation; when the ironed $E[h(\theta)|\mathcal{P}]$ is constant,

[17] Similarly, if $h(\theta)$ is constant, then there is full separation under DRS and full pooling under IRS. And if $h(\theta)$ is decreasing, then there may be multiple groups under DRS and there is full pooling under IRS. This last result can be shown by ironing the objective as in Myerson (1981).

the principal chooses full pooling. Thus there will always be full separation under welfare–maximisation, but there may be regions of pooling under profit–maximisation, if $MR(\theta)$ is badly behaved. This suggests welfare–maximisation leads to smaller groups than profit–maximisation. In comparison, Proposition 1 says that when we allow for different quality functions, the reverse is likely to be true. △

Example 4 (Generalised Average Quality). Suppose $Q(G) = \phi_1(E[\phi_2(\theta)|\theta \in G])$. If $\phi_1(\cdot)$ is concave and increasing, as suggested by the empirical analysis of Henderson, Mieszkowski, and Sauvageau (1978), then the quality function has increasing returns to scale, by Jensen's inequality. The profit–maximising group structure is then likely to exhibit some pooling and, by Proposition 1, be finer than the welfare–maximising group structure. △

Example 5 (Linear Quality). Suppose $Q(G) = \alpha \inf(G) + \beta \sup(G)$ and $\theta \sim U[\underline{\theta}, \overline{\theta}]$. One can verify that if $\mathcal{G} \in \Gamma$, then $Q(\theta, \mathcal{G}) = Q(\theta, \mathcal{I}(\mathcal{G}))$. Proposition 2(a) implies that if $\beta \leq \alpha$ (e.g. min–quality) there is decreasing returns to scale and welfare– and profit–maximisation will entail full separation. Conversely, if $\beta \geq \alpha$ (e.g. max–quality) there is increasing returns to scale and welfare– and profit–maximisation will generally induce some pooling. △

Example 6 (Multiplicative Quality). With multiplicative technology, $Q(G)$ exhibits decreasing returns to scale. Hence full separation is optimal if $\mathcal{P} \in \Gamma$.[18] Even with costly group formation, it will be optimal to have assortative matching when $\mathcal{P}$ is sufficiently fine. To see this, suppose G_1 and G_2 overlap. Then define disjoint G_1' and G_2' such that G_1' lies below G_2', $Q(G_1') = Q(G_1)$ and $Q(G_2') = Q(G_2)$. △

Example 7 (Intervals Not Optimal). Suppose $Q(G) = \sup(G) - E[\mathbf{1}_G]$, so the quality of the group depends upon its leader and the number of followers. This is one interpretation of the results Mas and Moretti (2006) and Falk and Ichino (2006). Here, groups will not take the form of intervals: it will be optimal to have lots of small groups, each with a very good leader. Since groups will overlap, Lemma 2(a) implies that they must all have the same quality. △

6 The Exclusion Effect

In Section 5 we examined the optimal way to segregate different types of agents when the principal serves all agents. In this section we extend the analysis to allow for exclusion. In the education example, these excluded agents may attend a public school or, in the case of universities, enter the workplace.

[18] $\mathcal{P} \in \Gamma$ if, for example, $\sigma(\mathcal{P})$ equals the Borel sets or $\mathcal{P}$ consists of intervals of equal measure.

An agent has an outside option of zero. Given a group structure $\mathcal{G}$, suppose $A \in \sigma(\mathcal{G})$ are excluded. Agents' rents can then be characterised by Lemma 1, where the quality function is given by $Q(\theta, \mathcal{G})\mathbf{1}_{\neg A}$.[19]

Lemma 4. *In any incentive compatible mechanism $\langle \mathcal{G}, A, y\rangle$ then*
(a) A is decreasing; and
(b) $A \in \sigma(\mathcal{I}(\mathcal{G}))$.

Proof. Follows from the monotonicity condition (Lemma 1(c)). □

6.1 Principal's Problem

There are two possible reasons to exclude an agent. First, the principal might wish to exclude θ if $h(\theta) < 0$. Second, the principal can exclude groups to 'monotonise' a non–monotonic quality function.

Formally, the *principal's problem* is to choose a group structure $\mathcal{G}$ and a set of excluded agents A to maximise

$$H = E\left[h(\theta)\, Q(\theta, \mathcal{G})\mathbf{1}_{\neg A}\right]$$

subject to $Q(\theta, \mathcal{G})\mathbf{1}_{\neg A}$ increasing in θ. Let $D^*_{\mathcal{G}}$ be the smallest decreasing set in $\sigma(\mathcal{I}(\mathcal{G}))$ such that $Q(\theta, \mathcal{G})$ is increasing on $[\underline{\theta}, \overline{\theta}]\backslash D^*_{\mathcal{G}}$, and let $Q^*(\theta, \mathcal{G}) := Q(\theta, \mathcal{G})\mathbf{1}_{\neg D^*_{\mathcal{G}}}$ be the induced quality function. Denote the positive and negative components of a function by $\phi(x)^+ := \max\{\phi(x), 0\}$ and $\phi(x)^- := -\min\{\phi(x), 0\}$.

Lemma 5. *Fix $\mathcal{G}$ and suppose $h(\theta)$ is quasi–increasing. Then the principal's maximal profits are given by*

$$H(\mathcal{G}) = E\Big[E[h(\theta)|I(\mathcal{G})]^+\, Q^*(\theta, \mathcal{G})\Big] \tag{6.1}$$

Proof. Fix $\mathcal{G}$. By Lemma 4, the excluded set A must be decreasing and measurable with respect to $\sigma(\mathcal{I}(\mathcal{G}))$. Given such a set, the monotonicity condition is satisfied if and only if $A \supset D^*_{\mathcal{G}}$. The principal's payoff is then given by

$$E\Big[h(\theta)Q(\theta, \mathcal{G})\mathbf{1}_{\neg A}\Big] = E\Big[E[h(\theta)Q(\theta, \mathcal{G})\mathbf{1}_{\neg A}\,|I(\mathcal{G})]\Big] = E\Big[E[h(\theta)|I(\mathcal{G})]\, Q(\theta, \mathcal{G})\mathbf{1}_{\neg A}\Big] \tag{6.2}$$

The first equality uses the law of iterated expectations, while the second uses the fact that $Q(\theta, \mathcal{G})$ and A are measurable with respect to $\sigma(\mathcal{I}(\mathcal{G}))$. The principal thus chooses $A \supset D^*_{\mathcal{G}}$ to maximise (6.2). Pointwise maximisation implies

$$A^* = D^*_{\mathcal{G}} \cup \{\theta : E[h(\theta)|I(\mathcal{G})] < 0\} \tag{6.3}$$

[19]Notation: $\neg A := \{\theta : \theta \notin A\}$.

Since $h(\theta)$ is quasi–increasing, (6.3) is a decreasing set, as required. This yields equation (6.1). □

Observe that Lemma 5 applies to the welfare–maximisation problem and, under (MON), to the profit–maximisation problem. Moreover, if we assume that the quality function is weakly increasing, then the principal need not exclude in order to 'monotonise' the quality function.

Lemma 6. *Suppose that $h(\theta)$ is quasi–increasing and $Q(G)$ is weakly increasing in G. Then the principal's payoffs are maximised by $\mathcal{G} \in \Gamma$, and are given by*

$$H(\mathcal{G}) = E\Big[E[h(\theta)|I(\mathcal{G})]^{+}\, Q(\theta,\mathcal{G})\Big] \tag{6.4}$$

Proof. Suppose $\mathcal{G} \notin \Gamma$ maximises the principal's payoff (6.1). Then form a new structure $\mathcal{G}'$ by pooling all excluded agents into one group. Since $Q(G)$ is weakly increasing, $\mathcal{G}' \in \Gamma$. The new optimal set of excluded agents is then given by

$$A^{*} = \{\theta : E[h(\theta)|I(\mathcal{G}')] < 0\}$$

The new group structure $\mathcal{G}'$ therefore attains a (weakly) greater payoff than $\mathcal{G}$, as required. □

6.2 Welfare– and Profit–Maximisation

The *principal's problem* is thus to choose $\mathcal{G}$ to maximise (6.1). Proposition 3 shows that the segregation effect extends to the case where the principal can exclude agents. Notably, this result places no restrictions on the sign of $MR(\theta)$.

