Essays on Social Networks, Information and Organisations

by

Andrew Harkins

A thesis submitted in partial fulfilment of the requirements
for the degree of
Doctor of Philosophy in Economics

Department of Economics

University of Warwick

September 2014

Contents

List of Figures v

List of Tables vi

Acknowledgements vii

Declaration viii

Abstract ix

1 Network Games with Perfect Complements 1

1.1 Introduction 1

1.2 Related Literature 3

1.3 Model 5

1.3.1 Cohesive Subgroups 7

1.3.2 The Pareto Dominant Equilibrium 10

1.3.3 Other Nash Equilibria 12

1.4 Stable Equilibria 13

1.4.1 Stability of the Pareto Dominant Equilibrium 14

1.4.2 The Potential Maximising Equilibrium 16

1.5 Socially Efficient Networks and the Price of Stability 21

1.5.1 Socially Efficient Networks with Linear Link Costs 23

1.6 Conclusion 24

2 Pre-Launch Publicity and Third Party Reviewers **34**
2.1 Introduction 34
2.2 Related Literature 36
2.3 The Optimal Launch Campaign 40
2.3.1 Information Control with Multiple Receivers and Private Signals 43
2.3.2 The Seller's Choice of Launch Campaign 49
2.3.3 Endogenous Pricing 55
2.4 A Simple Model of Third Party Reviewers 57
2.4.1 The Seller's Optimal Reviewer 60
2.4.2 The Buyers' Optimal Reviewer 64
2.5 Discussion and Conclusion 66

3 Costly Communication and Organisational Attention **79**
3.1 Introduction 79
3.2 Literature 81
3.3 Model 84
3.3.1 Endogenous Attention 90
3.3.2 Endogenous Attention in the Star Network 91
3.3.3 Endogenous Attention and Span of Control 94
3.4 Conclusion 96

List of Figures

1.1 The *k-cores* of a Network . . . 8

1.2 Coreness Profile of a Bridge Network . . . 10

1.3 Coreness and Bonacich Centralities of Nodes in a Bridge Network . . . 12

1.4 Some Equilibria of the Game for the Network in Figure 1.1 . . . 13

1.5 An Unstable Equilibrium . . . 14

1.6 Stable Coreness Profile . . . 16

1.7 Density Decomposition and Potential Maximising Partition . . . 21

2.1 Timing of the Model . . . 43

2.2 Summary of Private Signals and Launch Campaign . . . 47

2.3 Expected Buyers as a Function of the Posterior Belief μ_s . . . 48

2.4 Optimal Launch Campaign with $\lambda = 0.7$, $\mu = 0.7$ . . . 52

2.5 Optimal Launch Campaign with $\lambda = 0.5$, $\mu = 0.2$ . . . 53

2.6 The PDF and CDF for $\tilde{X}$ . . . 60

2.7 Effect of λ on the Seller's Choice of Reviewer . . . 63

3.1 A Three Person Organisation . . . 91

3.2 Tree Networks with Differing CEO Spans of Control . . . 94

List of Tables

2.1 Summary of Optimal Launch Campaigns 54

Acknowledgements

I would like to thank my supervisors, Bhaskar Dutta and Daniel Sgroi, for their advice and patience during the time I have spent at Warwick.

I also thank my wife Caroline. Without her support, dedication and encouragement, this would not have been possible.

Declaration

I declare that this thesis is my own work and has not been submitted for a degree at another university.

Andrew Harkins

September 2014

Abstract

In the first chapter of this thesis I bring two well-known concepts from sociology and network science into the literature on network games, by microfounding the notions of the *k-core* and *coreness*. I show that these concepts arise naturally out of a simple threshold game played on a network. I also analyse the stability properties of equilibria, borrowing ideas from evolutionary game theory. The Pareto dominant equilibrium is shown to be unstable and prone to unravelling, so vulnerable nodes in the network are identified. This model can be applied to technology adoption decisions within firms and user engagement in social networks.

The second chapter focuses on how firms manage public beliefs about the value of their product. I use a 'Bayesian persuasion' approach to investigate how they might optimally design a publicity campaign to maximise sales. The optimal campaign will depend on the accuracy of private information, with more broadly focused campaigns preferred when consumers are less certain of their valuation. I also look at how the behaviour of external reviewers influences this decision. Reviewers have varying standards to pass their test and so I analyse the optimal reviewer in this context, highlighting cases where buyers can prefer softer reviewers than sellers.

The third chapter unites these themes by examining how the internal structure of the firm influences its ability to adapt itself to the external environment. Following Herbert Simon, I assume that a major obstacle in this task is insufficient attention to the relevant information. Since attention is a costly and scarce resource for employees, I examine how their endogenous allocation of attention impacts the organisation's performance. I follow the 'team theory' approach to show that more broadly defined tasks exhibit inertia in decision making, also relating this to recent empirical results regarding how CEOs allocate their attention.

Chapter 1

Network Games with Perfect Complements

1.1 Introduction

A feature which characterises many of the decisions made by individuals, groups and firms is the importance of coordination of our action with the actions of others. Coordination problems of this sort are pervasive in economics and a recent literature has developed which examines the role of the structure of an individual's social network in determining the likely outcome of such problems. These strategic complementarities in action can often be imperfect, such as the case of investing effort in a team project, since extra effort by one person may be an imperfect substitute for another person's lack of effort. In many other situations however, efforts are not substitutable and the benefits derived will hinge on those who contribute least, a situation known in the literature as *perfect complementarity*. For example, the benefit gained from socialising with friends or co-workers is strictly limited by the socialising efforts of other individuals with whom you interact.

Another example of where this assumption applies is to user engagement in online social networks. Activity is costly in terms of time and generates content which is non-rival but non-substitutable, as increasing activity beyond that of others does not increase utility gained from interacting with others. Examples of this from within organisations could include the use of 'knowledge sharing' tools such as collaborative online workspaces, company wikis or document management systems, where content is created at a private cost but is only beneficial for the user if others also contribute. The model also applies to network goods and technology adoption more broadly, such as individuals investing effort in learning a specialised technical language or a new piece of software, with the hope of others using it.

In this model agents located in social network will invest in a capital or effort variable which is privately costly but can provide benefits if their neighbours also invest. The benefit accrued across any pairwise relationship between agents will depend on the minimum investment made by either party. In terms of their best response functions, individuals will decide to increase or decrease investment depending on the number of neighbours who have put in weakly higher levels of investment, since benefits will be determined by the least investment made by either party in a pairwise relationship.

The main contribution of the chapter is a characterisation of two salient equilibria of the game in terms of well known concepts from the social network analysis literature. The Pareto dominant equilibrium of the model is shown to depend on the *coreness* of an individual in the network, a concept used first in the sociology literature. Coreness can be regarded as a course measure of centrality, but it is distinct from other notions which have preceded it in literature. In a well known model, Ballester et al. (2006) show players in a linear-quadratic network game with linear best responses play equilibrium strategies which depend on their 'Bonacich centrality'. They therefore provide a 'microfoundation' of a well known and widely use notion from the sociology literature. I provide a similar microfoundation for another well known concept from the sociology and network science literature by showing that it arises naturally from a threshold game played on networks.

I also look at equilibrium stability and identify nodes which can be targeted for subsidies in order to prevent the more favourable Pareto dominant equilibrium from unravelling. Finally, I characterise the most stable equilibrium of the game in terms of a concept which is closely related to stochastic stability in the evolutionary game theory literature. I show that the equilibrium of the game which can be expected to arise most frequently in the long-run is related to a well known problem in computer science, known as the 'densest subgraph' problem. The contribution of this chapter therefore overlaps into the multiple different disciplines, including sociology and computer science.

This model falls under the class of 'threshold games', since the optimality of a particular effort level will depend on the number of individuals also playing that effort level or higher. As we increase the effort level, a larger number of neighbours is required to sustain this as an optimal decision. The main difference between the set-up of this model and that of Young (1998) for example, is that costs are only incurred once at the node level rather than at the edge level, and so thresholds for action depend on the absolute number of neighbours playing a given strategy, rather than a proportion. There will be a marked difference in equilibrium patterns of play as agents now wish to coordinate, not with a small cabal of insiders, but with larger 'core' groups of the network.

I now proceed with a review of the relevant literature on network games before moving on to a description of the model. The following sections then solve for equilibria and examine the stability properties of these equilibria.

1.2 Related Literature

The growing literature looking at games played on networks has been surveyed by Jackson (2008) and more recently by Jackson and Zenou (2012), so I present a brief summary of papers directly relevant to the present model.

The bulk of the literature examines the relationship between specific structural characteristics of the network and Nash equilibria of complete information games.[1] Earlier models which examine the relationship between the structure of a social network and the equilibria of coordination problems have often framed the decision as a binary one e.g. adopt new technology/don't adopt, revolt/stay at home, withdraw savings/don't withdraw etc. This chapter differs by examining how the structure of social interaction may influence decisions in the case where one-off investments are made in a continuous 'capital' variable by individuals at private cost, and benefits are realised through pair-

[1] A notable exception is Galeotti et al. (2010) who develop a tractable model of equilibrium play when knowledge of the network is incomplete.

wise interaction with other individuals' capital. In that respect, this chapter is most closely aligned with the model of Ballester et al. (2006), but considers neighbouring levels of investment to be perfectly complementary with one's own level.

Several key contributions have identified a link between the network position of agents and their equilibrium actions, with a particular focus on different measures of network centrality. A foundational contribution to this literature is Ballester et al. (2006), which was the first model to clearly identify the link between the equilibria of games played on networks and some variant of the eigenvector centrality of agents. In their model, the marginal utility from exerting effort is a linear function of their neighbours' efforts and costs are quadratic and privately borne. This gives rise to best response functions which are linear in the actions of other agents and can be solved uniquely by agents selecting their Bonacich centralities as efforts. Another model which links an agent's centrality in the network to their equilibrium action is Calvó-Armengol et al. (2014). They consider a different setting in which agents may invest in active and passive communication and have quadratic loss functions which are minimised when actions are matched with their local state and the actions of others. Equilibrium actions and communication efforts in their model are found to depend on a measure of centrality named the Invariant Method index, which bears some similarity to the Bonacich centrality measure.

A second strand of the literature on network games of strategic complementarity which is related to the results presented in this chapter is that associated with so-called 'threshold games'. A typical setting is one similar to that described by Granovetter (1978) where a group of individuals face a collective action problem in the form of a binary choice, e.g. either to strike or not to strike, but prefer only to take the action if at least some threshold percentage of the group do the same. Differences in thresholds can lead to cascading behaviours, where one individual switching strategy forces others to switch, leading to yet more switching until a new equilibrium is reached. Similar problems were also formally analysed in an earlier paper by Schelling (1973), who also discusses some consequences of the spatial 'configuration' on cascades and equilibrium outcomes.

Chwe (2000) examines how the communication structure of a society might enable individuals with heterogeneous thresholds to more easily coordinate on their preferred equilibrium in such a setting. His model shows that optimal networks can be formed from a series of interlocking cliques of agents, which allow all agents to take the risky action by making all locally important thresholds common knowledge.

The model presented in this chapter is perhaps most closely related to that of Morris (2000), who investigates the role of network structure in a binary decision threshold game when the modeller is concerned about the robustness of decision making with respect to contagion. Morris shows that only sufficiently inward looking groups of nodes can be resilient to an invading cascades of failures which occur in other areas of the network. The model of Young (1998) is also concerned with the structural conditions which enable different regimes of play in different areas of the network and arrive at a similar conclusion to Morris (2000).

Young (1998) also highlights a negative result with regard to the ability of any network structure to prevent a risk-dominant equilibrium from prevailing as the unique stochastically stable state of play. This is related to an earlier paper by Ellison (1993), which also examines the robustness of equilibria in a stag hunt game to trembles in decision making for some simple structures. The Ellison (1993) model has also been extended to the setting where the network is endogenous by Jackson and Watts (2002), who find that stochastically stable equilibria may arise which are neither risk-dominant nor Pareto-dominant when agents can select who they are linked to.

1.3 Model

Let N be a finite set of individuals indexed by $i = 1, \dots, n$ who are located on a network g, where g is a set of links or edges such that $(i, j) \in g$ implies that agent i and j are connected in the network. I use $g + ij$ to denote a graph g' which is formed from g plus the addition of an edge between nodes i and j, (i.e. $g' = g \cup (i, j)$). The neighbourhood of an agent i is denoted by $N_i(g) = \{j \in N \mid (i, j) \in g\}$ and their *degree* given by $d_i(g) = |N_i(g)|$. The network is undirected so $(i, j) \in g$ implies that agent

j and i are connected. These individuals gain utility from interaction with others to whom they are connected by selecting a level of activity $x_i \in [0, \bar{x}]$ but incur a private cost which is quadratic in x_i.[2] Unlike Ballester et al. (2006), who focus on linear best response functions where actions are local complements but global substitutes, this chapter considers the case where there is no substitutability in action across individuals. Utility functions therefore take the form

$$u_i(x_1, \ldots, x_n) = \sum_{j \in N_i} \min\{x_i, x_j\} - \frac{1}{2}x_i^2$$

The utility function is strictly concave and continuous in x_i and so achieves a unique maximum on $[0, \bar{x}]$ given a profile of the others' actions x_{-i}. Since the gross benefit from raising x_i increases linearly with the number of neighbours playing strictly higher levels of effort, let $N_i''(x_i, x_{-i})$ denote the set $\{j \in N_i \mid x_j > x_i\}$ and let $N_i'(x_i, x_{-i})$ denote $\{j \in N_i \mid x_j \geq x_i\}$. Accordingly, let $d_i'(g, x) \equiv |N_i'(x_i, x_{-i})|$ and $d_i''(g, x) \equiv |N_i''(x_i, x_{-i})|$.

This specification of utilities gives rise to a 'threshold game' where x_i can be a best response to x_{-i} only if the number of neighbours playing weakly higher effort is at least x_i, or only if

$$x_i \leq d_i'(g, x) \tag{1.1}$$

If this were not the case, then decreasing x_i lowers utility linearly along $d_i'(g, x)$ links, but lowers cost at rate x_i. Furthermore, since there is no substitutability in effort, increasing x_i brings benefits only along those links where agent i is pivotal (i.e. where $x_i < x_j$). Therefore, a best response x_i must also satisfy a second condition

$$x_i \geq d_i''(g, x) \tag{1.2}$$

[2] I assume throughout that $\bar{x} \geq \max_{i \in N} d_i(g)$

This second condition ensures that raising x_i cannot be beneficial and taken together with (1) these two sufficient conditions characterise the unique best response. The main difference between these conditions and those from other threshold games on networks is that optimal actions depend on the absolute number of neighbours playing a (weakly) higher action, rather than a proportion of a neighbourhood, since investment cost is split across all neighbours.

Since all players are best responding in equilibrium, the level of investment which can be sustained for agent i in equilibrium will depend not only on the number of neighbours playing a weakly higher action but also on the number of neighbours' neighbours who play a weakly higher action. This clearly implies that for a given agent to sustain a high x_i in equilibrium we require not only that they be highly connected, but that their neighbours and neighbours' neighbours be highly connected. As later results will demonstrate, dense and cohesive subgroups of the network will find it easier to sustain higher levels of investment in equilibrium.