Proposition 3 (Segregation Effect II)**.** *Suppose (MON) holds. For any welfare–maximising solution, $\mathcal{G}^W$, $\Pi(\mathcal{G}^W) \geq \Pi(\mathcal{G})$ on $\{\mathcal{G} : \mathcal{G} \subset \mathcal{G}^W\}$. Hence if any optimal solutions, $\mathcal{G}^W$ and $\mathcal{G}^\Pi$, are ordered in terms of set inclusion, then there exists a profit–maximising solution, $\mathcal{G}^{\Pi *}$, such that $\mathcal{G}^W \subset \mathcal{G}^{\Pi *}$.*

Proof. See Appendix A.5. □

There are two effects underlying Proposition 3. First, a profit–maximising principal cares relatively more about high value agents than a welfare–maximising principal (see Proposition 1). Second, a profit–maximising principal is more willing to exclude agents than a welfare–maximising agent (see Proposition 4). Hence the smaller group size provides additional flexibility to exclude some agents.

Proposition 4 (Exclusion Effect)**.** *Suppose that (MON) holds and either (a) $Q(G)$ is weakly increasing in G, or (b) $\mathcal{G}^W \subset \mathcal{G}^\Pi$. Then exclusion is higher under profit–maximisation than welfare–maximisation.*

Proof. Denote the types excluded under profit–maximisation by A^{Π} and those excluded under welfare–maximisation by A^W.

(a) Lemma 6 implies $A^W = \emptyset$, so that $A^W \subset A^{\Pi}$.

(b) Suppose $\mathcal{G}^W \subset \mathcal{G}^{\Pi}$. By Lemma 3, $\mathcal{I}(\mathcal{G}^W) \subset \mathcal{I}(\mathcal{G}^{\Pi})$. Let $\mathcal{I}^*$ be the coarsest partition such that $\Delta Q(\theta) := Q(\theta, \mathcal{G}^{\Pi}) - Q(\theta, \mathcal{G}^W)$ is quasi–increasing. As in the proof of Proposition 1, we have $E[\theta \Delta Q(\theta)|\mathcal{I}^*] \leq 0$. Yet if $A^{\Pi} \subsetneq A^W$, then $E[\theta \Delta Q(\theta)|I^*] > 0$ on the first interval in $\mathcal{I}^*$, yielding a contradiction. □

The exclusion effect is analogous to the standard monopoly distortion. Under profit–maximisation the principal would like to exclude agents with negative marginal revenue, whereas under welfare–maximisation the principal would like to exclude no agents (see Proposition 4(a)). The principal may also exclude agents in order to 'monotonise' the quality function. However, the profit–maximising principal is more likely to exclude agents than a welfare–maximising principal since she cares less about low value agents (see Proposition 4(b)).

Proposition 5 provides a characterisation of the excluded agents. A quality function $Q(G)$ is *increasing* in G if $Q(G_H) \geq Q(G_L)$ whenever G_H is larger than G_L in strict set order.[20]

Proposition 5. *Suppose $h(\theta)$ is increasing. Assume that either:*
(a) $Q(G)$ is weakly increasing in G and exhibits DRS; or
(b) $Q(G)$ is increasing in G and exhibits IRS.
Then the principal's objective is maximised by excluding the set $A^* = \{\theta : E[h(\theta)|\mathcal{P}] < 0\}$.

Proof. (a) Suppose $Q(G)$ exhibits DRS. Since $Q(G)$ is weakly increasing in G, Lemma 6 implies that profit is maximised by $\mathcal{G} \in \Gamma$. Also observe that, given $h(\theta)$ is increasing, $\overline{D} := \{\theta : E[h(\theta)|\mathcal{P}] < 0\}$ is a decreasing set.

First, suppose that $A \supsetneq \overline{D}$. Then form a new group structure $\mathcal{G}'$ by including $A \backslash \overline{D}$ as a single group. Since $Q(G)$ is weakly increasing, $\mathcal{G}' \in \Gamma$. Moreover, $\mathcal{G}'$ yields a (weakly) higher payoff than $\mathcal{G}$.

Next, suppose that $A \subsetneq \overline{D}$. Since $h(\theta)$ is increasing, and $Q(G)$ is weakly increasing and exhibits DRS, Proposition 2(a) implies that the principal's payoffs are maximised by full separation.[21] Lemma 6 then implies that the principal should exclude agents with $E[h(\theta)|\mathcal{P}] < 0$.

(b) Suppose $Q(G)$ exhibits IRS. The proof that the principal's payoff is maximised by $A \subset \overline{D}$ is the same as part (a). Next, suppose that $A \subsetneq \overline{D}$. Since $Q(G)$ is (weakly) increasing and exhibits IRS, Proposition 2(b) implies that the principal's payoffs are maximised when groups are intervals.[22] Denote the lowest included interval by I_0. By Lemma 6, we must have

[20]Definition: G_H is larger than G_L in strict set order if $\min\{\theta, \theta'\} \in G_L$ and $\max\{\theta, \theta'\} \in G_H$ for all $\theta \in G_L$ and $\theta' \in G_H$.

[21]Proposition 2(a) does not allow for exclusion, but the result immediately extends. Intuitively, with exclusion, smaller groups provide more flexibility and, via Jensen's inequality, further increase the principal's payoff.

[22]Proposition 2(b) does not allow for exclusion, but the proof is identical.

$E[h(\theta)|I_0] \geq 0$, so I_0 is the only interval that intersects with $\overline{D}$. Next, form a new group structure, $\mathcal{G}''$, by excluding $\overline{D}$. Since $Q(G)$ is (weakly) increasing, $\mathcal{G}'' \in \Gamma$. Since $Q(G)$ is increasing, we thus have $\mathcal{G}'' \in \Gamma$ and $Q(I_0 \backslash \overline{D}) \geq Q(I_0)$. Hence,

$$
\begin{aligned}
E\Big[h(\theta)Q(\theta,\mathcal{G})\,\Big|\,I_0\Big] &= E\Big[E[h(\theta)|\mathcal{P}]^+ Q(I_0) - E[h(\theta)|\mathcal{P}]^- Q(I_0)\,\Big|\,I_0\Big] \\
&\leq E\Big[E[h(\theta)|\mathcal{P}]^+ Q(I_0 \backslash \overline{D})\,\Big|\,I_0\Big]
\end{aligned}
$$

so that $\mathcal{G}''$ attains a (weakly) higher payoff than $\mathcal{G}$. □

At first sight, it seems reasonable to conjecture that the principal should exclude an agent if and only if $h(\theta) < 0$. There are two reasons why this may not be correct. First, the principal may exclude more agents in order to 'monotonise' the quality function. Second, the principal may exclude fewer agents if they exert a positive externality on the included agents, with $h(\theta) > 0$. Broadly speaking, Proposition 5 shows that both of these possibilities are ruled out if the quality function is increasing.

7 Costly Group Formation

The segregation effect (Proposition 1) states that groups will be finer under profit–maximisation than welfare–maximisation. With costly group formation this is countered by the appropriability effect: a profit–maximising principal cannot capture consumer surplus and may not invest enough in creating groups. Examples 8–9 illustrate how the appropriability effect can dominate the segregation effect. Proposition 6 then derives sufficient conditions for the segregation effect to dominate the appropriability effect.

In order to focus on the segregation effect, we suppose the principal cannot exclude any agents.[23] The *principal's problem* is thus to choose $\mathcal{G} \in \Gamma$ to maximise $H(\mathcal{G}) - c(\mathcal{G})$, where $H(\mathcal{G}) := E[h(\theta)Q(\theta,\mathcal{G})]$ and $c(\mathcal{G})$ is an arbitrary cost function.

Example 8 (Appropriability Effect I). Suppose $\theta \sim \text{Par}(\alpha,\beta)$, as in Example 1. Then the profit–maximising problem is to choose groups $\{G_i\}_{i=1}^N$ to maximise:

$$
(1-\alpha^{-1})\sum_{i=1}^{N} Q(G_i)\int_{G_i}\theta dF - c(\mathcal{G}) \tag{7.1}
$$

This coincides with the welfare–maximising problem if $\alpha = \infty$. Suppose that $c(\mathcal{G})$ only depends on $\mathcal{G}$ through the number of groups N, and is increasing in N (e.g. N is the number of teachers).

[23] One can allow for exclusion using the 'evil type' approach in footnote 10. Proposition 6 then holds no matter what the sign of $MR(\theta)$. Saying this, the result is less interesting when $MR(\theta) < 0$ since $\mathcal{G}^W$ and $\mathcal{G}^\Pi$ are unlikely to be ordered, as assumed in the final line.

It then follows from (7.1) that there will be more groups under welfare–maximisation than profit–maximisation.[24] △

Example 9 (Appropriability Effect II). Suppose that splitting a group increases everyone's quality (e.g. multiplicative quality) and that $MR(\theta) \geq 0$. Hence $\Delta Q(\theta) = Q(\theta, \mathcal{G}_H) - Q(\theta, \mathcal{G}_L) \geq 0$ $(\forall \theta)$, for $\mathcal{G}_L, \mathcal{G}_H \in \Gamma$ such that $\mathcal{G}_L \subset \mathcal{G}_H$. Since $\theta \geq MR(\theta) \geq 0$, we have $E[\theta \Delta Q(\theta)] \geq E[MR(\theta)\Delta Q(\theta)]$. That is, whenever a profit–maximiser splits a group, a welfare–maximiser will also split the group. △

Assumption (HR). $[1 - F(\theta)]/f(\theta)$ is decreasing in θ.

Proposition 6 (Weak Segregation Effect)**.** *Suppose (HR) holds and $Q(G)$ exhibits increasing returns to scale. For any welfare–maximising solution, $\mathcal{G}^W$, $\Pi(\mathcal{G}^W) \geq \Pi(\mathcal{G})$ on $\{\mathcal{G} \in \Gamma : \mathcal{G} \subset \mathcal{G}^W\}$. Hence if any optimal solutions, $\mathcal{G}^W$ and $\mathcal{G}^\Pi$, are ordered in terms of set inclusion, then there exists a profit–maximising solution, $\mathcal{G}^{\Pi *}$, such that $\mathcal{G}^W \subset \mathcal{G}^{\Pi *}$.*

Proof. Suppose $\mathcal{G}^W$ maximises welfare and fix $\mathcal{G} \subset \mathcal{G}^W$, such that $\mathcal{G} \in \Gamma$. Hence $E[\theta \Delta Q(\theta)] \geq c(\mathcal{G}^W) - c(\mathcal{G})$, where $\Delta Q(\theta) = Q(\theta, \mathcal{G}^W) - Q(\theta, \mathcal{G})$. By Lemma 3, $\mathcal{I}(\mathcal{G}) \subset \mathcal{I}(\mathcal{G}^W)$.

Let $\mathcal{I}^*$ be the coarsest partition on which $\Delta Q(\theta)$ is quasi–increasing. As in Proposition 2, IRS implies that $E[\Delta Q(\theta) | \mathcal{I}^*] \leq 0$. Since $\Delta Q(\theta)$ is quasi–increasing on each $I^* \in \mathcal{I}^*$, $E[\mathbf{1}_D \Delta Q(\theta)] \leq 0$ for any decreasing set D.

For decreasing sets $\{D_i\}$ and positive constants $\{a_i\}$, $i \in \{1, \ldots, m\}$, $E[\sum_i a_i \mathbf{1}_{D_i} \Delta Q(\theta)] \leq 0$. Since (HR) implies that $[1 - F(\theta)]/f(\theta)$ is decreasing, we can define $\{D_i\}$ such that $\sum_i a_i \mathbf{1}_{D_i} \to [1 - F(\theta)]/f(\theta)$ as $m \to \infty$. Hence

$$E\left[\frac{1 - F(\theta)}{f(\theta)} \Delta Q(\theta)\right] \leq 0 \tag{7.2}$$

Equation (7.2) implies that

$$\Pi(\mathcal{G}^W) - \Pi(\mathcal{G}) = E[MR(\theta)\Delta Q(\theta)] \geq E[\theta \Delta Q(\theta)] \geq c(\mathcal{G}^W) - c(\mathcal{G})$$

as required. □

The appropriability effect states that a profit–maximising principal cannot capture consumer surplus and may not invest enough in group formation. Under (HR) and IRS, consumer surplus is maximised by complete pooling, so a profit–maximiser will be willing to invest more in group formation than a welfare–maximiser.

[24] There are other variants of this result. For example, if $c(\mathcal{G}_H) \geq c(\mathcal{G}_L)$ for $\mathcal{G}_L \subset \mathcal{G}_H$, then $W(\mathcal{G}^\Pi) \geq W(\mathcal{G})$ on $\{\mathcal{G} \in \Gamma : \mathcal{G} \subset \mathcal{G}^\Pi\}$.

Proposition 6 is more restrictive than the original segregation effect (Proposition 1). First, it assumes that the distribution of types satisfies (HR) rather than (MON). Defining $h(\theta, 0) = \theta$ and $h(\theta, 1) = MR(\theta)$, (MON) implies that $h(\theta, t)$ is log–supermodular, while the stronger (HR) assumption is required for $h(\theta, t)$ to be supermodular. Second, the result assumes that $Q(G)$ satisfies IRS, overcoming the problem in Example 9.

Example 10 provides a tractable numerical illustration of Proposition 6. Observe that Example 10 exhibits constant returns to scale, so the conditions of Proposition 6 are stronger than necessary.

Example 10 (Average Quality). Suppose $Q(G) = E[\theta \mid G]$ and $c(\mathcal{G})$ only depends on $\mathcal{G}$ through the number of groups, N. By Proposition 2(b), the optimal group structure consists of intervals. The principal then chooses cutoffs $\{\theta_i\}_{i=0}^{N}$ to maximise welfare or profit (5.4). Assume $\sigma(\mathcal{P})$ equals the Borel sets, enabling the use of calculus. When $\theta \sim U[\underline{\theta}, \overline{\theta}]$, the FOCs for $\{\theta_i\}_{i=1}^{N-1}$ reduce to $(\theta_{i+1} - \theta_i) = (\theta_i - \theta_{i-1})$ under *both* welfare– and profit–maximisation. If $\underline{\theta}$ is sufficiently high, such that exclusion is not desirable under either objective, then marginal welfare from an extra group is $dW/dN = (\overline{\theta} - \underline{\theta})^2/6N^3$, while the marginal profit from an extra group is $d\Pi/dN = (\overline{\theta} - \underline{\theta})^2/3N^3$. Since $d\Pi/dN \geq dW/dN$, a profit–maximising principal will choose to have more groups.[25]

This example shows that, once again, profit–maximisation exhibits excessive segregation. However, conditional on choosing the same number of groups, the welfare– and profit–maximising principals will choose to divide agents in the same manner. This makes regulation relatively easy: the government need only restrict the total number of tariffs; the principal will then choose the welfare–maximising group structure.[26] △

8 Group Size and Relative Position

In this section we investigate how the size of groups changes with the types of agents. These results are of particular interest in the education market, where they enable us to assess how class composition should change (a) with ability and (b) over time.