1.3.1 Cohesive Subgroups

Notions of group cohesiveness in social networks have long been studied in the sociology literature and there are many intuitive concepts such as cliques, clans and clubs which are defined in standard texts such as Wasserman and Faust (1994). A variant of these which has been used in the economics literature is the notion of a p-cohesive subset, defined in Morris (2000). Formally, a subset of nodes is said to be p-cohesive if every node within that subset has (at least) a proportion p of their neighbours within that subset. A related idea is also found in Young (1998) where a subset of nodes S is called r-close-knit if for every $S' \subseteq S$ the proportion of links originating in S' and ending in S is at least r. Intuitively, p-cohesiveness is a condition on the degrees of nodes, whereas r-close-knittedness is a condition on links within a subgroup and therefore a p-cohesive subgroup is $p/2$-close-knit.

This chapter will employ a particularly useful concept originally defined by Seidman (1983) known as the *k-core*. Seidman (1983) considers subgraphs of g which can be induced by repeatedly pruning nodes of low degrees from the network in order uncover

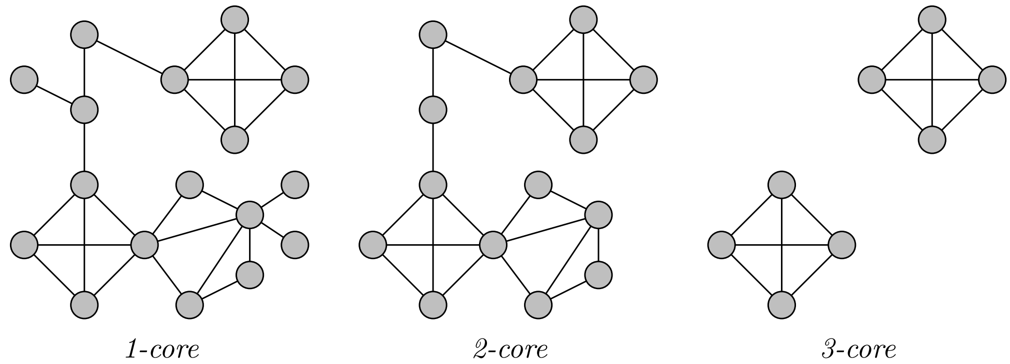


Figure 1.1: The *k-cores* of a Network

groups of densely connected individuals. The graph which is obtained by iteratively removing all nodes of degree less than k is known as a *core* of order k, or a k-core.[3] For any subgraph $g_k \subseteq g$, I will use N_k to denote the set of agents who have a strictly positive degree in that subgraph. A precise definition of a k-core of a graph g now follows:

Definition 1.1. A *k-core* of a graph g is a subgraph $g_k \subseteq g$ such that $d_i(g_k) \geq k$ for each $i \in N_k$.

A k-core is therefore a subgraph of g where every agent with positive degree in that subgraph has at least degree k. When I state that a group of nodes 'form' a k-core, this means that the subgraph consisting of these nodes and links between them is itself a k-core. If an agent i is contained within a k-core then this implies that they have at least k neighbours of degree k or greater.

Every connected graph trivially contains a 1-core, whilst the 2-core which can be formed using the least possible number of edges is the ring network. Also note that the definition implies that nodes belonging to k-cores of high orders are also members of lower orders, permitting a nested *k-core decomposition* of any given network. Figure 1.1 shows such a k-core decomposition of a graph for $1 \leq k \leq 3$.

There have been a number of applications of the concept of the k-core outside of economics, for example, in the analysis of protein networks in bioinformatics (e.g.

[3] Seidman (1983) in fact refers to the k-core as maximal subgraph which can be obtained by iteratively removing nodes of lower degree. I follow Wasserman and Faust (1994) and the more recent literature by referring to any core of order k as a k-core.

Bader and Hogue (2003) and Wuchty and Almaas (2005)) and in the visualisation of large networks in computer science (e.g. Baur et al. (2004)).

Since cores of successive orders are nested within the previous core we can define a *coreness* value for each $i \in N$:

Definition 1.2. A node $i \in N$ has *coreness* $c_i(g) = k$ if it is contained in a core of order k but not in a core of order k' for $k' > k$.

The coreness of a node can be interpreted as a coarse measure of its centrality, as we can view it as a condition on its degree centrality and the degree centrality of other nodes in their neighbourhood. Nodes with high coreness are likely to have important roles in the network since they have neighbours with high degrees (who in turn have neighbours with high degrees, etc). High coreness can often indicate that a given node is a member of a dense and cohesive subset of the network, since cliques of size n immediately form an $(n-1)$-core. The vector of coreness for all agents $i \in N$ will be denoted by $\mathbf{c}(g) = (c_1(g), \ldots, c_n(g))$.[4] Although the coreness profile of a network can give an indication of dense subsets of g, more information will be needed to establish cohesiveness of the network as a whole (e.g. the links between different cores).

As can be seen in the example in Figure 1.2, the coreness of individual nodes can depend on structural characteristics of the network which are relatively 'far away'. In this example, the addition of a single link between the remaining pair of nodes with degree 2 would raise the coreness of all nodes to 3. Although adding a link can increase the corenesses of distant nodes, the following lemma shows that increases in the coreness of any given node can be at best be directly proportional to increases in their degree.

Lemma 1.1. $c_l\,(g + ij) \leq c_l\,(g) + 1,\ \forall l \in N\ ,\forall ij.$

Proof. All proofs are contained in the Appendix. □

Adding an edge between i and j can increase c_i and c_j by at most 1. Although the removal of one link can have a cascading effect which influences all nodes (e.g the transition from ring to line network), we can also interpret Lemma 1.1 as saying that the

[4] In what follows I refer to each agent's coreness in the network g as simply c_i. I will indicate by using $c_i(g')$ when considering subgraphs of g.

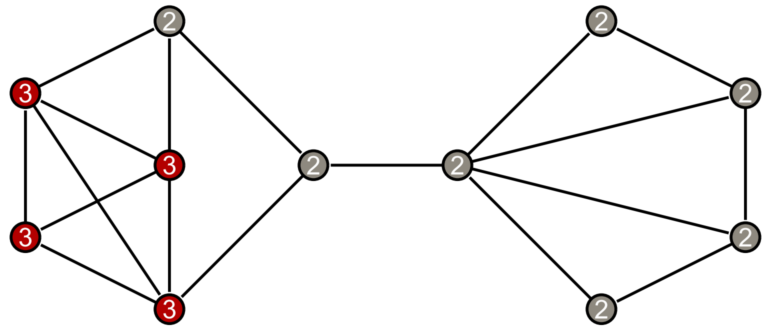


Figure 1.2: Coreness Profile of a Bridge Network

removal of an edge ij cannot lower the coreness of any agent in the network by more than 1. With these definitions in hand, I now discuss the properties of equilibrium action profiles.

1.3.2 The Pareto Dominant Equilibrium

Now that the coreness of nodes has been defined, the model can be solved for action profiles $\mathbf{x} \in X = [0, \bar{x}]^n$ which constitute a Nash equilibrium of the game $\Gamma = \langle N, X, u \rangle$. The game will have multiple equilibria and so I focus in this section on the Pareto dominant Nash equilibrium. An observation which can be made immediately is that Γ is a supermodular game.[5] Since the game Γ is supermodular, the results of Milgrom and Roberts (1990) show us that firstly a greatest and least equilibrium must exist, and secondly, if $\mathbf{x}$ and $\mathbf{x}'$ are equilibria of Γ where $\mathbf{x} \geqslant \mathbf{x}'$ then $\mathbf{x}$ Pareto dominates $\mathbf{x}'$. These statements follow directly from Theorems 5 and 7 (respectively) of Milgrom and Roberts (1990).[6] Since the set of equilibria of a supermodular game forms a complete lattice (Zhou, 1994), this immediately implies that the greatest equilibrium Pareto dominates all others. With the existence of a Pareto dominant equilibrium established, we can now characterise this equilibrium in terms of the coreness of agents.

Proposition 1.1. *The Pareto dominant Nash equilibrium of* Γ *is* $\mathbf{x}^* = \mathbf{c}(g)$

[5]Firstly, the strategy set $X = [0, \bar{x}]^n$ is a complete lattice using the usual partial order $\mathbf{x} \geqslant \mathbf{x}'$ if $x_k \geq x'_k$ for all $k = 1, \ldots, n$. By the definition of Milgrom and Roberts (1990), the game is supermodular since u_i has increasing differences in (x_i, x_{-i}), u_i is supermodular in x_i for fixed x_{-i}, and u_i is upper semi-continuous in X_i and order continuous in X_{-i} with a finite upper bound.

[6]The existence of an equilibrium can be trivially established since $x_i = 0$ for all i is always an equilibrium.

Intuitively, a coreness of c_i for agent i guarantees that they have at least c_i neighbours, who have at least c_i neighbours, etc, who could feasibly play c_i in equilibrium. Furthermore, having coreness c_i implies that $x_i > c_i$ can never be played in equilibrium as there is an insufficient number of supporting nodes along paths from i. Supermodularity allows us to infer that the equilibrium where $\mathbf{x}^* = \mathbf{c}$ is Pareto dominant, since it is maximal in terms of investment. The reader can also use the supermodularity property to verify that beginning with action profile $\bar{\mathbf{x}} = (\bar{x}, \ldots, \bar{x})$ and iterating the best responses of agents we arrive at $x_i = c_i$ for all i, implying that $\mathbf{x}^* = \mathbf{c}(g)$ is the maximal equilibrium (see Milgrom and Roberts (1990)).

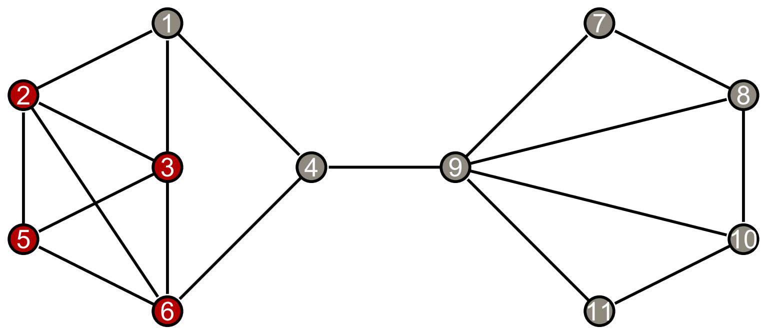
Proposition 1.1 therefore establishes a link between the centrality of a node in the network and the effort levels in the Pareto dominant equilibrium of this threshold game. A question which naturally arises is whether there is some relationship between coreness and the Bonacich centrality measures which play a key role in the models of Ballester et al. (2006) and Ballester and Calvó-Armengol (2010).

The Bonacich centrality of a node defined in Bonacich (1987) measures the number of paths of all lengths which originate at node i, weighted by a decay factor δ which decreases with the length l of the path. Formally, let $\mathbf{G}$ be the adjacency matrix of network g and define the Bonacich centrality of a node i as $b_i = \sum_{l=0}^{\infty} \delta^l \sum_{j \in N} g_{ij}^l$ where $\sum_{j \in N} g_{ij}^l$ is the sum of all paths of length l from i, (i.e. the sum across the i^{th} row of $\mathbf{G}^l$). The Bonacich centrality is finite whenever $\delta < \frac{1}{|\lambda_{max}(g)|}$ where $|\lambda_{max}(g)|$ is the absolute value of the largest eigenvalue of $\mathbf{G}$. This condition ensures that the sum does not grow too quickly as we iterate on powers of $\mathbf{G}$.

Since coreness says something about the number of paths of different lengths which originate at i, is it generally the case that $c_i \geq c_j$ implies that $b_i \geq b_j$? The answer turns out to be 'no', but we will be able to put a lower bound on the Bonacich centralities of nodes.

Proposition 1.2. *If node i has coreness $c_i(g) = k$ then*

$$b_i(g, \delta) \geq \frac{1}{1 - \delta k} \tag{1.3}$$



Node	Coreness	Bonacich Centrality ($\delta = 0.1$)	Bonacich Centrality ($\delta = 0.25$)
3	3	1.620	10.298
5	3	1.486	8.696
6	3	1.621	10.186
9	2	1.700	7.615

Figure 1.3: Coreness and Bonacich Centralities of Nodes in a Bridge Network

This bound is tighter when δ is low and the proof of Proposition 1.2 demonstrates that when g itself is a regular graph then (1.3) is met with equality. Although nodes with higher coreness will usually have higher Bonacich centralities this is not always the case, as demonstrated in Figure 1.3. Since the degree of node 9 in Figure 1.3 is greatest we find that $b_9(g,\delta) > b_j(g,\delta)$ for any other $j \in N$ when $\delta = 0.1$, since the paths of length 2 are weighted less by a factor of 10 when compared with paths of length 1.

1.3.3 Other Nash Equilibria

The extreme complementarity of actions in this model often results in multiple Nash equilibria. For example, the reader will notice that the profile $(0, 0, \ldots, 0)$ is always an equilibrium in any network. In fact, the game will have infinitely many Nash equilibria for any non-empty network.[7] This result is a consequence of the perfect complementarity assumption. The fact that each agent's effort cannot be used as in imperfect substitute for another agent's lack of effort leads to an inertia for a large number of action profiles. Figure 1.4 illustrates three equilibria for the network presented earlier in Section 3. The

[7] To show this, pick any equilibrium $\mathbf{x} \geqslant 0$ and consider the subset of agents playing highest effort in that equilibrium. Reducing x_i by a sufficiently small ε for each agent in this subset must also be a Nash equilibrium as conditions (1.1) and (1.2) must still hold. Similarly, if $\mathbf{x} = 0$ then increasing the effort of all agents by a sufficiently small ε is also an equilibrium.

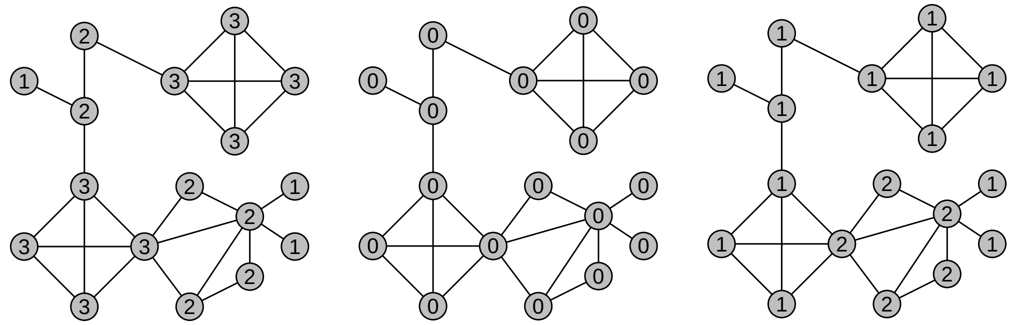


Figure 1.4: Some Equilibria of the Game for the Network in Figure 1.1

first panel shows the Pareto best equilibrium which is given by the coreness profile, whereas the final panel shows a Pareto inferior equilibrium. Observe that there are some equilibria where those with lower coreness play higher action in equilibrium.

Any action profile where each x_i lies in the interval $[d''_i(g,x), d'_i(g,x)]$ for each i will be a Nash equilibrium, yet some equilibria may seem more stable than others. In particular, in any graph, the equilibrium profile $(0, 0, \ldots, 0)$ requires only one node to deviate upwards to cause a cascade of increases, whereas this is not necessarily the case for equilibrium profile $(1, 1, \ldots, 1)$. This discussion highlights the fact that some equilibria may be more stable in the face of random shocks than others. It also therefore motivates a closer look at the stability of the Pareto dominant equilibrium and an attempt to refine some the predictions of the model.