Our results can be summarised as follows. In Section 8.1, we show that, under costless group formation, higher types will tend to be in larger groups. In Section 8.2, we show this result extends to costly group formation if the quality function exhibits increasing returns to scale, but reverses under decreasing returns. Finally we relate our results to those of Lazear (2001) and discuss the implications for education.

[25] See Web Appendix. http://www.economics.utoronto.ca/board/papers/groups-webappendix.pdf.

[26] Although the models are very different, this result resembles Epple and Romano (1998, Propsition 4(i)) which showed that, conditional on the number of schools, a private system segregates students optimally. Their paper also showed that the business stealing effect tends to lead to excessive entry of private schools.

8.1 Costless Group Formation

In order to examine how group size changes with agents' types, we consider the following experiment. First, suppose that types are initially distributed according to $\theta \sim f(\theta)$ on $[\underline{\theta}, \overline{\theta}]$. We then examine the effect of an upwards shift in the distribution so that $\theta \sim f(\theta - t)$ on $[\underline{\theta} + t, \overline{\theta} + t]$. We then compare the size of the group containing θ in the initial distribution to that containing $\theta + t$ in the shifted distribution. For an arbitrary group, G, under the initial distribution, let $G(t) := G + t$. Similarly define $\mathcal{G}(t), \Gamma(t)$ and $\mathcal{P}(t)$ relative to the new distribution.

Assumption (LIN). A vertical shift affects group quality linearly: $Q(G(t)) = Q(G(0)) + \lambda t$.

The (LIN) assumption is satisfied by average–quality ($\lambda = 1$) and linear–quality ($\lambda = \alpha + \beta$) among others. The *principal's problem* is then to choose $\mathcal{G}(t) \in \Gamma(t)$ to maximise[27]

$$H = \int_{\underline{\theta}+t}^{\overline{\theta}+t} h(\theta, t) Q(\theta, \mathcal{G}(t)) f(\theta - t) \, d\theta.$$

Under (LIN) we can change variables to $\tilde{\theta} = \theta - t$. Under welfare–maximisation, $h(\theta, t) = \theta$, so the objective becomes $h(\tilde{\theta} + t, t) = \tilde{\theta} + t$. Under profit–maximisation, $h(\theta, t) = \theta - [1 - F(\theta - t)]/f(\theta - t)$ so the objective becomes $h(\tilde{\theta} + t, t) = MR(\tilde{\theta}) + t$. Putting this together, $h(\tilde{\theta} + t, t) = h(\tilde{\theta}) + t$. Denoting $\mathcal{G} = \mathcal{G}(0)$, the *principal's problem* is thus to choose $\mathcal{G} \in \Gamma$ to maximise

$$H(\mathcal{G}, t) = \int_{\underline{\theta}}^{\overline{\theta}} [h(\tilde{\theta}) + t][Q(\tilde{\theta}, \mathcal{G}) + \lambda t] f(\tilde{\theta}) \, d\tilde{\theta}. \tag{8.1}$$

Proposition 7. *Suppose $h(\theta) + t$ is positive and increasing in θ, and that quality satisfies (LIN). Fix $t_H > t_L$. For any t_H–optimal solution, $\mathcal{G}^H$, $H(\mathcal{G}^H, t_L) \geq H(\mathcal{G}, t_L)$ on $\{\mathcal{G} \in \Gamma : \mathcal{G} \subset \mathcal{G}^H\}$. Hence if any optimal solutions, $\mathcal{G}^L$ and $\mathcal{G}^H$, are ordered in terms of set inclusion, then there exists a t_L–optimal solution, $\mathcal{G}^{L*}$, such that $\mathcal{G}^H \subset \mathcal{G}^{L*}$.*

Proof. The function $h(\theta) + t$ is positive and increasing in θ, and is therefore log–submodular in (θ, t). The rest of the proof is identical to Proposition 1. □

Proposition 7 says that higher types will tend to be in larger groups under welfare or profit–maximisation. To understand the result, take a group $[\theta_L + t, \theta_H + t]$ and consider a split that reduces the quality of the low types a lot, while raising the quality of the high types a little. When the agents' types are low (i.e. t is low), the ratio between the highest and lowest types in the group, $(\theta_H + t)/(\theta_L + t)$, is large and this split may increase welfare/profit. Yet when the agents' types are high (i.e. t is high), the ratio between the highest and lowest types in the group is small and the split is less likely to be beneficial.

[27] This assumes the principal cannot exclude. As in footnote 10, this is without loss if $h(\theta, t)$ is always positive.

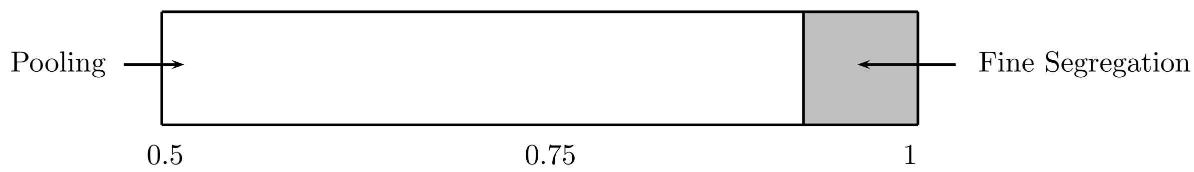

Figure 4: Optimal Group Formation: Cobb Douglas Quality

Our result concerns the group structure as the entire distribution of types shifts. It also suggests that higher types will be in larger groups than lower types within a given distribution if the relative ratio of high types to low types remains constant throughout the distribution (e.g. the density is uniform, ignoring boundary problems). This can be seen in Figure 3, where higher types are in larger groups under both welfare– and profit–maximisation.

Proposition 7 assumes that the quality function satisfies (LIN). Without this assumption the result may be overturned, and high types may be in smaller groups than low types. For example, Figure 4 illustrates the welfare–maximising partition under Cobb–Douglas quality.[28] In this case, agents below a certain cutoff are pooled into one giant group, while all other types are very finely segregated. Intuitively, for low types the quality function is very concave and the returns to scale are large; for high types the quality function is less concave and the returns to scale are small.

8.2 Costly Group Formation

With costly group formation, the *principal's problem* is to choose $G \in \Gamma$ to maximise $H(\mathcal{G}, t) - c(\mathcal{G})$, where $H(\mathcal{G}, t)$ is defined by (8.1) and $c(\mathcal{G})$ is an arbitrary cost function.

Proposition 8. *Suppose $h(\theta) + t$ is increasing in θ, and that quality satisfies (LIN). Suppose either:*
(a) there are increasing returns to scale and fix $t'' < t'$; or
(b) there are decreasing returns to scale and fix $t'' > t'$.
Then for any t'–optimal solution, $\mathcal{G}'$, $H(\mathcal{G}', t'') \geq H(\mathcal{G}, t'')$ on $\{\mathcal{G} \in \Gamma : \mathcal{G} \subset \mathcal{G}'\}$. Hence if any optimal solutions, $\mathcal{G}'$ and $\mathcal{G}''$, are ordered in terms of set inclusion, there exists a t''–optimal solution, $\mathcal{G}''^*$, such that $\mathcal{G}' \subset \mathcal{G}''^*$.