1.4 Stable Equilibria

The focus so far on the Pareto dominant equilibrium $\mathbf{x}^*$ can be justified when considering environments which allow some degree of pre-play communication or third party mediation. Since actions are complements, it is in all agent's interests to coordinate on the Pareto dominant equilibrium. However, without communication or mediation it may seem unlikely that individuals could tacitly coordinate on the Pareto dominant equilibrium, especially if n is large. Furthermore, if tacit coordination is somehow achieved, then the question of stability of $\mathbf{x}^*$ with respect to random shocks to actions is also raised. The removal of a single link from the network can lead to a cascade of falling coreness values for all nodes in the network. In a similar manner, a temporary

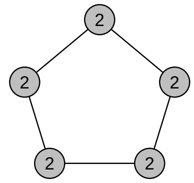


Figure 1.5: An Unstable Equilibrium

shock to the action of certain critical nodes in the network may result in a sequence of best responses which do not return to the original equilibrium $\mathbf{x}^*$.

Take the example in Figure 1.5, the coreness of all nodes in the network is $c_i = 2$ but the sequence of best responses following a temporary drop in the action of any node would converge to a new equilibrium where $x_i = 1$ for all i. A Pareto dominant equilibrium such as the one displayed in Figure 1.5 could not reasonably be considered stable in an environment where decisions are subject to infrequent random shocks.

To study the stability properties of equilibria in this game I now examine two environments. In the first, I examine the properties of networks where the Pareto dominant equilibrium is stable when encountering rare and isolated shocks to individual actions. Following this I will also apply a common refinement technique for similar 'minimum effort' games which uses the potential function to select equilibria which are most likely to be observed in the long run in the presence of persistent random shocks.

1.4.1 Stability of the Pareto Dominant Equilibrium

The notion of stability used in this subsection will be based on the convergence of sequences of best responses following a unilateral shock to some agent $i \in N$. A shock to equilibrium profile $\mathbf{x}$ is a profile $\hat{\mathbf{x}} \in X$ such that $\hat{x}_i = x_i + \varepsilon$ for exactly one $i \in N$ and $\hat{x}_i = x_i$ for all other $j \in N$, where $\varepsilon \in [-x_i, \infty)$. Following a shock, assume that play evolves over discrete time periods $t = \{0, 1, 2, \dots\}$ and define the best response dynamic as a sequence $\{\mathbf{x}^t\}$ in X such that for each $i \in N$ we have that $x_i^{t+1} = \text{argmax}_{x_i \in X_i} u_i\left(x_i, \mathbf{x}_{-i}^t\right)$. An equilibrium profile $\mathbf{x}$ will be considered stable against unilateral shocks if this sequence of myopic best responses returns to $\mathbf{x}$ following any shock $\hat{\mathbf{x}}$.

Definition 1.3. A profile $\mathbf{x}$ is *stable against unilateral shocks* if for any shock $\hat{\mathbf{x}}$, the sequence of myopic best responses with initial condition $\hat{\mathbf{x}}$ converges to $\mathbf{x}$

Focusing on the Pareto dominant equilibrium $\mathbf{x}^*$, we need not consider positive shocks to this profile as the supermodularity of Γ will guarantee that all resulting sequences of best responses converge back to the largest equilibrium $\mathbf{x}^*$. As I am interested in the stability of $\mathbf{x}^*$, I therefore restrict attention to cases where $\varepsilon \in [-x_i^*, 0]$.

Recall that g_k denotes a subgraph of g such that $d_i(g_k) \geq k$ for all $i \in N_k$. For a given $i \in N$ with coreness $c_i(g) = k$, let $\mathcal{G}_i$ be the intersection of all subgraphs $g_k \subseteq g$ such that $c_i(g_k) = c_i(g)$.

Definition 1.4. A node i is *critical* for a node j, written $i \to j$, if $(i, j) \in \mathcal{G}_j$

If $i \to j$ then the removal of edge (i, j) from the network will lower the coreness of node j. Moreover, if $i \not\to j$ then (i, j)'s removal cannot lower j's coreness as there exists a subgraph g' such that $c_j(g') = c_j(g)$ but $(i, j) \notin g'$. Note that there is always at least one critical node in any network, since $d_j(g) = c_j(g)$ for some $j \in N$, all $i \in N_j$ are critical for j, since the removal of any link $(i, j) \in g$ will lower j's degree and hence their coreness.

Using the relation $\to$ it is possible to construct a directed graph $\hat{g}$ where $(i, j) \in \hat{g}$ if and only if $(i, j) \in g$ and $i \to j$. The directed graph $\hat{g}$ is non-empty and identifies transmission paths of shocks which may propagate through the network as a result of the temporary lowering of the action of a given node. Since this graph is directed, it becomes necessary to define in-degrees and out-degrees for nodes in this network. I therefore use $d_i^+(\hat{g})$ to denote the in-degree of a node in $\hat{g}$, and $d_i^-(\hat{g})$ to denote their out-degree, with corresponding neighbourhoods $N_i^+(\hat{g})$ and $N_i^-(\hat{g})$.

We can also define a directed graph using the non-critical links, denoted by $\bar{g}$ where $(i, j) \in \bar{g}$ if and only if $(i, j) \in g$ but $i \not\to j$. A high out-degree for a node i in $\bar{g}$ will imply that there are a large number of neighbours who will not lower their action, even if i does. As will be shown, this will play a crucial role for the stability of an equilibrium action profile $\mathbf{x}$. Before stating the proposition in full, let $c_i^-(\bar{g})$ denote the out degree coreness of a node i in graph $\bar{g}$ in an analogous manner to the undirected degree coreness

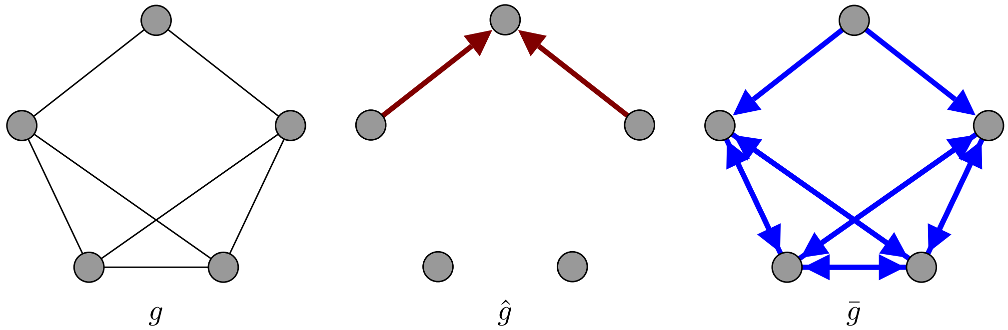


Figure 1.6: Stable Coreness Profile

measure from Section 1.3.[8]

Proposition 1.3. *The equilibrium* $\mathbf{x}^* = \mathbf{c}(g)$ *is stable against unilateral shocks if and only if* $c_i(g) = c_i^-(\bar{g})$ *for all* $i \in N$

Proposition 1.4 states that if there is a core whose members mutually support each other following a shock then this core is stable. Intuitively, it is a high out degree in $\hat{g}$ which prevents the action profile from converging back to $\mathbf{x}^*$ following a shock, as i's lower effort simultaneously causes the efforts of many neighbours to fall, preventing i from reverting back to x_i in later periods.

An alternative interpretation of Proposition 1.4 is that if we wish to protect certain nodes in the network from shocks or perhaps subsidise their efforts, then we should target those where $c_i(g) > c_i^-(\bar{g})$. If these nodes receive a shock to their action then this will lead to failures of other nodes which cannot be recovered from. If we wish to avoid even temporary disruptions to the action profile $\mathbf{x}^*$ then we should protect or subsidise all *critical* nodes in the network, since it is only these nodes which can propagate the initial shocks.

1.4.2 The Potential Maximising Equilibrium

The setting for the previous subsection has assumed that shocks are infrequent and idiosyncratic, and so the issue of correlated shocks or persistent randomness actions has yet to be addressed. It should be noted that since there is a continuum of equilibria

[8]That is, $c_i^-(\bar{g}) = k$ for $i \in N$ if and only if there exists a directed subgraph $\bar{g}_k \subseteq \bar{g}$ such that $d_i^-(\bar{g}_k) \geq k$ and $d_j^-(\bar{g}_k) \geq k$ for all j where $d_j^-(\bar{g}_k) > 0$, but no such subgraph exists for $k + 1$.

in the model, standard refinement techniques such as Nash tattonement (Bramoullé and Kranton, 2007) or asymptotic stability cannot be of use, since for any equilibrium $\mathbf{x}$ there always exists another within any arbitrarily small neighbourhood. With this in mind, I now discuss a refinement which has been successful for similar games in the experimental economics literature, and which takes into account some notion of persistent shocks to actions.

Experimental studies in Van Huyck et al. (1990), Goeree and Holt (2005), Chen and Chen (2011) and others[9] have identified the remarkable effectiveness of the potential function as a tool for equilibrium selection in 'minimum effort' coordination games. Despite the infinite number of equilibria, these studies demonstrate that the Nash equilibria which maximise the potential function of the game tend to be observed experimentally for a variety of parameters. Crawford (1991) puts forward an evolutionary explanation of these results, which is further strengthened by Anderson et al. (2001), who show that the distribution of strategies in the logit equilibrium maximises their stochastic potential function.

For games with discrete action sets, Blume (1993) has also shown that set of stochastically stable equilibria under the logit best response dynamic is equal to the set of potential maximizers as noise levels tend to zero. As noted by Goeree and Holt (2005), one can view the potential maximizing equilibria as being a close analogue of the stochastically stable equilibria for the case of continuous action sets, since the potential maximizing equilibria have the largest basin of attraction for an important class of evolutionary dynamics (see Sandholm (2010)).[10]

As defined by Monderer and Shapley (1996), an exact potential function of a game is a function $\rho : X \to \mathbb{R}$ such that $\forall x_i, x_i' \in X_i$ and $\forall x_{-i} \in X_{-i}$

$$\rho\left(x_i, x_{-i}\right) - \rho\left(x_i', x_{-i}\right) = u_i\left(x_i, x_{-i}\right) - u_i\left(x_i', x_{-i}\right), \qquad (\forall i \in N)$$

It is readily checked that an exact potential function for the game Γ is given by:

[9] See Appendix F of Chen and Chen (2011) for details.

[10] For a suitably discretised version of the game, the equilibria which maximise the potential function coincide with the stochastically stable equilibria under the logit learning dynamic. See Blume (1993) and Young (1998).

$$\rho(\mathbf{x}) = \sum_{(i,j)\in g} \min\{x_i, x_j\} - \frac{1}{2}\sum_{i\in N} x_i^2 \tag{1.4}$$

Any action profile which maximises ρ in each coordinate direction is also a Nash equilibrium and so the set of profiles which globally maximise ρ are a subset of these equilibria. I now use the potential function as a tool for equilibrium refinement, as first suggested by Monderer and Shapley (1996). The results of Blume (1993) and Young (1998) have shown that in games with discrete actions, the action profiles which maximise the potential function are the only stochastically stable outcomes of the game. Therefore, I study the set of profiles $\mathbf{x}$ which globally maximise $\rho(\mathbf{x})$, viewing this as an approximation to the stochastically stable outcome of the game, if it were suitably discretised.[11]

Notice that the potential function in (1.4) necessarily inherits the properties of the utility functions u_i (e.g. it has increasing differences in x_i). The following lemma shows that it is also supermodular and strictly concave on X.

Lemma 1.2. *The potential function ρ is:*

(a) *Supermodular on X*

(b) *Strictly concave on X*

Since the ρ is a supermodular function defined on a lattice we can establish that a maximum exists (Topkis (1998)) and so the potential function ρ has a unique maximum given strict concavity.

In order to find the potential maximising equilibrium I take advantage of the hierarchical nature of the equilibrium profiles. The optimal actions of those playing the largest x_i in equilibrium cannot be influenced by the actions of those playing strictly lower x_i, since only those playing the lowest x_i along any edge (i, j) are pivotal. This fact can be exploited to find the Nash equilibrium which maximises the potential function ρ by first partitioning the set of agents N according to their equilibrium actions in the potential maximising profile $\tilde{\mathbf{x}}$. Therefore, given a profile $\tilde{\mathbf{x}}$, let $\{\tilde{S}_1, \ldots, \tilde{S}_K\}$ be a

[11]The use of stochastic stability as a refinement tool is common in the literature, however the concept of stochastic stability has yet to be extended to the case of continuous actions.

partition of N such that $i \in \tilde{S}_k$ and $j \in \tilde{S}_k$ if and only if $\tilde{x}_i = \tilde{x}_j$, indexing these subsets in descending order of their equilibrium action. Let $\tilde{g}_k \subseteq g$ denote the subgraph such that $(i,j) \in \tilde{g}_k$ if and only if $i \in \tilde{S}_k$ and $j \in \tilde{S}_{k'}$ for $k' \leq k$.

Focusing on the subset of agents who are in $\tilde{S}_1$, their optimal decision does not depend on the actions of agents in other subsets of N. The equilibrium actions of $\tilde{S}_1$ are only pivotal along links to other members of $\tilde{S}_1$, so we may optimise for members of $\tilde{S}_1$ whilst ignoring the actions of other subsets. Setting $\tilde{x}_i = \tilde{x}_j \equiv \tilde{x}_1$ for i and $j \in \tilde{S}_1$, the maximisation problem becomes

$$\max_{x_1} \sum_{(i,j)\in\tilde{g}_k} |\tilde{g}_1|\, x_1 - \left|\tilde{S}_1\right| \frac{1}{2} x_1^2$$

The solution to this maximisation problem is $\tilde{x}_1 = \frac{|\tilde{g}_1|}{|\tilde{S}_1|}$. At the optimum, an agent in $\tilde{S}_2$ is pivotal in along all links with others $\tilde{S}_2$ and links to those in $\tilde{S}_1$,and so applying a similar logic for $\tilde{S}_2$ we get $\tilde{x}_2 = \frac{|\tilde{g}_2|}{|\tilde{S}_2|}$. In general for $\tilde{S}_k$ we have

$$\tilde{x}_k = \frac{|\tilde{g}_k|}{\left|\tilde{S}_k\right|} \tag{1.5}$$

The profile $\tilde{\mathbf{x}}$ therefore gives a candidate maximiser of ρ. However, in order to characterise the partition $\{\tilde{S}_1, \ldots, \tilde{S}_K\}$ I must first introduce another piece of terminology.

In a similar manner to a *k-core*, let $D(k) \subseteq g$ denote the largest subgraph of g such that $\frac{|D(k)|}{|N_{D(k)}|} \geq k$ for $k \in \mathbb{R}^+$.[12] Intuitively, this means that $D(k)$ is the largest subgraph such that the average degree of nodes in that subgraph is at least k. Whereas the definition of a *k-core* placed a restriction on the minimum degree of nodes in a subgraph, we now place a restriction on the average degree of nodes in the subgraph. Furthermore, like the *k-core decomposition* from Section 1.3.1, we can use the set of all such $D(k)$ to construct a *density decomposition* of G, since all such $D(k)$ are nested within subgraphs of lower densities.[13] For each $i \in N$, let δ_i be the largest k such that

[12] Largest in this case means there does not exist an alternative $D'(k) \subseteq g$ such that $D(k) \subset D'(k)$. Note that $D(k)$ is unique, otherwise we could take the union of the two or more subgraphs which satisfy this property.