Proof. Suppose IRS holds and fix $t' > t''$. Consider a t'–optimal solution, $\mathcal{G}'$, and consider $\mathcal{G} \subset \mathcal{G}'$. IRS implies that $E[\Delta Q(\theta)] \leq 0$, where $\Delta Q(\theta) := Q(\theta, \mathcal{G}') - Q(\theta, \mathcal{G})$. Observe that $H(\mathcal{G}', t) - H(\mathcal{G}, t) = E[(h(\theta) + t)\Delta Q(\theta)]$ and

$$E[(h(\theta) + t'')\Delta Q(\theta)] - E[(h(\theta) + t')\Delta Q(\theta)] = (t'' - t')E[\Delta Q(\theta)] \geq 0$$

[28]In Figure 4, $\theta \sim U[0.5, 1]$, $Q(G) = (E[\theta|G] - 0.5)^{0.3}$ and $\sigma(\mathcal{P})$ equals the Borel sets. In the finely segregated part, the groups are around 0.0001 wide.

Hence $H(\mathcal{G}',t') - H(\mathcal{G},t') \geq c(\mathcal{G}') - c(\mathcal{G})$ implies $H(\mathcal{G}',t'') - H(\mathcal{G},t'') \geq c(\mathcal{G}') - c(\mathcal{G})$, as required. The proof for DRS is identical. □

Proposition 8 says that (a) under IRS, higher types are in larger groups; and (b) under DRS, higher types are in smaller groups. In comparison, if there is costless group formation then (a) under IRS, higher types are in larger groups (Proposition 7); and (b) under DRS, there is full separation (Proposition 2). To understand this result, consider the IRS case. Splitting a group has an efficiency effect, reducing the mean group quality, and a distributional effect, benefiting high types while hurting low types. When all types are higher, then the ratio between the highest and lowest types in a group declines, and the distributional effect becomes less important. Hence the efficiency effect becomes paramount, leading to an increase in group size.

Proposition 8 considers a shift of the entire distribution of types. Example 11 shows that, under the uniform–linear model, a similar result applies within a given distribution of types.

Example 11 (Linear–Quality). Suppose $Q(G) = \alpha \inf(G) + \beta \sup(G)$ and $c(\mathcal{G})$ only depends on $\mathcal{G}$ through the number of groups, N. By Example 5, the optimal group structure consists of intervals. The welfare–maximising principal then chooses cutoffs $\{\theta_i\}_{i=0}^N$ to maximise (5.1). Assume $\theta \sim U[\underline{\theta}, \overline{\theta}]$ and let $\sigma(\mathcal{P})$ equal the Borel sets, enabling the use of calculus. Under constant returns to scale ($\alpha = \beta$), the FOCs for $\{\theta_i\}_{i=1}^{N-1}$ reduce to $(\theta_{i+1} - \theta_i) = (\theta_i - \theta_{i-1})$, as in Example 10, so groups are the same size for all types. Under DRS (i.e. $\alpha \geq \beta$), then $(\theta_{i+1} - \theta_i) \leq (\theta_i - \theta_{i-1})$, so groups are smaller for higher types. Under IRS (i.e. $\alpha \leq \beta$), then $(\theta_{i+1} - \theta_i) \geq (\theta_i - \theta_{i-1})$, so groups are larger for higher types.[29] △

These results have implications for education markets. When considering the optimal classroom size, the assumption of decreasing returns seems reasonable. Proposition 8 then suggests that more able students should be in smaller classes. Intuitively, when students are more able, they have more to gain from a reduction in class size. This results seems consistent with the 35% reduction is U.S. pupil–teacher ratio over the last half–century (Hanushek (1999)).

This result can also be contrasted to Lazear (2001, Proposition 1) which shows that, with multiplicative quality, groups are larger for higher types. The reason for Lazear's result is that, under multiplicative quality, there are significant decreasing returns to scale when θ is low, but approximately constant returns as θ approaches one. This is the reverse of the logic behind the Cobb–Douglas example in Figure 4.

When considering the optimal school composition, holding class size constant, Henderson, Mieszkowski, and Sauvageau (1978) suggest that increasing returns may be the appropriate assumption. Propositions 7–8 then imply that selective grammar schools were of more use in the 1950s, when education levels were relatively low, than today.

[29]See Web Appendix. http://www.economics.utoronto.ca/board/papers/groups-webappendix.pdf.

9 Conclusion

This paper has analysed how a principal will divide agents into groups in the presence of peer effects. With costless group formation, we showed that a profit–maximising principal will segregate agents more finely than is socially optimal (the segregation effect) and exclude too many agents (the exclusion effect). We also analysed how the optimal group structure depends upon the returns to scale of the peer technology. With costly group formation, we demonstrated that a profit–maximising firm may not invest enough in group formation (the appropriability effect). However, under increasing returns to scale, the segregation effect dominates the appropriability effect.

Our analysis has direct implications for public policy. The large growth in private communities suggests that these developments are filling a gap in the market, leading to welfare gains for parts of society. Our model is consistent with this fact: even when agents benefit from living in varied communities (i.e. under increasing returns to scale) then the welfare–maximising outcome will exhibit assortative matching, consisting of different tiers of communities. Nevertheless, the segregation effect illustrates that private community development will often lead to excessively homogenous neighbourhoods. This suggests that, in cases where a few local developers have market power, the government should be especially careful to ensure new developments contain a wide range of housing stock.

This paper also informs the debate on the role of private schools. Much of the discussion over vouchers and public–private partnerships centres on the mantra of parental choice. However, choice is not the aim in itself. This paper has shown that when the options are designed by an organisation with market power, then private provision may provide too much choice, introducing excessive segregation. On the positive side, given knowledge of these distortions, there is no reason why an alert regulatory agency cannot mitigate their impact.

A Omitted Material

A.1 Multiplicative Technology

Kremer (1993) and Lazear (2001) consider a group of agents, $G = \{p_1, \ldots, p_{\mu(G)}\}$, where agent i makes a mistake with probability $p_i = 1 - \theta_i$. For example, one can think of a project that requires $\mu(G)$ jobs to be completed. The probability the project is completed successfully is then

$$Q(G) = \prod_{i=1}^{\mu(G)} (1 - p_i)$$

We now consider a continuous type analogue to this quality function. Suppose agents types are distributed according to absolutely continuous measure μ, where $\mu([\underline{\theta}, \overline{\theta}]) = m$. Let $f(\theta) = d\mu(\theta)/m$ be the normalised density. The quality of a group $G \subset [\underline{\theta}, \overline{\theta}]$ is determined as follows. First, as in the discrete model, suppose that a project requires $\mu(G)$ jobs to be completed. Second, break each job into k equal tasks. Third, draw k agents independently from G, where each agent makes a mistake with probability p_i/k. Then let each of these agents do one of the k tasks for each of the $\mu(G)$ jobs. The probability the project is completed successfully is then

$$Q_k(G) = \left[\prod_{j=1}^{k} \left(1 - \frac{p_i}{k}\right) \right]^{\mu(G)}$$

Lemma 7. *As the number of tasks grows, $k \to \infty$,*

$$Q_k(G) \xrightarrow{p} \exp\left(-\int_G (1-\theta) d\mu\right) = \exp\left(-m \int_G (1-\theta) dF\right).$$

Proof. Define

$$\Delta_k(p_i) := \frac{\ln(1) - \ln(1 - p_i/k)}{p_i/k}$$

For each k we draw a new set of agents with error probabilities $\{p_i\}_{i=1}^k$, so $\Delta_k(p_i)$ is a triangular array. Observe that for a given p_i,

$$\lim_{k \to \infty} \Delta_k(p_i) = \frac{d}{dx} \ln(x)|_{x=1} = 1$$

Since $\Delta_k(p_i) \in [1, \Delta_k(1)]$, the law of large numbers (e.g. Durrett (1995, p. 41)) implies

$$\begin{aligned} \ln(Q_k(G)) &= \mu(G) \sum_{i=1}^{k} \ln(1 - p_i/k) \\ &= \mu(G) \, \frac{1}{k} \sum_{i=1}^{k} -p_i \left[\frac{\ln(1) - \ln(1 - p_i/k)}{p_i/k} \right] \\ &= \mu(G) \, \frac{1}{k} \sum_{i=1}^{k} -p_i \, \Delta_k(p_i) \\ &\xrightarrow{p} -\mu(G) \, E[p \mid G] \end{aligned}$$

where the second line uses $\ln(1) = 0$. □

A.2 Proof of Lemma 3

Pick $I_H \in \mathcal{I}(\mathcal{G}_H)$. By construction, there exists a $G_H \in \mathcal{G}_H$ such that $I_H = I_{\mathcal{G}_H}(G_H)$. Since $\mathcal{G}_L \subset \mathcal{G}_H$, there exists $\{G_{H_j}\}_{j \in J}$ such that $G_{H_j} \in \mathcal{G}_H$ and $G_H \in \cup_{j \in J} G_{H_j} = G_L$ for some $G_L \in \mathcal{G}_L$. Clearly,

$$I_H = I_{\mathcal{G}_H}(G_H) \subset \cup_{j \in J} I_{\mathcal{G}_H}(G_{H_j})$$

Since $I_{\mathcal{G}}(G)$ is the smallest interval containing G, $I_{\mathcal{G}_H}(G_{H_j}) \subset I_{\mathcal{G}_H}(\cup_{j \in J} G_{H_j})$ for each $j \in J$. Thus,

$$\cup_{j \in J} I_{\mathcal{G}_H}(G_{H_j}) \subset \cup_{j \in J} I_{\mathcal{G}_H}(\cup_{j \in J} G_{H_j}) = I_{\mathcal{G}_H}(\cup_{j \in J} G_{H_j}) = I_{\mathcal{G}_H}(G_L)$$

Since $\mathcal{G}_L \subset \mathcal{G}_H$, the definition of $I_{\mathcal{G}}(G)$ implies that

$$I_{\mathcal{G}_H}(G_L) \subset I_{\mathcal{G}_L}(G_L) \in \mathcal{I}(\mathcal{G}_L)$$

We have thus shown that $I_H \in \mathcal{I}(\mathcal{G}_L)$, as required.

A.3 Proof of Corollary 1

Suppose $\mathcal{G}^W \subset \mathcal{G}^\Pi$. Lemma 3 implies that $\mathcal{I}(\mathcal{G}^W) \subset \mathcal{I}(\mathcal{G}^\Pi)$. Let $\mathcal{I}^*$ be the coarsest partition such that $\Delta Q(\theta) := Q(\theta, \mathcal{G}^\Pi) - Q(\theta, \mathcal{G}^W)$ is quasi–increasing.

Lemma 8. $E[\Delta Q(\theta) | \mathcal{I}^*] \leq 0$.

Proof. As in Proposition 1, we have $E[\theta \Delta Q(\theta) | \mathcal{I}^*] \leq 0$. For any $I^* \in \mathcal{I}^*$, it follows that

$$0 \geq E[\theta \Delta Q(\theta) | I^*] \geq E[\theta | I^*] E[\Delta Q(\theta) | I^*]$$

where the second inequality comes from the fact that a quasi–increasing function is positively correlated with an increasing function (e.g. Persico (2000, Lemma 1)). □

Fix $I^* \in \mathcal{I}^*$. Denote the distribution function of $Q(\theta, \mathcal{G}^\Pi)$, conditional on $\theta \in I^*$, by $F_\Pi(q) := E[\mathbf{1}_{Q(\theta,\mathcal{G}^\Pi)\leq q}|I^*]$ Similarly define the distribution function of $Q(\theta, \mathcal{G}^W)$, conditional on $\theta \in I^*$, by $F_W(q) := E[\mathbf{1}_{Q(\theta,\mathcal{G}^W)\leq q}|I^*]$.

Lemma 9. *For any $I^* \in \mathcal{I}^*$, $F_W(q) - F_\Pi(q)$ is weakly quasi–increasing.*

Proof. $Q(\theta, \mathcal{G}^W)$ and $Q(\theta, \mathcal{G}^\Pi)$ are increasing, so denote the inverses by $Q_W^{-1}(q) := \inf\{\theta : Q(\theta, \mathcal{G}^W) > q\}$ and $Q_\Pi^{-1}(q) := \inf\{\theta : Q(\theta, \mathcal{G}^\Pi) > q\}$. $Q(\theta, \mathcal{G}^\Pi) - Q(\theta, \mathcal{G}^W)$ is quasi–increasing on I^*, so $Q_W^{-1}(q) - Q_\Pi^{-1}(q)$ is weakly quasi–increasing. The difference between the distribution functions is

$$F_W(q) - F_\Pi(q) = E[\mathbf{1}_{\theta\leq Q_W^{-1}(q)} - \mathbf{1}_{\theta\leq Q_\Pi^{-1}(q)}|I^*]$$

Hence $F_W(q) - F_\Pi(q)$ is weakly quasi–increasing. □

For $I^* \in \mathcal{I}^*$, Lemmas 8–9 imply that $[Q(\theta, \mathcal{G}^W)|I^*] \geq_{icv} [Q(\theta, \mathcal{G}^\Pi)|I^*]$, where $\geq_{icv}$ denotes the increasing–concave order (Shaked and Shanthikumar (1994, Theorem 3.A.12(b))). The increasing–concave order is closed under mixtures so $Q(\theta, \mathcal{G}^W) \geq_{icv} Q(\theta, \mathcal{G}^\Pi)$ (Shaked and Shanthikumar (1994, Theorem 3.A.5(b))).

A.4 Monotone Comparative Statics used in Appendix A.5

The function $h(\theta, t)$ is *extended–log–supermodular* if for $\theta_H \geq \theta_L$ and $t_H \geq t_L$,

$$h(\theta_H, t_H)h(\theta_L, t_L) \geq h(\theta_H, t_L)h(\theta_L, t_H) \tag{A.1}$$

If $h(\theta, t)$ is also positive, then it is *log–supermodular.*

Lemma 10. *Suppose $h(\theta, t)$ is extended–log–supermodular, weakly quasi–increasing in θ and weakly quasi–increasing in $-t$. Then for any partition $\mathcal{I}$, $E[h(\theta, t)|\mathcal{I}]^+$ is log–supermodular.*

Proof. First, we show that the properties of $h(\theta, t)$ carry over to $\psi(\theta, t) := E[h(\theta, t)|\mathcal{I}]$. Suppose $h(\theta, t)$ is extended–log–supermodular and pick $t_H > t_L$ and $\theta_H > \theta_L$, where $\theta_H \in I_H$ and $\theta_L \in I_L$.

$$\begin{aligned}\psi(\theta_H, t_H)\psi(\theta_L, t_L) &- \psi(\theta_L, t_H)\psi(\theta_H, t_L)\\ &= \int_{I_L}\int_{I_H} [h(\theta_H, t_H)h(\theta_L, t_L) - h(\theta_L, t_H)h(\theta_H, t_L)]\, dF(\theta_H)dF(\theta_L) \geq 0\end{aligned}$$

using the extended–log–supermodularity of $h(\theta, t)$. Similarly, if $h(\theta, t)$ is weakly quasi–increasing in a parameter then $\psi(\theta, t)$ has the same property.