[13] To see why $k' > k$ implies that $D(k') \subseteq D(k)$, observe that if we had subgraphs $D(k)$ and $D(k')$ where $D(k') \nsubseteq D(k)$ for $k' > k$, then taking the union of these two graphs gives us a graph with average

$i \in N_{D(k)}$ for some $D(k) \subseteq g$ and then define the *density decomposition* of g as follows:

Definition 1.5. A *density decomposition* of a graph g is a partition $\mathcal{D} = \{D_1, \ldots, D_K\}$ of N such that $i \in D_k$ and $j \in D_k$ if and only if $\delta_i = \delta_j$

Using Definition 1.5, I index the subsets of this density decomposition in descending order of their induced subgraph densities. Finding the density decomposition of the network is closely related to a similar problem in the computer science literature as the densest k subgraph problem. This was studied first by Goldberg (1984), who examines the problem of finding the densest subgraph which can be induced using only k nodes. He shows that this can be solved by using a version of the celebrated max-flow min-cut theorem from studies of network flows. My problem is similar, in that I am looking for a hierarchical decomposition of the network in terms of subgraph density.

Using this definition, I now characterise the Nash equilibrium profile $\tilde{\mathbf{x}}$ which maximises the potential function ρ:

Proposition 1.4. *The potential maximising action profile induces a partition* $\{\tilde{S}_1, \ldots, \tilde{S}_K\}$ *of* N *such that* $\tilde{x}_i = \frac{|\tilde{g}_k|}{|\tilde{S}_k|}$ *for each* $i \in \tilde{S}_k$*, where* $\{\tilde{S}_1, \ldots, \tilde{S}_K\}$ *is the density decomposition of* g*.*

Whilst the Pareto dominant equilibrium partitioned nodes into nested subgraphs based on minimum degree, the potential maximising equilibrium partitions nodes into nested subgraphs based on average degree. In the potential maximising equilibrium, agents in the densest subgraph of g will play the highest action, followed by those in the second most dense subgraph, followed by those in the third, and so on. Since costs are incurred at nodes but benefits are received along edges, agents in subgraphs which have a large number of edges spanned by a small number of nodes (i.e. high density) should be expected to play higher equilibrium actions.

An example of a potential maximising equilibrium is shown in Figure 1.6. The densest subgraph of the network displayed in Figure 1.6 is the subgraph formed by nodes in A. This subgraph has a mean internal degree of 3.25, and so their equilibrium effort is $\tilde{x}_A = \frac{13}{8} = 1.625$. The subgraph formed by nodes in $A \cup B$ is the second most

degree that exceeds k, and therefore contradicts the definition of $D(k)$.

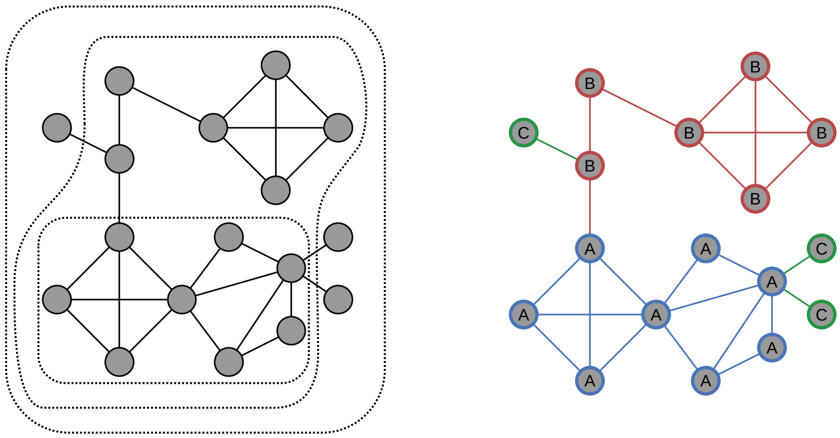


Figure 1.7: Density Decomposition and Potential Maximising Partition

dense in g and so agents in B play $\tilde{x}_B = \frac{9}{6} = 1.5$. Finally, the network as a whole is the third most dense and so agents in C play $\tilde{x}_C = 1$. It is worth noting that agents with the higher corenesses do not necessarily play higher actions in the potential maximising equilibrium. In Figure 1.6 there is a subset of agents in B with coreness 3 who play a lower action than the subset of agents in A with coreness 2.

1.5 Socially Efficient Networks and the Price of Stability

I now consider the features of networks which maximise equilibrium effort and overall welfare. Considering first the case where a network designer can costlessly add edges between nodes, it is worth noting that each utility function $u_i(x_i, \mathbf{x}_{-i})$ displays increasing differences in its own degree d_i.[14] Similarly, the potential function ρ also exhibits increasing differences in the vector of degrees $\mathbf{d}$. A straightforward application of Theorem 2.8.1 from Topkis (1998) shows that if $d_i(g) \geq d_i(g')$ for each i then $\mathbf{x}^*(g) \geqslant \mathbf{x}^*(g')$ and $\tilde{\mathbf{x}}(g) \geqslant \tilde{\mathbf{x}}(g')$. Although it is obvious that the coreness of nodes cannot decrease by adding more links to a network, the impact on the density decomposition is perhaps less clear. The application of Topkis's theorem therefore means that the complete network will always permit the highest action profile in either the Pareto dominant or potential maximising equilibrium.

Defining a utilitarian social welfare function $U(\mathbf{x}) = \sum_{i \in N} u_i(\mathbf{x})$ we can examine the

[14] In other words, fixing the profile of others' actions at $\mathbf{x}_{-i} \in X_{-i}$, $u_i(x_i, \mathbf{x}_{-i}, d_i) - u_i(x_i', \mathbf{x}_{-i}, d_i) \geq u_i(x_i, \mathbf{x}_{-i}, d_i') - u_i(x_i', \mathbf{x}_{-i}, d_i')$ for $x_i > x_i'$ and $d_i > d_i'$

difference in welfare between the social optimum and Pareto dominant Nash equilibrium $\mathbf{x}^*$. As shown in Proposition 1.6 below, in the special case of perfect complements, the Pareto dominant Nash equilibrium also maximises social welfare for a subset of networks.

Proposition 1.5. *The Pareto dominant Nash equilibrium maximises social welfare if and only if the network is a regular graph*

When the network is regular (i.e. $d_i(g) = d_j(g)$ for all i and j) then a network designer can implement the socially optimal level of effort in equilibrium without resorting to transfers of any kind. In order to find the socially optimal profile of action for any network it is possible to directly apply this ideas from Section 1.4.2 since $U(\mathbf{x})$ has an almost identical structure to $\rho(x)$. It is straightforward to verify that the optimal solution will again depend on the density decomposition of G and will result in efforts such that $x_k = 2\frac{|D(k)|}{|N_{D(k)}|}$ for each $k \in \{1, \ldots, K\}$.

A social planner may be interested in how the divergence between Nash and socially optimal outcomes varies across different types of networks. This issue is examined using a concept known as the *price of stability*, which is defined as the ratio of the total utility surplus in the best Nash equilibrium to the total surplus at the social planner's optimum.[15]

As demonstrated in Proposition 1.5, the price of stability $PoS \equiv U(x^*) / \max_{x \in X} U(x)$ is equal to 1 in the case of complete networks and so the decentralisation of decisions to agents leads to no loss of welfare. However, this is only true for regular networks. As is shown in the next proposition, the price of stability is bounded below by $\frac{3}{4}$.

Proposition 1.6. *The price of stability lies in the interval* $(\frac{3}{4}, 1]$.

As shown in the proof of Proposition 1.6, the lower bound of $\frac{3}{4}$ is achieved in the limit for a star network with a very large number of spokes. The price of stability in this game can be viewed as a measure of the redundancy of links in the network. Whilst the total surplus achievable by a social planner depends on the density decomposition of the network (i.e. the average degrees of subgraphs of the network), the surplus achievable

[15] A related concept known as the *price of anarchy* is defined as the ratio of the surplus in the worst equilibrium to that of the social optimum. This is 0 for all networks as $x_i = 0$ for all agents is always stable.

in the best Nash equilibrium depends on the core decomposition (i.e. the minimum degrees of these subgraphs). Thus, when there is a large disparity in the degrees of agents within each of the subgraphs induced by $\mathcal{D}$ then the price of stability is close to $\frac{3}{4}$ since the average degree may be high relative to the minimum degree. On the other hand, when the price of stability is close to 1 this implies that the average degree and the minimum degree of nodes within each of the subgraphs induced by $\mathcal{D}$ is very close.

With reference to the results on equilibrium stability in Section 1.4, although redundant links increase the stability of the best equilibrium in the face of random shocks, they also increase the PoS and so reduce the benefit from decentralising decisions to individual agents. Therefore, a trade off exists from the perspective of a network designer as redundant links may be costly to maintain (both in terms of a link cost and the PoS), yet they bring stability with them.

1.5.1 Socially Efficient Networks with Linear Link Costs

I now consider the case where the maintenance of edges incurs a link cost γ which is linear in the degree of each agent. Assuming that equilibrium $\mathbf{x}^*$ is played, does the addition of link costs imply that networks other than the complete network might be optimal from the perspective of the network designer? Substituting in the Pareto dominant equilibrium actions $x_i^* = c_i$ the problem for the network designer is

$$\max_{g \in G} \sum_{i \in N} \sum_{j \in N_i} \min\left(c_i, c_j\right) - \frac{1}{2} c_i^2 - \gamma d_i$$

Although networks in which agents have high coreness will produce larger surpluses, higher cores will require a larger number of links to construct. The structure of the optimal network will depend on the rate at which the number of links needed increases as we increase the coreness of agents in the graph. However, as Proposition 1.7 shows, this effect does not dominate the benefit received from increasing coreness and so complete networks will be optimal provided the link cost is low enough:

Proposition 1.7. *The optimal network with linear link cost γ is the complete network if $\gamma \leq \frac{1}{2}(n-1)$ and the empty network if $\gamma \geq \frac{1}{2}(n-1)$*

The intuition for this result is that we can ensure that n agents have $c_i = k$ by constructing a k-regular graph using $\lceil \frac{n \cdot k}{2} \rceil$ links. Therefore the marginal cost of increasing the coreness for these n agents is approximately constant in k (ignoring integer problems) and is given by

$$\gamma \left(\lceil \frac{n\,(k+1)}{2} \rceil - \lceil \frac{nk}{2} \rceil \right) = \begin{cases} \gamma \frac{n}{2} & \text{if } n \text{ even} \\ \gamma \lceil \frac{n}{2} \rceil & \text{if } n \text{ odd and } k \text{ even} \\ \gamma \lfloor \frac{n}{2} \rfloor & \text{if } n \text{ odd and } k \text{ odd} \end{cases} \tag{1.6}$$

Since the benefits from raising c_i are increasing in k this indicates why the optimal network must be the complete network.

1.6 Conclusion

This chapter has provided an analysis of the equilibrium properties of games played on networks where agents' actions are perfect complements. I show that well known concepts from the sociology and network science literatures arise naturally out of a simple game of perfect complements played on networks. This chapter can therefore be seen as providing a 'mircofoundation' for the concepts of the *k-core* (Seidman, 1983), (a measure of subgroup importance), and the related concept of *coreness* (a coarse measure of centrality).

The first main contribution of the chapter is therefore a characterisation of actions in the Pareto dominant equilibrium using the notion of *coreness*. However, as I show in the chapter, the Pareto dominant equilibrium can be unstable and prone to unravelling for certain networks. I therefore examine which nodes are more vulnerable to shocks and hence identify where the resources of a social planner can be best directed in order to ensure high actions in equilibrium.

I also analyse which equilibria are most likely to persist, in the long-run, in the face of continual perturbations to actions. Relating this to the concept of stochastic stability from evolutionary game theory, I characterise the potential maximising equilibrium for this game and relate it to a well-studied problem in computer science, known as the ‘densest subgraph’ problem. Whilst the Pareto dominant equilibrium induces a nested decomposition of the network based on the minimum degree of nodes, I show that the potential maximising Nash equilibrium of the game provides a decomposition of the network based on their average degree. Agents who are members of the densest subgraph (in terms of average degree) will play the highest action in this regime.

Although the setting is related to that of a ‘threshold game’, the results presented here differ from those of Morris (2000) and Young (1998), and give new insights into the structural factors which may influence equilibrium decisions. In an exogenously given network, agents who are located in dense but large subgroups will select high levels of investment. Peripheral nodes who inhabit sparsely connected areas of the network will be unable to support high levels of investment even if they have high degrees themselves.

Future work may wish to further consider the role of a social planner in targeting nodes or links to subsidise. For example, a designer could pay a given node to increase their effort, raising the actions of others in equilibrium. Considering nodes who have coreness c_i but are first to be removed in the iterative pruning process used to uncover the $c_i + 1$ core, transfers could be provided to these nodes to increase their efforts and hence sustain a higher equilibrium. Alternatively, a social planner could wish to identify particular links which, if added, would bring the greatest increase to the coreness of agents. Similar questions have been addressed in Bhawalkar et al. (2012) and remain an open area for study.

Appendix - Proofs

Proof of Lemma 1.1. First note that adding an edge cannot lower the coreness of any node in the network. Now suppose that edge (i, j) increased c_i from k to $k+m$ for some $m \geq 2$, this implies that i now has at least $k+m$ neighbours with coreness $k+m$ in the network $g+ij$. However if we remove this newly added edge this would leave i with $k+m-1$ neighbours with at least coreness $k+m-1$, contradicting our assumption that c_i was initially k in g.

To prove for $l \neq i, j$, suppose again that edge (i, j) raised c_l from k to $k+m$ for some $m \geq 2$. Focusing on the subset $\mathcal{K} \subseteq N$ who form the largest k-core in g, we can notice that if either i or j were members of this k-core, then their degrees have only increased by 1 in $g+ij$ and, as established, their coreness increases by at most 1. Since $c_i\left(g+ij\right) \leq k+1$ and $c_j\left(g+ij\right) \leq k+1$, they cannot form part of the new $(k+m)$-core needed to support $c_l\left(g+ij\right) = k+m$. If i and j were not members of $\mathcal{K} \subseteq N$ then $d_\kappa\left(g\right) = d_\kappa\left(g+(i,j)\right)$ for all $\kappa \in \mathcal{K}$ and hence $c_l\left(g\right) = c_l\left(g+ij\right)$.

$\square$

Proof of Proposition 1.1. I first show that $\mathbf{x}^* = \mathbf{c}(g)$ is an equilibrium. Partition the agents into subsets $\{S_1, S_2, \ldots, S_K\}$ based on their coreness such that $c_i = k$ for all $i \in S_k$. Take the set of agents with the largest coreness S_K, since subset are indexed by their coreness, there must exist a connected graph of at least $K+1$ such agents. Since $x_i^* = K \leq d_i'\left(g, x\right)$ for all $i \in S_K$ condition (1.1) is satisfied. We just need to show that (1.2) holds for S_K, but since the subgraph composed of agents playing a strictly higher action is empty this condition is satisfied. Now take S_{K-1} and note again that $x_i^* = K-1 \leq d_i'\left(g, x\right)$ for all $i \in S_{K-1}$ since all agents have coreness $K-1$. Since agents $i \in S_{K-1}$ cannot join a higher core, condition (1.2) must also hold as the number of individuals playing a higher action cannot exceed $K-1$. This reasoning holds for all subsets of lower coreness and so the action profile $\mathbf{x}^* = \mathbf{c}\left(g\right)$ is a Nash equilibrium.