Second, we show that $\psi(\theta,t)^+$ is log–supermodular. Pick $\theta_H > \theta_L$ and $t_H > t_L$. If $\psi(\theta_L,t_H) \leq 0$, then $\psi(\theta,t)^+$ is trivially log–supermodular. If $\psi(\theta_L,t_H) > 0$, then $\psi(\theta_H,t_H) \geq 0$ and $\psi(\theta_L,t_L) \geq 0$ from the monotonicity properties of $\psi(\theta,t)$. The log–supermodularity of $\psi(\theta,t)^+$ follows from the extended–log–supermodularity of $\psi(\theta,t)$. □

Lemmas 11–12 are variants of Karlin and Rubin (1956, Lemma 1). The method of proof is identical.

Lemma 11. *Consider groups $\mathcal{G}_L, \mathcal{G}_H$ such that $\mathcal{I}(\mathcal{G}_L) \subset \mathcal{I}(\mathcal{G}_H)$ and assume $Q(\theta,\mathcal{G}) \geq 0$. Suppose that $E[h(\theta,t)|\mathcal{I}(\mathcal{G}_H)]^+$ is log–supermodular in (θ,t) and decreasing in t. Consider the partition $\mathcal{I}^* \subset \mathcal{I}(G_L)$ such that $Q(\theta,\mathcal{G}_L) - Q(\theta,\mathcal{G}_H)$ is quasi–increasing on every $I \in \mathcal{I}^*$. Then*

$$E\Big[E[h(\theta,t)|\mathcal{I}(\mathcal{G}_H)]^+ Q(\theta,\mathcal{G}_H)\Big|\mathcal{I}^*\Big] - E\Big[E[h(\theta,t)|\mathcal{I}(\mathcal{G}_L)]Q(\theta,\mathcal{G}_L)\Big|\mathcal{I}^*\Big] \tag{A.2}$$

is quasi–increasing in t.

Proof. Write $\psi(\theta,t) := E[h(\theta,t)|\mathcal{I}(\mathcal{G}_H)]$ and $\Delta Q(\theta) := Q(\theta,\mathcal{G}_H) - Q(\theta,\mathcal{G}_L)$. Rewriting (A.2) we wish to show that

$$E\Big[\psi(\theta,t)^+\Delta Q(\theta)\Big|\mathcal{I}^*\Big] + E\Big[\psi(\theta,t)^- Q(\theta,\mathcal{G}_L)\Big|\mathcal{I}^*\Big] \tag{A.3}$$

is quasi–increasing in t. By way of contradiction, suppose there exists $t_H > t_L$ and an interval $I \in \mathcal{I}^*$

$$E\Big[\psi(\theta,t_L)^+\Delta Q(\theta)\Big|I\Big] + E\Big[\psi(\theta,t_L)^- Q(\theta,\mathcal{G}_L)\Big|I\Big] \geq 0 \tag{A.4}$$

and

$$E\Big[\psi(\theta,t_H)^+\Delta Q(\theta)\Big|I\Big] + E\Big[\psi(\theta,t_H)^- Q(\theta,\mathcal{G}_L)\Big|I\Big] < 0 \tag{A.5}$$

Since $\Delta Q(\theta)$ is increasing on I, so we can break it up into positive and negative components. That is, $\Delta Q(\theta) \geq 0$ on some $I^+ \in \mathcal{I}_H$ and $\Delta Q(\theta) < 0$ on $I^- := I \backslash I^+$. For notational convenience, restrict the state space to I and rewrite (A.4) and (A.5) as

$$E\Big[\psi(\theta,t_L)^+\Delta Q(\theta)^+\Big] + E\Big[\psi(\theta,t_L)^- Q(\theta,\mathcal{G}_L)\Big] \geq E\Big[\psi(\theta,t_L)^+\Delta Q(\theta)^-\Big] \tag{A.6}$$

and

$$E\Big[\psi(\theta,t_H)^+\Delta Q(\theta)^-\Big] > E\Big[\psi(\theta,t_H)^+\Delta Q(\theta)^+\Big] + E\Big[\psi(\theta,t_H)^- Q(\theta,\mathcal{I}_L)\Big] \tag{A.7}$$

There are two possible cases. First, suppose that the left hand side of (A.6) equals zero. Then (A.6) implies the left hand side of (A.6) is also zero and, since $\psi(\theta,t)$ is decreasing in t, the left hand side of (A.7) is zero. We thus obtain a contradiction. Second, we suppose the left hand side of (A.6) is positive. Multiplying (A.6) and (A.7),

$$E\Big[\psi(\theta,t_H)^+\Delta Q(\theta)^-\Big]E\Big[\psi(\theta,t_L)^+\Delta Q(\theta)^+\Big]+E\Big[\psi(\theta,t_H)^+\Delta Q(\theta)^-\Big]E\Big[\psi(\theta,t_L)^-Q(\theta,\mathcal{I}_L)\Big] \quad \text{(A.8)}$$
$$> E\Big[\psi(\theta,t_H)^+\Delta Q(\theta)^+\Big]E\Big[\psi(\theta,t_L)^+\Delta Q(\theta)^-\Big]+E\Big[\psi(\theta,t_L)^+\Delta Q(\theta)^-\Big]E\Big[\psi(\theta,t_H)^-Q(\theta,\mathcal{I}_L)\Big]$$

We now show that (A.8) also yields a contradiction. This follows from two facts. First, using the log–supermodularity of $\psi(\theta,t)^+$,

$$E\Big[\psi(\theta,t_H)^+\Delta Q(\theta)^+\Big]E\Big[\psi(\theta,t_L)^+\Delta Q(\theta)^-\Big]-E\Big[\psi(\theta,t_H)^+\Delta Q(\theta)^-\Big]E\Big[\psi(\theta,t_L)^+\Delta Q(\theta)^+\Big] \quad \text{(A.9)}$$
$$=\int_{I_L^-}\int_{I_L^+}[\psi(\theta_H,t_H)^+\psi(\theta_L,t_L)^+-\psi(\theta_H,t_L)^+\psi(\theta_L,t_H)^+]\Delta Q(\theta_H)^+\Delta Q(\theta_L)^-dF(\theta_H)dF(\theta_L)\geq 0$$

Second, $\psi(\theta,t)$ is decreasing in t. Hence $\psi(\theta,t_L)^+\geq\psi(\theta,t_H)^+$ and $\psi(\theta,t_H)^-\geq\psi(\theta,t_L)^-$. This means

$$E\Big[\psi(\theta,t_L)^+\Delta Q(\theta)^-\Big]E\Big[\psi(\theta,t_H)^-Q(\theta,\mathcal{G}_L)\Big]\geq E\Big[\psi(\theta,t_H)^+\Delta Q(\theta)^-\Big]E\Big[\psi(\theta,t_L)^-Q(\theta,\mathcal{G}_L)\Big] \quad \text{(A.10)}$$

Together, (A.9) and (A.10) contradict (A.8), as required. □

Lemma 12. *Suppose $\Delta Q(\theta)$ is quasi–increasing on I. In addition, suppose that $h(\theta,t)$ is log–supermodular in (θ,t) and decreasing in t. Then $E[h(\theta,t)\Delta Q(\theta)|I]$ is quasi–increasing in t.*

Proof. Follows from Lemma 11. □

A.5 Proof of Proposition 3

The method of proof is the same as in Proposition 1. Suppose $\mathcal{G}^W$ maximises welfare and pick G such that $\mathcal{G}\subset\mathcal{G}^W$. We wish to show that $\Pi(\mathcal{G}^W)\geq\Pi(\mathcal{G})$. By Lemma 3, $\mathcal{I}(\mathcal{G})\subset\mathcal{I}(\mathcal{G}^W)$. Denote the benefit from splitting, conditional on $\mathcal{I}(\mathcal{G})$, by

$$\begin{aligned}\Delta W(\theta) &:= E\Big[\theta Q^*(\theta,\mathcal{G}^W)\Big|\mathcal{I}(\mathcal{G})\Big]-E\Big[\theta Q^*(\theta,\mathcal{G})\Big|\mathcal{I}(\mathcal{G})\Big]\\ \Delta\Pi(\theta) &:= E\Big[E[MR(\theta)|\mathcal{I}(G^W)]^+Q^*(\theta,\mathcal{G}^W)\Big|\mathcal{I}(\mathcal{G})\Big]-E\Big[E[MR(\theta)|\mathcal{I}(\mathcal{G})]^+Q^*(\theta,\mathcal{G})\Big|\mathcal{I}(\mathcal{G})\Big]\end{aligned}$$

Let $\Delta Q^*(\theta):=Q^*(\theta,\mathcal{G}^W)-Q^*(\theta,\mathcal{G})$. Since $\mathcal{G}^W$ maximises welfare, $E[\Delta W(\theta)]\geq 0$. Observe that $\Delta Q^*(\theta)$ is increasing on each $I\in\mathcal{I}(\mathcal{G})$ and let $\mathcal{I}^*$ be the coarsest partition such that $\Delta Q^*(\theta)$ is quasi–increasing for all $I^*\in\mathcal{I}^*$.