Now to show that this equilibrium is maximal and therefore Pareto dominant, assume that there exists another equilibrium vector of actions $\mathbf{x}'$ such that $\mathbf{x}' \geqslant \mathbf{x}^*$. Take any $i \in N$ playing $x_i' > x_i^* = c_i$ in equilibrium, the best response condition (1.1) implies that i has at least $\lceil x_i' \rceil$ neighbours playing $x_j' \geq x_i'$. Moreover, these agents j must have at least $\lceil x_j' \rceil$ neighbours playing $x_k' \geq x_j'$. Continuing with this reasoning contradicts the assumption that the coreness of node i was $c_i = x_i^* < x_i'$ since we can now construct a subgraph containing i (using these nodes and links only) where each node has at least degree $\lceil x_i' \rceil$ within that subgraph.

The fact that the maximal equilibrium is also Pareto dominant follows from Theorem 7 of Milgrom and Roberts (1990). □

Proof of Proposition 1.2. Since the Bonacich centrality of a node is increasing in the number of paths emanating from it, I focus on the case where all nodes reached on paths from i have exactly degree k. When this is the case I minimise the number of possible paths from i under the constraint that $c_i = k$. If $c_i = k$ then there are at least k^l paths of length l from i to other nodes in the network. Summing paths of all lengths we get $b_i(g, \delta) = \sum_{l=0}^{\infty} \delta^l \sum_{j \in N} g_{ij}^l \geq \sum_{l=0}^{\infty} (\delta k)^l$ and provided $|\delta k| < 1$ this implies that $\sum_{l=0}^{\infty} (\delta k)^l = \frac{1}{1-\delta k}$. To show that this limit is well defined (i.e. $\delta < 1/k$) we rely on an elementary result from spectral graph theory to bound the largest eigenvalue of $\mathbf{G}$.

I assume without loss of generality that the graph is connected since both the coreness and the Bonacich centrality of a node can only depend on structural properties within the same component. Since we assumed that $d_j = k$ for all nodes reachable from i, these nodes form a k-regular graph (with adjacency matrix $\mathbf{G}_k$). We can therefore conclude that the largest eigenvalue of $\mathbf{G}_k$ cannot exceed k, since the largest eigenvalue is always bounded from above by the largest degree of any agent in the network (see Brualdi, 2011) . The condition that $\delta < \frac{1}{|\lambda_{max}(g)|}$ now means that $\frac{1}{1-\delta k}$ is well defined as $\frac{1}{1-\delta k} > 1$ for any δ where b_i is itself well defined. □

Proof of Proposition 1.3. In order to proceed with the proof I will first show that assuming $c_i(g) = c_i^-(\bar{g})$ for all $i \in N$ implies that $\hat{g}$ will be acyclic. Assume to the contrary that a directed cycle $(i \to j \to \cdots \to k \to i)$ in $\hat{g}$ exists. This implies that $c_i(g) \geq c_j(g)$, and similarly for all subsequent nodes in the cycle, implying that $c_i(g) = c_j(g)$. However, if $i \to j$ but $c_i(g) = c_i^-(\bar{g})$ this means that i is linked to a larger number of individuals with coreness at least c_i than agent j. To see why, note that i has $c_i(g) = c_i^-(\bar{g})$ but $(i,j) \notin \bar{g}$ by assumption, so i has at least $c_i + 1$ neighbours of coreness c_i or higher (since j also has coreness c_i). However, since $i \to j$ this means that j has exactly c_i neighbours of coreness c_i since $c_j(g - ij) < c_j(g)$ and the coreness of others cannot be affected by the removal of (i,j) from g due to the assumptions that $c_i(g) = c_i^-(\bar{g})$ for all $i \in N$, and $(i,j) \notin \bar{g}$. Iterating this logic fully along the cycle we reach a contradiction and so $\hat{g}$ must be acyclic.

I now show that $c_i(g) = c_i^-(\bar{g})$ implies stability of $\mathbf{x}^*$. Assume that node i_0 lowers action to 0 at period $t = 0$. Since $c_{i_0}(g) = c_{i_0}^-(\bar{g})$, a sufficient number of neighbours are unaffected by this and permit i_0 to revert back to playing c_{i_0} in period $t = 1$. For nodes $i_1 \in N_{i_0}^-(\hat{g})$, their action must also fall in period $t = 1$ but return to c_{i_1} in period $t = 2$ for the same reason. Since the graph $\hat{g}$ is acyclic, each node can only be affected by a cascading shock a finite number of times and so the assumption that $c_i(g) = c_i^-(\bar{g})$ always ensures that nodes revert to playing c_i the period after they receive a shock to their action. After all nodes revert following their final shock the profile has converged back to $\mathbf{x}^*$.

Finally I show that if $c_i(g) > c_i^-(\bar{g})$ for some $i \in N$ then $\mathbf{x}^*$ cannot be stable. Pick any individual $i \in N$ with $c_i(g) = c_i^-(\bar{g})$ and lower their action to 0 in $\hat{\mathbf{x}}$. In period $t = 1$ i's action returns to c_i but the actions of neighbours in the set $\{j \in N_i \mid i \to j\}$ must fall to $c_j - 1$. In period $t = 2$, i's action again falls to $c_i^-(\bar{g})$, causing the actions of neighbours in the set $\{j \in N_i \mid i \to j\}$ to fall back to $c_j - 1$ in period $t = 3$. The pattern of periods 2 and 3 then cycles and so $\mathbf{x}^t$ does not return to $\mathbf{x}^*$. $\square$

Proof of Lemma 1.2. For (a) we need that $\rho(\mathbf{x}' \vee \mathbf{x}'') + \rho(\mathbf{x}' \wedge \mathbf{x}'') \geq \rho(\mathbf{x}') + \rho(\mathbf{x}'')$ for any $\mathbf{x}'$ and $\mathbf{x}''$. Applying (1.5) we can see that costs immediately cancel on both sides, leaving

$$\sum_{(i,j)\in g} \min\left\{\max\{x_i', x_i''\}, \max\{x_j', x_j''\}\right\} + \min\left\{\min\{x_i', x_i''\}, \min\{x_j', x_j''\}\right\} \geq$$

$$\sum_{(i,j)\in g} \min\left\{x_i', x_j'\right\} + \min\left\{x_i'', x_j''\right\},$$

which clearly holds. Since $-\frac{1}{2}\sum_{i\in N} x_i^2$ is strictly concave in X it suffices to check for each that $\sum_{(i,j)\in g}\ \min\{x_i, x_j\}$ is not convex. As can be easily verified, $\sum_{(i,j)\in g}\ \min\{x_i, x_j\}$ can increase at most linearly with $\mathbf{x}$. □

Proof of Proposition 1.4. Assume that $\{\tilde{S}_1, \ldots, \tilde{S}_K\}$ is the optimal partition of N where $\tilde{x}_i$ satisfies (1.6) for all $i \in N$. It is obvious that $\tilde{x}_k = \frac{|\bar{E}(S_k)|}{|S_k|}$ is a necessary condition for optimality of $\tilde{\mathbf{x}}$ if $\tilde{x}_i = \tilde{x}_j$ for all i and $j \in \tilde{S}_k$. To show that $\{\tilde{S}_1, \ldots, \tilde{S}_K\}$ is the density decomposition of G, I show first that $\tilde{S}_1 = D_1$. If $\tilde{S}_1 = D_1$ then it must be the case that no subset of nodes can optimally play a higher action in equilibrium. Since the optimal action of any $\tilde{S}_1$ cannot be influenced by the actions of other nodes and since D_1 forms the densest subgraph, (1.5) $\implies \tilde{S}_1 = D_1$. To show that $\tilde{S}_2 = D_2$ we can apply the same logic. Taking the actions of $\tilde{S}_1 = D_1$ as given, if $\tilde{S}_2 = D_2$ then no other subset of nodes can play a strictly higher action as a joint best response to D_1, which is again guaranteed by necessary condition (1.5).

Noting that changes in x_i by nodes playing lower effort in equilibrium cannot influence incentives, we can continue this reasoning downwards for all other subsets in the partition to complete the proof. □

Proof of Proposition 1.5. To show the 'if' part for a *d-regular* graph we can note that $x_i = x_j$ for all $i, j \in N$ in the social welfare maximising profile $\mathbf{x} = \text{argmax} \sum_{i\in N} u_i(\mathbf{x})$. This follows from the fact all actions are weak complements and all agents are identical.

The problem then becomes

$$\max_{x} \quad 2\,|g|\,x - \tfrac{|N|}{2}\,x^2$$

First order conditions imply that the social maximise profile of actions satisfies $x_i = \frac{2|E|}{|N|} = d = c_i$. To show the 'only if' part I use a constrained optimisation formulation of the problem:

$$\max_{\mathbf{x}} 2 \sum_{(i,j)\in g} \min\{x_i, x_j\} - \frac{1}{2} \sum_{i\in N} (x_i)^2$$

I reformulate the problem using a dummy variable y_{ij} for $\min\{x_i, x_j\}$:

$$\begin{array}{ll} \max\limits_{\mathbf{x},\mathbf{y}} & 2\sum_{(i,j)\in g} y_{ij} - \frac{1}{2}\sum_{i'\in N} (x_{i'})^2 \\ \text{subject to} & y_{ij} \leq x_i \\ & y_{ij} \leq x_j \end{array}$$

In the above reformulation, at least one of the constraints must bind with equality. The Lagrangian for reformulated problem is:

$$\mathcal{L} = 2 \sum_{(i,j)\in g} y_{ij} - \frac{1}{2} \sum_{i'\in N} (x_{i'})^2 - \sum_{(i,j)\in g} \lambda_{ij} (y_{ij} - x_i) - \sum_{(i,j)\in g} \mu_{ij} (y_{ij} - x_j)$$

The first order conditions with respect to some y_{ij} and some $x_{i'}$ give:

$$\frac{\partial \mathcal{L}}{\partial y_{ij}} = 2 - \lambda_{ij} - \mu_{ij} = 0$$

$$\frac{\partial \mathcal{L}}{\partial x_{i'}} = \sum_{(i,j)\ni i'} \lambda_{ij} - x_i = 0$$

Rearranging and summing these over all edges and N respectively gives $2\,|g| = \sum_{(i,j)\in g} (\lambda_{ij} + \mu_{ij})$ and $\sum_{i'\in N} x_{i'} = \sum_{(i,j)\in g} (\lambda_{ij} + \mu_{ij})$ and so $\sum_i x_i = 2\,|g| = \sum_i d_i$ at any social optimum. If the network is not regular then $c_i < d_i$ for some $i \in N$ and so there is too little effort relative to the social optimum. $\square$

Proof of Proposition 1.6. The proof proceeds in three steps:

First I establish that the price of stability is bounded from below by the ratio of the surplus generated in the potential maximising equilibrium of the game to that of the socially efficient action profile. Using the results of Proposition 1.4 and Proposition 1.5 we get that if $\mathbf{x}$ is the action profile in the socially efficient case, then $\mathbf{x} = 2\tilde{\mathbf{x}}$. Since the surplus in the potential maximising equilibrium cannot exceed the surplus in the Pareto dominant equilibrium we have that

$$PoS\left(g\right) \geq \frac{\sum_{(i,j)\in g} 2 \,\min\left\{\tilde{x}_i, \tilde{x}_j\right\} - \sum_i \frac{1}{2}\tilde{x}_i^2}{\sum_{(i,j)\in g} 4 \,\min\left\{\tilde{x}_i, \tilde{x}_j\right\} - \sum_i 2\tilde{x}_i^2} \equiv \underline{PoS}\left(g\right). \tag{1.7}$$

Since the action in the Pareto dominant equilibrium is weakly higher than $\tilde{\mathbf{x}}$, the price of stability cannot be lower than the expression given in (1.7).

I now show that the price of stability in a star network is equal to $\underline{PoS}(g)$. Since the coreness of all agents in any star network is $c_i = 1$ and since the star as a whole is the densest subgraph of the network, the potential maximising equilibrium action in a star network with 1 hub and M spokes is given by

$$\tilde{x}_i = \frac{M}{M+1}$$

Taking the limit of this as $M \to \infty$ gives us

$$\lim_{M\to\infty} \frac{M}{M+1} = 1 = c_i$$

Therefore, the potential maximising equilibrium profile approaches the Pareto dominant one as M gets large. For large M, the price of stability approaches

$$\lim_{M\to\infty} PoS = \lim_{M\to\infty} \frac{2M - \frac{1}{2}(M+1)}{4M - 2(M+1)} = \lim_{M\to\infty} \frac{3M-1}{4M-2} = \frac{3}{4}$$

Finally, I show that for all other networks, $\underline{PoS}\left(g\right) \geq \frac{3}{4}$. Assume to the contrary that $\underline{PoS}(g) < \frac{3}{4}$ for some g, substituting from (1.7) reduces this to

$$\sum_{(i,j)\in g} \min\{\tilde{x}_i, \tilde{x}_j\} > \sum_{i\in N} \tilde{x}_i^2 \tag{1.8}$$

Therefore, $\underline{PoS}(g) < \frac{3}{4}$ if, in the potential maximising equilibrium for g, the surplus gained along links exceeds twice the sum of the costs accrued at each node.

I restrict attention to complete networks, as if there exists some network g which satisfies (1.8), then the complete network also satisfies (1.8). Since $\sum_{(i,j)\in g} \min\{\tilde{x}_i, \tilde{x}_j\} - \sum_{i\in N} \tilde{x}_i^2$ has increasing differences in $(\mathbf{x}, \mathbf{d})$, then by applying the insights of Topkis (1998), the x which maximises this function is increasing as we add more links, strictly increasing $\sum_{(i,j)\in g} \min\{\tilde{x}_i, \tilde{x}_j\} - \sum_{i\in N} \tilde{x}_i^2$.

In the complete network, actions are symmetric and given by $\tilde{x} = \frac{|g|}{n} = \frac{n(1-n)}{2n}$ which implies that the condition in (1.8) becomes

$$\frac{n(1-n)}{2}\frac{n(1-n)}{2n} > n\left(\frac{n(1-n)}{2n}\right)^2$$

Which after cancelling becomes $\frac{n(1-n)}{2n} > \frac{n(1-n)}{2n}$, a contradiction. □

Proof of Proposition 1.7. To find the optimal network g in the feasible set of graphs G we are given the problem

$$\max_{g\in G} \sum_{i\in N}\sum_{j\in N_i} \min(c_i, c_j) - \frac{1}{2}c_i^2 - \gamma d_i$$

The utility surplus generated by all nodes in the complete network is $n\left((n-1)\left(\frac{1}{2}(n-1)-\gamma\right)\right)$ since $d_i = c_i = n-1$. This surplus is non-negative when $\gamma \le \frac{1}{2}(n-1)$. Now consider a g' which is not complete and an agent i in that network. Agent i has coreness c_i and therefore has at least c_i neighbours with coreness c_i or higher. The maximum possible surplus generated at i in g' is $d_i c_i - \frac{1}{2}c_i^2 - \gamma d_i = c_i\left(d_i - \frac{1}{2}c_i\right) - d_i\gamma$. Since $c_i \le d_i$ in any network this surplus cannot exceed $\frac{1}{2}d_i^2 - d_i\gamma$, which ensures that each node has non-negative surplus only if $\gamma \le \frac{1}{2}d_i$. If $\gamma \le \frac{1}{2}(n-1)$ then the maximum possible surplus at any node $i \in g'$ is $\frac{1}{2}d_i(d_i - \gamma)$, strictly less than

$\frac{1}{2}(n-1)((n-1)-\gamma)$ for each node in the complete network. If on the other hand $\frac{1}{2}(n-1) < \gamma$ then no node can contribute a non-negative surplus in any non-empty network since $d_i \leq (n-1)$ for any $g \in G$. $\square$

Chapter 2

Pre-Launch Publicity and Third Party Reviewers

2.1 Introduction

This chapter examines the role of pre-launch publicity in providing information for buyers regarding the value of a good offered by a seller. For example, firms in the movie industry benefit from publicity campaigns and positive initial reviews from film critics to increase opening week sales. Smartphone manufacturers and software companies selectively allow access to early versions of their product to provide pre-launch product demonstrations and generate 'buzz'. Politicians might rely on favourable interviews or articles in the media in order to enhance their image before selecting a policy platform.