Lemma 13. $E[\Delta W(\theta)|\mathcal{I}^*]\geq 0$.

Proof. Same as Proposition 1. □

Lemma 14. $E[\Delta\Pi(\theta)|\mathcal{I}^*] \geq 0$ *and hence* $E[\Delta\Pi(\theta)] \geq 0$.

Proof. Let us divide the state space into two. $MR(\theta)$ is quasi–increasing in θ, so $E[MR(\theta)|\mathcal{I}(\mathcal{G})]$ is positive on some increasing set $I_B \subset [\underline{\theta}, \overline{\theta}]$, and strictly negative on the compliment, I_A.

First, $E[\Delta W(\theta)|\mathcal{I}^*] \geq 0$ implies $E[\Delta W(\theta)\mathbf{1}_{I_B}|\mathcal{I}^*] \geq 0$. To see this, notice that the function $h(\theta, t) = \theta\mathbf{1}_{\theta \geq t}$ is log–supermodular and decreasing in t. Let $h(\theta, 0) = \theta$ and $h(\theta, 1) = \theta\mathbf{1}_{I_B}$ and apply Lemma 12.

Second, let $h(\theta, 0) = \theta\mathbf{1}_{I_B}$ and $h(\theta, 1) = MR(\theta)\mathbf{1}_{I_B}$. Under (MON), $h(\theta, t)$ is extended–log–supermodular in (θ, t), weakly quasi–increasing in θ and decreasing in t. Lemma 10 implies that $E[h(\theta, t)|\mathcal{I}(\mathcal{G}^W)]^+$ is log–supermodular and decreasing in t. Hence, by Lemma 11, $E[\Delta W(\theta)\mathbf{1}_{I_B}|\mathcal{I}^*] \geq 0$ implies that $E[\Delta\Pi(\theta)\mathbf{1}_{I_B}|\mathcal{I}^*] \geq 0$.

Third, $E[MR(\theta)\mathbf{1}_{I_A}|\mathcal{I}(\mathcal{G})]^+ = 0$, so $E[\Delta\Pi(\theta)\mathbf{1}_{I_A}|\mathcal{I}^*] \geq 0$. Thus $E[\Delta\Pi(\theta)\mathbf{1}_{I_B}|\mathcal{I}^*] \geq 0$ implies $E[\Delta\Pi(\theta)|\mathcal{I}^*] \geq 0$. □

References

Alchian, A. A., and H. Demsetz (1972): "Production, Information Costs, and Economic Organization," *American Economic Review*, 62(5), 777–795.

Arnott, R., and J. Rowse (1987): "Peer Group Effects and Educational Attainment," *Journal of Public Economics*, 32(3), 287–305.

Benabou, R. (1993): "Workings of a City: Location, Education, and Production," *Quarterly Journal of Economics*, 108(3), 619–652.

Caucutt, E. M. (2002): "Educational Vouchers When There Are Peer Group Effects—Size Matters," *International Economic Review*, 43(1), 195–222.

Damiano, E., and H. Li (2005): "Competing Matchmaking," Working paper, University of Toronto.

——— (2006): "Price Discrimination and Efficient Matching," *Economic Theory*, forthcoming.

Durrett, R. (1995): *Probability: Theory and Examples*. Duxbury Press.

Epple, D., and R. E. Romano (1998): "Competition between Private and Public Schools, Vouchers, and Peer-Group Effects," *American Economic Review*, 88(1), 33–62.

Falk, A., and A. Ichino (2006): "Clean Evidence on Peer Effects," *Journal of Labor Economics*, 24(1), 39–57.

Hanushek, E. A. (1999): "Some Findings from an Independent Investigation of the Tennessee STAR Experiment and from Other Investigations of Class Size Effects," *Education Evaluation and Policy Analysis*, 21(2), 143–163.

Hanushek, E. A., J. F. Kain, J. M. Markman, and S. G. Rivkin (2003): "Does Peer Ability Affect Student Achievement?," *Journal of Applied Econometrics*, 18(5), 527–544.

Helsley, R. W., and W. C. Strange (2000): "Social Interactions and the Institutions of Local Government," *American Economic Review*, 90(5), 1477–1489.

Henderson, J. V., P. Mieszkowski, and Y. Sauvageau (1978): "Peer Group Effects and Educational Production Functions," *Journal of Public Economics*, 10(1), 97–106.

Karlin, S., and H. Rubin (1956): "The Theory of Decision Procedures for Distributions with Monotone Likelihood Ratio," *The Annals of Mathematical Statistics*, 27(2), 272–299.

KREMER, M. (1993): "The O-Ring Theory of Economic Development," *Quarterly Journal of Economics*, 108(3), 551–575.

LADD, H. F. (2002): "School Vouchers: A Critical View," *Journal of Economic Perspectives*, 16(4), 3–24.

LAZEAR, E. P. (2001): "Educational Production," *Quarterly Journal of Economics*, 116(3), 777–803.

MAS, A., AND E. MORETTI (2006): "Peers at Work," Working paper, UC Berkeley.

MAS-COLELL, A., M. D. WHINSTON, AND J. R. GREEN (1995): *Microeconomic Theory.* Oxford University Press.

MILGROM, P., AND I. SEGAL (2002): "Envelope Theorems for Arbitrary Choice Sets," *Econometrica*, 70(2), 583–601.

MYERSON, R. (1981): "Optimal Auction Design," *Mathematics of Operations Research*, 6(1), 58–73.

NECHYBA, T. J. (2000): "Mobility, Targeting, and Private-School Vouchers," *American Economic Review*, 90(1), 130–146.

PERSICO, N. (2000): "Information Acquisition in Auctions," *Econometrica*, 68(1), 135–148.

PESENDORFER, W. (1995): "Design Innovation and Fashion Cycles," *American Economic Review*, 85(4), 771–792.

RAYO, L. (2005): "Monopolistic Signal Provision," *B.E. Journals in Theoretical Economics: Contributions.*

SACERDOTE, B. (2001): "Peer Effects with Random Assignments: Results for Dartmouth Roommates," *Quarterly Journal of Economics*, 116(2), 681–704.

SCOTCHMER, S. (2002): "Local Public Good and Clubs," *in Handbook of Public Economics*, pp. 1997–2042, (A. J. Auerback and M. Feldstein, eds.).

SHAKED, M., AND J. G. SHANTHIKUMAR (1994): *Stochastic Orders and Their Applications.* Academic Press, San Diego.

ZIMMERMAN, D. (2003): "Peer Effects in Academic Outcomes: Evidence from a Natural Experiment," *Review of Economics and Statistics*, 85(1), 9–23.