When a firm brings a new product to market they generally have two methods of releasing information about the product prior to launch. One option for the firm is to release information directly to buyers via advertising campaigns, public product demonstrations, press releases, and similar marketing activities. The other option is to release the product to an independent third party who can make public their assessment of the product. These two channels perform different, although complementary, roles in attracting buyers to purchase the product. On the one hand, information released directly by the firm tends to be descriptive in nature, conveying details about the particular features of the product. The seller may choose to put a focus on information which they deem as likely to be viewed favourably by buyers and obfuscate other details in order to maximise the appeal of the product.

The seller may find it difficult, however, to directly make claims about the objective quality of their product, as these claims are generally not seen as credible, since they can be costlessly mimicked by low quality sellers. Evaluations of the quality of the good being sold are therefore generally conducted by independent third parties. Although third-party reviews may also be descriptive in nature, the crucial role which they perform is in evaluating the quality of the product, often giving a purchasing recommendation to the buyer.

In this chapter, I first examine how a seller of a good would optimally design their launch campaign in the face of a trade-off between generating favourable posterior beliefs and maximising the probability that the campaign is a success. I assume that the seller has full flexibility over the design of the information structure and draw on the insights of the recent literature on 'Bayesian persuasion' (Kamenica and Gentzkow (2011)).

The seller's optimal campaign will take one of three possible forms: a *mass market* campaign which (if successful) targets all buyers, a *niche* campaign which targets only high-type buyers, and a *segmented* campaign which is a hybrid of the two. I examine when these campaigns are appropriate and how the seller benefits by withholding information from buyers in each case.

Buyers in this model have private information about their likely valuation for the product. The pre-launch publicity campaigns which sellers optimally choose will depend on the accuracy of this private information. In particular, when buyers have weak private information about the product's likely value, this induces sellers to design campaigns which are more revealing. In contrast to much of the existing literature, the seller never wishes to subject themselves to 'all or nothing' tests where information is either completely obscured or perfectly revealed.

In this chapter I also examine how the differing standards of product reviewers can affect the amount of pre-launch publicity carried out by the seller. Reviewers can provide a valuable service to buyers (and sellers), but, as will be shown, reviewers who are excessively soft will limit the ability of firms to manipulate buyers' beliefs via their launch campaign. For this reason, sellers who perfectly observe their quality as high may benefit from picking reviewers who give a less useful signal of product quality.

My model therefore builds on the work of Gill and Sgroi (2012) who analyse the optimal choice of pre-launch reviewer. Perhaps counter-intuitively, there can be situations where buyers prefer overly soft reviewers, whereas sellers prefer overly harsh ones. Softer reviewers will force the seller to reveal more information about the product pre-launch, benefiting buyers, yet harsh reviewers can be more attractive to the seller since they allow them to separate from lower quality sellers.

The remainder of the chapter is structured as follows: First I explain how this work relates to the previous literature before discussing the application of the Bayesian persuasion framework of Kamenica and Gentzkow (2011) to the case of multiple receivers with private information, utilising the recent insights provided by Alonso and Câmara (2014). I then provide an exposition of the main results on the optimal choice of launch campaign for the seller. Following this I look at a simple model of third party reviewers and extend the model to account for endogenous selection of a pre-launch reviewer. I end the chapter with a discussion of the main findings of the model in the context of the other results in the literature.

2.2 Related Literature

This chapter is related to the emerging literature on Bayesian persuasion and information control. Kamenica and Gentzkow (2011) study the design of optimal signal structures which persuade the receiver of the signal to take actions which will benefit the sender. They provide conditions under which a sender will benefit from persuasion and provide a tractable framework which has opened up the study of optimal signal design. The authors have also extended their model to multiple senders who compete in providing information for a receiver (Gentzkow and Kamenica, 2012) and to the case of costly signals (Gentzkow and Kamenica, 2014). Another recent paper by Alonso and Câmara (2014) examines the case of information control where the sender and receiver have differing prior beliefs, extending the original work of Kamenica and Gentzkow (2011) to a broader class of problems.

Although the literature on Bayesian persuasion is still in its nascent stages, there have been a sequence of important contributions on persuasion and information revelation by monopolists which have preceded this literature. Work on 'persuasion games' was initiated by Milgrom (1981) and Milgrom and Roberts (1986) who examine the release of verifiable reports about the quality of a good provided by a seller with private information. In this setting a seller must make a verifiable statement about the quality of the form "the product is of quality at least q", which buyers must then choose how to interpret, before making a purchasing decision. The central conclusion of these papers is that scepticism by buyers ensures that the payoff from full information revelation cannot be improved upon by attempting to misrepresent the product. These early contributions therefore suggest that sellers may have little to gain from this form of persuasion.

Lewis and Sappington (1994) were the first to specifically address how a monopolist might design an information structure in order to extract value from consumers. In their paper, the monopolist has control over how much consumers can learn about their tastes for the product by varying the noise of a signal which the firm transmits. Following the design of the signal structure by the monopolist, buyers observe their individual signals and decide whether or not to purchase. In the two focal results of the paper, they show that the monopolist will prefer extreme information structures where consumers are provided with the most informative signal or no signal at all.

Another related contribution is that of Ottaviani and Prat (2001). Their model builds on the insights of Lewis and Sappington (1994) and those of Milgrom and Weber (1982) to the case where an uninformed monopolist can commit to publicly revealing information about product quality. Their setting is more general than Lewis and Sappington (1994) as they assume weaker conditions on the distribution of the signals, yet they focus solely on the release of public information. Ottaviani and Prat (2001) show that the monopolist always benefits from the release of public information which is affiliated with buyer's private valuations for the product. Moreover, the monopolist will prefer to have more informative public signals, and so signals which perfectly reveal product quality are superior to partial revelation. This paper again suggests that the value of information control for sellers may be limited.

A final key contribution to this strand of the literature is Johnson and Myatt (2006) who develop a theoretical framework for the analysis of transformations of demand curves. This framework is then applied to the study of advertising and the release of product information by a monopolist. They note that the release of product information causes a dispersion of buyer valuations in expectation and so causes the demand curve to rotate. The authors go on to show that extreme information structures are optimal (i.e those which reveal no information or allow maximal dispersion of private beliefs) echoing the previously mentioned contributions. Taken as a whole, this strand of the literature seems to suggest an 'all or nothing' position for the seller, such that they will either wish to publicly disclose the maximal amount of information or nothing at all.

Another branch of the literature which is relevant to this chapter is the research that has been carried out on the endogenous selection of reviewers by monopolists. The most closely related paper in this branch is that of Gill and Sgroi (2012), who consider the optimal choice of pre-launch review in the case when firms can condition their pricing decision on the outcome of a public test. They show in their model that when the seller perfectly observes their quality, the high quality seller will always wish to select the toughest or softest possible reviewer, where 'toughness' is defined as the propensity to fail high quality products. The low type seller always pools on this decision and so the optimal choice of pre-launch reviewer for the firm is at one of the two extremes.

As pointed out by Gill and Sgroi (2012), the literature has, until recently, largely ignored the role of selection of pre-launch reviewers by sellers, when reviewers vary in their toughness. An exception to this is Lerner and Tirole (2006) who examine the joint decision of product design and certifier for a monopolist who must submit their product for certification by a third party such as a standard setting organisation. They find that firms generally prefer a weak certifier who have standards which are just low enough to endorse the product, in order to minimise the concessions made to the certifier. Although, they also show that there is an inverse relationship between the credibility of the chosen certifier and the pre-test belief about product quality. This chapter will build on both Gill and Sgroi (2012) and Lerner and Tirole (2006) by examining the choice of a publicly observable pre-launch reviewer in the case when the seller has some private information about the quality of the good for sale.

Another paper which looks at similar phenomena is Hvide (2009), who analyses the case of competition between testers of differing standards who charge fees for certification. The analysis in his model shows that by setting fees for certification, testers of differing standards can segment the market. He shows that high quality sellers (who desire tough tests to separate from low quality sellers) will be willing to pay larger fees to tested by tougher certifiers. His model differs in focus from mine as it looks at reviewers with exogenous toughness setting prices in order to extract surplus from sellers.

Other models such as Gill and Sgroi (2008) and Sgroi (2002) have considered how a monopolist may be able to influence early opinions about a product via pre-launch tests. In particular, Gill and Sgroi (2008) analyse how the choice of pre-launch test can be used to manipulate the process of observational learning which occurs when a product of unknown quality is released. The favourable opinions which are induced following the passing of a tough test are of value to the seller, as they maximise the possibility of an informational cascade on buying. The authors show that a seller will optimally select reviewers who deliver beliefs which are just enough to induce informational cascades without being so tough as to significantly lower the probability of passing the test.

To summarise this area of the literature, the work on the optimal choice of pre-launch reviewer/certifier is still underdeveloped and the key insights are unclear. On the one hand, the most extensive treatment of this issue by Gill and Sgroi (2012) highlights the attractiveness of reviewers who are either very tough or very soft, yet other papers have shown that found intermediate reviewers to be optimal. A key difference between this paper and Gill and Sgroi (2012) is that the latter permit the monopolist to condition the price of the product on the result of the signal. In my case, the signals are privately observed by buyers and so cannot be conditioned upon, as I will now discuss.

2.3 The Optimal Launch Campaign

The model described in this section will consider a seller who is introducing a new product to a unit mass of buyers that must choose to either accept or reject a seller's product offer. The seller produces their good at zero cost and is randomly drawn a product quality $q \in \{0, 1\}$ by Nature, which is then privately observed by the seller. The buyers and the seller share identical prior beliefs about the product quality and assign probability $\lambda \in (0, 1)$ to $q = 1$. Each buyer i will also receive an informative private signal about their *match type* $\theta_i \in \{0, 1\}$, where the buyer has prior belief $\mu \in (0, 1)$ that $\theta = 1$ before private signals are received. This is common knowledge and so the seller expects a fraction μ of the population to have match type $\theta = 1$. If buyer i accepts the product offer by taking action $a_i = 1$ then they receive utility $u(a_i) = \theta_i q$, while if they reject the offer by taking action $a_i = 0$ they receive their outside option which gives utility $\underline{u} < 1$. I focus on pure strategy equilibria throughout and assume that the buyer accepts the seller's offer if $\mathrm{E}[u_i] = \underline{u}$.

The assumption that the seller's offer must meet a reservation utility, as opposed to allowing sellers to set a price, is appropriate in some circumstances and less appropriate in others. In the running example of a movie production company who must design a publicity campaign this assumption fits well, as generally the price of admission is fixed and buyers must select between competing movies. The case where a monopolist sets their price conditional on the result of a public test is examined in Gill and Sgroi (2012). In my case, the only tool which the seller has at their disposal in order to increase payoff following the review is to engage in persuasive marketing of their product.

The payoff for the seller is given by the total fraction of buyers who accept the product offer and is denoted by π. I assume that each buyer i learns about their match type via a noisy private signal $\sigma_i \in \{L, H\}$ with accuracy $\alpha \in (0.5, 1)$, such that $\Pr(\sigma_i = H \mid \theta_i = 1) = \Pr(\sigma_i = L \mid \theta_i = 0) = \alpha$ and $\Pr(\sigma_i = H \mid \theta_i = 0) = \Pr(\sigma_i = L \mid \theta_i = 1) = 1 - \alpha$.[1] This results in updated beliefs μ^L and μ^H where $\alpha \in (0.5, 1)$ ensures

[1] I will follow the convention of distinguishing between buyers by the private information which they hold. I shall therefore refer to σ_i as a buyer's 'type', which is known to buyer, as opposed to their unknown match type θ_i.

that $\mu^L < \mu^H < 1$.

In addition to the private signal about their match type, buyers will be able to learn about their expected valuation via a *launch campaign* (conducted by the seller) and a product review (conducted by a third party reviewer). The role of the seller's launch campaign is to manipulate the buyer's beliefs about their match type through the release of information about the product. The launch campaign can be thought of as generating a signal for each buyer as a result of the seller's advertising campaign, promotional work (such as interviews with industry magazines) or public product demonstrations. I model the combination of these activities as a signal generating mechanism (a set of possible signal realisations and a collection of conditional probabilities over these realisations) which induces a distribution over the possible posterior beliefs of the buyers.

This form of information revelation differs from that considered in Ottaviani and Prat (2001). My approach has two distinguishing features: firstly, each buyer receives their own private signal which is independent of other signals; and secondly, some buyer types are more likely to receive certain signals than other buyers. In that respect, although the seller only designs a single publicly observed launch campaign, it is best viewed as a 'signal generating machine' which can output possibly different signals to each buyer.

I assume throughout that the seller's product is an experience good and so the quality can only be gauged following consumption. Due to this fact, the seller's launch campaign only allows buyers to gain information about their likely match type, even though the seller also has private information about product quality. Although the seller may benefit from the release of this private information, it is assumed that they have no way of credibly doing so without being mimicked by low quality sellers. Hence quality certification must be conducted by an independent third-party reviewer.

The third-party reviewer subjects the product to a publicly observable and costlessly verifiable test, the result of which conveys information about the product quality q.[2] As in Gill and Sgroi (2012), the test which the product is subjected to can vary in

[2] In an alternative interpretation of the model, the review could also release information about the match type in the form of the private signal which the buyers receive.

its toughness, where tough tests are those which are more difficult to pass but deliver strong posterior beliefs in the event of a pass. Soft tests are easier to pass and so a product which fails a soft test is more likely to be bad quality than a product which fails a tough test. I assume that the result of the reviews are announced pre-launch, so the seller's decision about the launch campaign can be conditioned on the result of the review. For now I will assume that the toughness of the reviewer's test is exogenously given, but examine this in more detail in Section 2.4.

This description of the setting fits the distinction provided by Johnson and Myatt (2006) between *hype* and *real information* as being different forms of advertising for the firm. In my case, reviews provide *hype* for the product, which corresponds to advertising which informs consumer about any unambiguously desirable property. The seller provides *real information* which allows consumers to learn about their subjective preferences.

As an example of where this model might apply, consider a movie production company who are launching a new film and must decide on the optimal amount of information (e.g. about the setting, plot, and so on) to release to audiences following the results of early reviews. Although I will generally refer to the two parties as the seller and the buyers, this setting also fits other applications. For example, the seller of the good could be a politician seeking endorsement by a group of voters where θ represents the ideological position of the politician, and q their 'valence' or overall competence. In this case, the publicly observable pre-launch review could represent a televised interview or debate, whereas the launch campaign could represent campaign speeches made by the politician.

The decision problem for the seller is one of signal design and so I build on the framework of Kamenica and Gentzkow (2011) (henceforth KG) to allow the seller to design their launch campaign under minimal restrictions on the structure of the signal itself. Since buyers may have differing beliefs about the match type pre-launch, the extension of this framework by Alonso and Câmara (2014) (henceforth AC) to the case of heterogeneous priors allows me to extend the standard framework to the case of

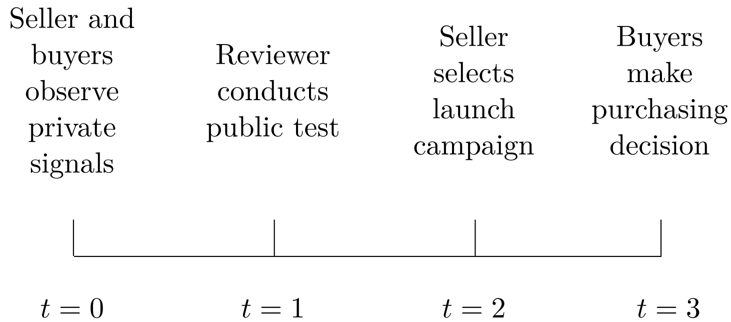


Figure 2.1: Timing of the Model

multiple receivers with private information.[3]

I summarise the timing of the model in Figure 2.1 before discussing the optimal launch campaign in more detail.

2.3.1 Information Control with Multiple Receivers and Private Signals

For the remainder of this section I shall fix the behaviour of the third-party reviewer and return to this aspect of the model in Section 2.4. I therefore assume that the result of the review has been released and has endowed the seller and buyers with shared belief λ about product quality. The task for the seller is to manipulate the buyers' beliefs about their match type, in order to maximise the number of buyers who purchase the product. In principle, the seller can release all the relevant information about the product at launch, so in the absence of information control, the buyers receive a signal which allows them to perfectly identify their match type. From this perspective, we should view the information control decision as being concerned with how much we garble the information about the details of the product.

As in the standard Bayesian persuasion model, the seller manipulates the beliefs of buyers by specifying how the signals which buyers get are generated. This involves defining a finite set of signal outcomes S and a collection of conditional probabilities $\{\Pr(\cdot \mid \theta)\}_{\theta\in\{0,1\}}$ over S. This set S and collection of probabilities $\{\Pr(\cdot \mid \theta)\}_{\theta\in\{0,1\}}$ shall be referred to as the seller's *launch campaign*. The launch campaign is therefore

[3]This set-up allows more flexibility over the possible posterior beliefs of the buyers when compared to the approach of Lewis and Sappington (1994), who restrict attention to specific signal structures.

not a signal itself but a signal generating mechanism which outputs an informative private signal for each individual buyer.[4]

In the simplified setting of KG, a seller and a single buyer would share a common prior belief $\mu \in (0,1)$ about the match type of the buyer, which is then updated following the signal realisation $s \in S$.[5] By specifying the conditional probabilities $\{\Pr(\cdot \mid \theta)\}$, each launch campaign would imply a marginal distribution over S and hence induce a distribution τ over the possible posterior beliefs μ_s of the buyer. Since the action of the buyer will depend on these posterior beliefs, the seller must optimally design the signal environment in order to induce posterior beliefs (and a distribution τ over these beliefs) which maximises their expected payoff.[6]

An important contribution of KG is to reduce the complexity of this problem to one of solely selecting the posterior beliefs which the seller wishes to induce. To ensure that any chosen posterior beliefs are consistent with Bayesian rationality, Proposition 1 of KG provides a characterisation of the admissible posterior beliefs and distributions τ over these beliefs. They show that if the sets of states and signal realisations are finite, then the martingale property $\mathrm{E}_\tau[\mu_s] = \mu$ ('Bayesian rationality' in KG) is a necessary and sufficient condition for the existence of conditional probabilities $\{\Pr(\cdot \mid \theta)\}$ which generate those beliefs according to Bayes' rule. This allows us to focus directly on the posterior beliefs which are induced by the signal without concerning ourselves with the details of the signal which generates them.

Applying this framework to the design of an optimal launch campaign, the seller commits to S and $\{\Pr(\cdot \mid \theta)\}$, which then generates a signal s_i for each buyer i. Signals are independent across buyers and I assume that each buyer's signal realisation is private information. Although the details of the launch campaign are common knowledge, buyers can have different pre-launch private signals and therefore different beliefs about

[4]The specification of the launch campaign is common knowledge.

[5]In my model there is a continuum of buyers, however this presents no problem since the assumption of a continuum of buyers performs essentially the same function as the 'concavification' of the payoff function which is used in KG.

[6]This is similar to the setting of Bergemann and Pesendorfer (2007) who examine how an auctioneer might design an information structure for buyers prior to auction. The main difference is that the Bayesian persuasion approach does not allow the seller to tailor the structure of each signal to individual buyer.

their match type θ_i prior to launch. The posterior belief about θ_i from a buyer who sees signal s_i would differ from a buyer j who received $s_j = s_i$, but received private signal $\sigma_j \neq \sigma_i$ pre-launch.

This creates a potential problem in directly choosing posteriors for the buyers subject to $\mathrm{E}_\tau [\mu_s] = \mu$, as it is no longer clear which μ and τ should be used. Furthermore, the existence of multiple different posterior beliefs for any given launch campaign also appears to increase the dimensionality of the problem, making it potentially less tractable. These issues are addressed by Alonso and Câmara (2014), who extend the framework of KG to the case of heterogeneous priors. They show that it is possible to reformulate the problem in terms of selecting the posterior beliefs of one party and then use a bijective function to map these into the resulting posterior beliefs of the others. Selecting Bayes rational posterior beliefs from the perspective of one party therefore selects Bayes rational (although different) posterior beliefs from the perspective of the other party.

Since the buyer's action space and the state space are binary, the possible signal realisations can be restricted to be either a 'good' signal $g \in S$ or 'bad' signal $b \in S$ where $\mu_g \geq \mu_b$.[7] The Bayesian rationality condition implies that by restricting attention to two signal realisations, we have that $\mu_g \geq \mu$ and $\mu_b \leq \mu$.

Taking advantage of the insight of Alonso and Câmara (2014), the task for the seller is therefore to select posterior beliefs μ_g and μ_b for the marginal buyer (i.e. averaged across both buyer types), subject to $\tau\mu_g + (1-\tau)\mu_b = \mu$, in order to maximise $\mathrm{E}[\pi]$.[8]

This choice of posteriors also results in a choice of μ_g^σ and μ_b^σ for $\sigma \in \{H, L\}$ using (2.1), which is found by applying Proposition 1 of Alonso and Câmara (2014):

$$\mu_s^\sigma = \frac{\mu_s \frac{\mu^\sigma}{\mu}}{\mu_s \frac{\mu^\sigma}{\mu} + (1-\mu_s)\frac{(1-\mu^\sigma)}{(1-\mu)}} \tag{2.1}$$

For any launch campaign such that $S = \{g, b\}$, we can use the Bayesian rationality condition $\tau^\sigma \mu_g^\sigma + (1-\tau^\sigma)\mu_b^\sigma = \mu^\sigma$ to back out $\tau^\sigma \in [0,1]$. This gives the probability

[7] See Proposition 2 of Alonso and Câmara (2014).

[8] In fact, since the function mapping the seller's posteriors to those of either buyer type is a bijection, we could equivalently select posteriors for either buyer type which are consistent from their point of view (i.e. satisfy Bayesian rationality with respect to their prior beliefs μ^H or μ^L) and then transform back in to corresponding beliefs for the seller using (2.1).

that the launch campaign will be successful in generating a 'good' signal for a buyer who has private pre-launch belief μ^σ as

$$\tau^\sigma = \frac{\mu^\sigma - \mu_b^\sigma}{\mu_g^\sigma - \mu_b^\sigma}$$

Applying a similar logic for the seller with marginal belief μ, the probability of a successful launch campaign is therefore

$$\tau = \frac{\mu - \mu_b}{\mu_g - \mu_b} \tag{2.2}$$

Examining (2.2) reveals an important feature of the seller's design problem, namely that a campaign is more likely to deliver a good signal when the belief induced following a good signal is lower.[9] Similarly, the higher the posterior belief following a bad signal, the less likely it is that a good signal will be drawn. When selecting the optimal launch campaign the seller will face a clear trade off between maximising the probability that a favourable signal is drawn and generating beliefs which are sufficient to encourage the buyers to purchase.

A second important feature about the launch campaign is also worth highlighting at this point. Not only do high type buyers find themselves with higher posterior beliefs following the same launch signal as a low type consumer, but they also perceive the launch campaign as *a priori* more likely to be successful. In other words, the same launch campaign with fixed conditional probabilities $\{\Pr(\cdot \mid \theta)\}$ is more likely to be persuasive for buyers who already have higher prior beliefs. I now summarise the structure of the private signals and launch campaign in Figure 2.2 before moving on to discuss the seller's payoff from any launch campaign.

The Seller's Payoff Function

The extension of KG to the case of heterogeneous priors allows us to express the seller's payoff function solely in terms of the posterior (post-launch) beliefs of a reference player.

[9] Although the ex-ante probability of a good signal being drawn will differ across types, increasing μ_s^σ for type σ will also do so for the seller, meaning that τ^σ and τ move in the same direction following a change of μ_s^σ.

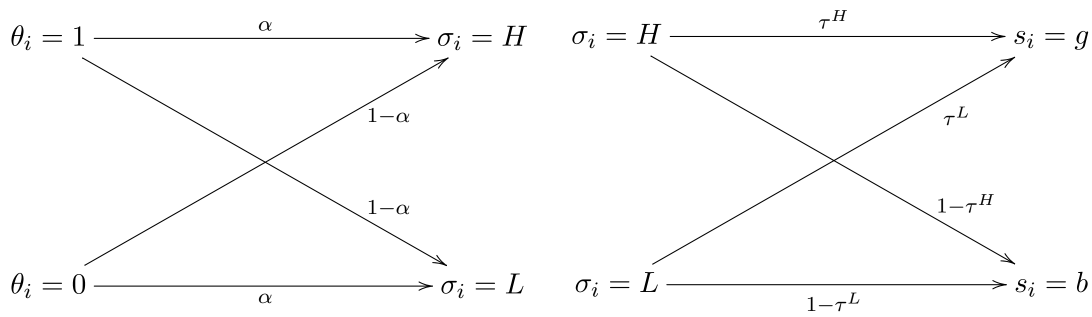


Figure 2.2: Summary of Private Signals and Launch Campaign

To illustrate how this helps, assume momentarily that each launch signal is observable by the seller. Prior to launch, each buyer has private signal σ_i, which is informative about their match type. Although this private signal is not observable by the seller, we can let μ_s denote the belief of the seller about buyer i's match type if they were to observe launch signal s being drawn for buyer i. I shall therefore refer to μ_s as the *reference belief*, which the seller will be selecting when designing the optimal signal. Whether the seller actually observes the launch signal or not is of little consequence, as this only aids us in formulating the seller's design problem.

Since the private signal has accuracy α, the pre-launch beliefs of the high and low type buyers are given by

$$\mu^H = \frac{\alpha\mu}{\alpha\mu + (1-\alpha)(1-\mu)} \qquad \mu^L = \frac{(1-\alpha)\mu}{(1-\alpha)\mu + \alpha(1-\mu)}$$

Substituting this in to (2.1) and simplifying gives the following beliefs μ_s^H and μ_s^L after the launch campaign signal s:

$$\mu_s^H = \frac{\alpha\mu_s}{\alpha\mu_s + (1-\alpha)(1-\mu_s)} \qquad \mu_s^L = \frac{(1-\alpha)\mu_s}{(1-\alpha)\mu_s + \alpha(1-\mu_s)} \tag{2.3}$$

These probabilities are the posterior beliefs held by each by type, written as a function of the (seller's) reference belief μ_s. We can also interpret the updated beliefs in (2.3) by changing the order in which private signals are received, so that μ_s is the

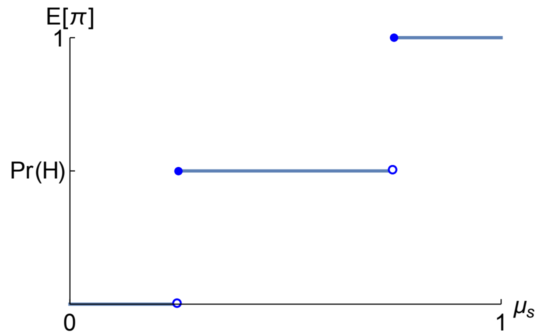


Figure 2.3: Expected Buyers as a Function of the Posterior Belief μ_s

post launch belief for all buyer types, who then receive an additional private signal $\sigma_i \in \{H, L\}$.

Having updated their beliefs to μ_s^σ, the buyer would gain expected surplus $\mu_s^\sigma \lambda$ from accepting the seller's product when the public belief about product quality is λ. Therefore a buyer of type σ takes action $a_i = 1$ if $\mu_s^\sigma \lambda \geq \underline{u}$ and action $a_i = 0$ otherwise. By defining $\mathbf{1}_{\underline{u}}(x)$ as an indicator function which is 1 when $x \geq \underline{u}$ and 0 otherwise, we can more easily express the seller's expected payoff. If the launch campaign were designed such that it always generated belief μ_s then the seller's expected fraction of buyers is given by

$$\mathrm{E}[\pi] = \Pr(H)\,\mathbf{1}_{\underline{u}}\left(\frac{\alpha\mu_s}{\alpha\mu_s + (1-\alpha)(1-\mu_s)}\lambda\right) + (1-\Pr(H))\mathbf{1}_{\underline{u}}\left(\frac{(1-\alpha)\mu_s}{(1-\alpha)\mu_s + \alpha(1-\mu_s)}\lambda\right)$$

By focusing on pure strategy seller preferred equilibria (i.e. if buyers are indifferent then they choose to purchase), $\mathrm{E}[\pi]$ is upper semicontinuous in μ_s. Substituting (2.3) in to the above, the expected payoff from any launch campaign is therefore

$$\Pr(H)\left[\tau^H \mathbf{1}_{\underline{u}}\left(\mu_g^H \lambda\right) + (1-\tau^H)\mathbf{1}_{\underline{u}}\left(\mu_b^H \lambda\right)\right] + \Pr(L)\left[\tau^L \mathbf{1}_{\underline{u}}\left(\mu_g^L \lambda\right) + (1-\tau^L)\mathbf{1}_{\underline{u}}\left(\mu_b^L \lambda\right)\right]$$

As shown by Proposition 2 of Alonso and Câmara (2014), an optimal signal must exist by Berge's maximum theorem.

2.3.2 The Seller's Choice of Launch Campaign

I now consider the choice of launch campaign by the seller, focusing on high type seller preferred perfect Bayesian equilibria. Although the seller has private information about their quality $q \in \{0, 1\}$, the mechanism through which information about quality is revealed by the reviewer is currently exogenous. Since the choice of launch campaign is conditioned on the result of the review, the low type seller will always wish to replicate the launch campaign of the high type seller, for any public belief λ which is generated.[10]

An assumption which is made about the launch campaign is that it is costless to provide. If it is costly for the seller to engage in promotion of their product, then this opens up the possibility of signalling when the seller has private information about the quality of the good.[11] I wish to focus purely on the design of the launch campaign itself: who it targets, which features of the product it seeks to mask, and how the behaviour of external reviewers influences this decision. For this reason, I maintain the assumption that the launch campaign is costless throughout.

Buyers will purchase the product if their private belief about their match type following the launch campaign meets some threshold belief which is determined by λ and $\underline{u}$. From the buyers' perspective, these correspond to $\mu_s^H \lambda = \underline{u}$ and $\mu_s^L \lambda = \underline{u}$ for the high and low types respectively. I denote by $\bar{\mu}^H$ and $\bar{\mu}^L$ these threshold beliefs from the perspective of the seller (i.e. in terms of the marginal reference belief). Using (2.3) we can see that these are given by

$$\bar{\mu}^H = \frac{(1-\alpha)(\underline{u}/\lambda)}{(1-\alpha)(\underline{u}/\lambda) + \alpha(1-\underline{u}/\lambda)} \qquad \bar{\mu}^L = \frac{\alpha(\underline{u}/\lambda)}{\alpha(\underline{u}/\lambda) + (1-\alpha)(1-\underline{u}/\lambda)}$$

Or equivalently as

$$\bar{\mu}^H = \frac{(1-\alpha)\underline{u}}{(1-\alpha)\underline{u} + \alpha(\lambda - \underline{u})} \qquad \bar{\mu}^L = \frac{\alpha\underline{u}}{\alpha\underline{u} + (1-\alpha)(\lambda - \underline{u})} \tag{2.4}$$

As seen from (2.4), whenever $\underline{u} < \lambda$, the threshold beliefs are such that $\bar{\mu}^H < \bar{\mu}^L < 1$.

[10]In order to restrict attention to high type preferred equilibria I assume throughout that off path beliefs are such that buyers assume that any deviation from the high type's optimal choice is believed to be a deviation by the low type. I discuss this choice of refinement later in Section 2.4.1.

[11]The signalling impact of advertising has been analysed elsewhere and is surveyed by Bagwell (2007).

As highlighted previously, a launch campaign which induces stronger beliefs following a good signal must produce good signals less frequently in order to satisfy Bayesian rationality. This is the sense in which it will be easier to persuade high types to purchase than it will be for low types, as the reference belief needed to encourage them to buy is lower, and so a launch campaign which generates this signal will do so more often.

Moreover, it is for this reason that a fully informative launch campaign which reveals all product information (i.e. $\mu_g = 1$ and $\mu_b = 0$) may not be optimal, as it could reduce the probability that the campaign succeeds. As I show below, a fully informative campaign is only optimal in one special case:

Lemma 2.1. *The seller strictly benefits from information control if and only if* $\underline{u} < \lambda$

Corollary 2.1. *Full revelation is only optimal only if* $\underline{u} = \lambda$

If $\underline{u} > \lambda$ there is no launch campaign which can persuade buyers to purchase, so information control does not benefit the seller. Whenever $\underline{u} < \lambda$ then the seller will always benefit from manipulating how much the buyers can learn about the product.

The only case where full revelation is optimal is when $\underline{u} = \lambda$. In this case, buyers will only purchase if they are certain they are match type $\theta = 1$, therefore a fully revealing campaign is the seller's only option to persuade them to purchase. Ignoring this case, the seller never wishes to allow buyers to discover their match type with certainty. In particular, since the seller never wishes to induce posterior $\mu_g = 1$, this implies that they find it desirable to have some agents with match type $\theta_i = 0$ receive a good signal.

Another point to note about the structure of the optimal launch campaign is that the seller wishes to conduct an informative campaign if and only if $\lambda\mu^L < \underline{u}$. Clearly, if λ is such that $\underline{u} \leq \mu^L\lambda < \mu^H\lambda$ then all types buy and an informative launch signal cannot increase the payoff of the seller. If $\mu^L\lambda < \underline{u} \leq \mu^H\lambda$ then the seller can benefit from a launch campaign such that $\mu_b^H\lambda = \underline{u}$ and $\mu_g^L\lambda = \underline{u}$ as all high types will still purchase, and the low types now have a non-zero probability of purchasing. Similarly, for $\mu^L\lambda < \mu^H\lambda \leq \underline{u}$, neither type purchases and so a campaign which has a non-zero probability of persuading either type is always beneficial.

The seller will therefore only wish to provide extra information to buyers if the public belief about quality is sufficiently pessimistic, as the result of especially negative or uninformative reviews. However, if the public belief about quality is too low following the review then it may be impossible to persuade either of the two types to buy as $\underline{u} > \lambda$ and consequently $\bar{\mu}^L > 1$ and $\bar{\mu}^H > 1$. I will return to these cases later in the chapter but for now I assume that the review has endowed buyers with belief λ such that $\underline{u} < \lambda$ and information control is beneficial. Ignoring the above mentioned case where it is optimal for the seller to provide a completely uninformative launch campaign (i.e. $\underline{u} \leq \mu^L\lambda < \mu^H\lambda$), I now examine the two main cases where it is beneficial for the seller to commit to an informative launch campaign:

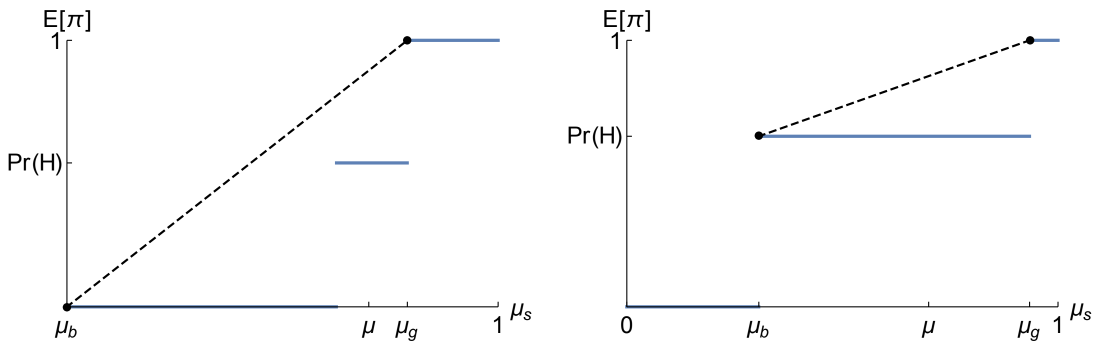
Case 1: $\mu^L\lambda < \underline{u} < \mu^H\lambda$

In this case, low type buyers will not endorse the product unless the seller conducts a persuasive launch campaign. Here the optimal launch campaign takes one of two possible forms. The belief induced following a successful launch must always be such that the low types purchase - if this were not the case (e.g. a good launch signal only persuaded the high type to purchase) then the seller gains nothing from the launch campaign. The optimal campaign is therefore such that a good signal encourages the low types to buy.

Bayesian rationality ensures that the seller faces a trade-off when picking the posterior belief μ_b. On the one hand, the seller wishes to select μ_b such that all high types will still purchase after a bad signal, yet lowering μ_b also increases the probability that a good signal will be drawn. I show below that the seller will prefer to pick an information structure which yields more extreme posterior beliefs when the private information of the buyers is less accurate.

Proposition 2.1. *If $\mu^L\lambda < \underline{u} < \mu^H\lambda$ then there exists an $\bar{\alpha}$ such that the optimal launch campaign satisfies:*

- $\mu_g^L\lambda = \underline{u}$ and $\mu_b\lambda = 0$ if $\alpha \leq \bar{\alpha}$
- $\mu_g^L\lambda = \underline{u}$ and $\mu_b^H\lambda = \underline{u}$ if $\alpha > \bar{\alpha}$



(a) $\alpha = 0.6$

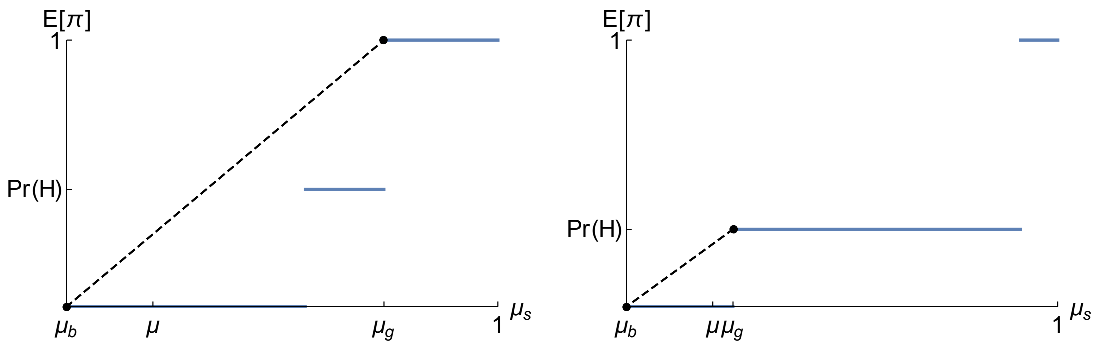
(b) $\alpha = 0.85$

Figure 2.4: Optimal Launch Campaign with $\lambda = 0.7$, $\mu = 0.7$

This result demonstrates that in this case two types of launch campaign can be desired by the seller. In both campaigns $\mu_g^L \lambda = \underline{u}$ is satisfied, meaning that all types buy following a good signal. The optimal belief which is induced after a bad signal will either be just enough to persuade the high types to purchase, or it will deliver a posterior belief of $\mu_b = 0$, maximising the chance of a good signal.

When the private information of buyers is more accurate this has two effects on the design of the signal. Firstly, a larger α means that high type buyers have a strong pre-launch belief that they are match type $\theta = 1$. This means that the seller can design a launch campaign which delivers a very negative signal and the high types will still wish to purchase. In other words, almost nothing that the seller could reveal about the product would be able to put the high types off. Inducing a more pessimistic belief when the launch campaign goes badly will be desirable for the seller since this actually increases the probability that a good type signal is drawn. When α is low, the beliefs of consumers are less dispersed and so a campaign which targets the high types following a bad signal is less likely to be successful, since the beliefs which the high types hold are close to the prior. This makes the realisation of a bad launch signal more likely.

Secondly, a higher α can also result in there being a larger fraction of high types in the population if $\mu > 0.5$. If $\mu > 0.5$, then the larger the fraction of high types, the more desirable it is to design a campaign which does not alienate them in the event that it sends them a bad signal.



(a) $\alpha = 0.6$ (b) $\alpha = 0.85$

Figure 2.5: Optimal Launch Campaign with $\lambda = 0.5$, $\mu = 0.2$

Case 2: $\mu^L\lambda < \mu^H\lambda < \underline{u}$

In this case λ is sufficiently low that no buyers wish to purchase unless they are persuaded to do so by the seller's launch campaign. Again, the optimal campaign takes one of two possible forms:

Proposition 2.2. *If $\mu^L\lambda < \mu^H\lambda < \underline{u}$ then the optimal launch campaign satisfies:*

- $\mu_g^L\lambda = \underline{u}$ and $\mu_b\lambda = 0$ if $\alpha \leq \bar{\alpha}$
- $\mu_g^H\lambda = \underline{u}$ and $\mu_b\lambda = 0$ if $\alpha > \bar{\alpha}$

Again there are two distinct launch campaigns which the seller may wish to commit to. When α is low, the seller wishes to have a launch campaign such that both types purchase following a good signal, yet no one purchases following a bad signal. When buyers have sufficiently strong private information, the seller adopts a niche marketing strategy where only high types are targeted, but with a higher probability of success than in the case where private information is low. A niche campaign is more beneficial when α is high but μ is low, since low types will need to have a signal which delivers posterior close to 1, making the campaign very unlikely to succeed.

As shown in Propositions 2.1 and 2.2, higher public beliefs about quality will induce different launch campaigns for the seller. The public belief about product quality is endowed by pre-launch reviews, which a feature of the model that is currently exogenous. With a view to later endogenising the behaviour of reviewers, I examine how the expected payoff for the seller will vary in λ.

	$\mu^L\lambda < \underline{u} < \mu^H\lambda$	$\mu^L\lambda < \mu^H\lambda < \underline{u}$
$\alpha \leq \bar{\alpha}$	$\mu_g^L\lambda = \underline{u}$; $\mu_b\lambda = 0$ (*Mass market*)	$\mu_g^L\lambda = \underline{u}$; $\mu_b\lambda = 0$ (*Mass market*)
$\alpha > \bar{\alpha}$	$\mu_g^L\lambda = \underline{u}$; $\mu_b^H\lambda = \underline{u}$ (*Segmented*)	$\mu_g^H\lambda = \underline{u}$; $\mu_b\lambda = 0$ (*Niche*)

Table 2.1: Summary of Optimal Launch Campaigns

When buyers have weak private information ($\alpha \leq \bar{\alpha}$), the seller will always opt for a *mass market* campaign. This launch campaign aims to persuade all buyers to purchase if successful and maximises the probability of success by setting $\mu_b^H\lambda = 0$. Therefore, the expected payoff for the seller from this launch campaign is simply $\Pr(g)$. When buyers have strong private information ($\alpha > \bar{\alpha}$), the seller will always opt for a *niche* campaign or a *segmented* campaign, yielding payoffs $\Pr(g, H)$ and $\Pr(g) + \Pr(b, H)$ respectively. The following lemma demonstrates that the seller always benefits from higher public belief λ for any campaign which they run.

Lemma 2.2. *Conditional on $\lambda > \underline{u}$, the seller's expected payoff is strictly increasing in λ*

The seller will always prefer to have a higher λ since it makes the job of persuading buyers to purchase easier. In the case where private information is weak, increasing λ means that it is easier to persuade low type buyers to purchase, since they will now have a lower threshold belief. Lowering the threshold belief increases the probability that the launch campaign is successful, since μ_g is now closer to μ.

In the case of strong private information, two effects work in different directions in the case of a *segmented* launch campaign. On the one hand, a higher λ means that the launch campaign is more likely to be successful for a larger proportion of buyers, but on the other hand, a higher λ means that there are more high types (in expectation), who would otherwise purchase and could be discouraged by a poor launch campaign signal. I show that increasing λ is always beneficial for the seller for all possible launch campaigns.

2.3.3 Endogenous Pricing

The results in this section stand in partial contrast to those in the existing literature. The findings of Lewis and Sappington (1994), Johnson and Myatt (2006) and Ottaviani and Prat (2001) would suggest that a seller should prefer 'all or nothing' launch campaigns, where buyers are allowed to discover their match type with certainty. Propositions 2.1 and 2.2 demonstrate that the seller never wishes to commit to 'all or nothing' campaigns in this setting (except when $\lambda = \underline{u}$), as when μ_g is close to 1, increases in μ_g will increase the probability that the campaign fails without increasing the payoff when it succeeds.

This result is a consequence of the assumption that the seller cannot set prices, and so is not able to extract the extra surplus which might arise from a higher μ_g. I now drop this assumption and allow for the seller to be a monopolist who to jointly picks their price and launch campaign. I show that allowing the monopolist full flexibility in setting their price restores the incentive to fully reveal information.

I assume that the seller is not able to condition their price on the outcome of the launch campaign. This assumption differs from Gill and Sgroi (2012), who allow conditional pricing, but is justified in this particular setting since s_i is privately observed by each buyer, unlike a public signal of quality. Therefore, the monopolist acts in two stages: first they set their price p and second they commit to their launch campaign by designing a signal. Buyers then observe their private signal σ_i, their launch signal s_i and the seller's price p; choosing to purchase if $p \leq \lambda \mathrm{E}\left[\theta_i \mid \sigma_i, s_i\right]$.

The analysis proceeds by replacing $\underline{u}$ from the previous case with p and then expressing the seller's payoff as a function of p. Conditional on the seller's choice of p, the optimal launch campaign is identical to the previous case, as the seller simply wishes to maximise the number of buyers who purchase at price p, rather than subject to a reservation utility $\underline{u}$. I maintain the assumption that the pre-launch reviewer endows a public belief of λ and so the seller never wishes to price at $p > \lambda$.

When buyers have weak private information ($\alpha \leq \bar{\alpha}$), then Propositions 2.1 and 2.2 have shown that the seller commits to a *mass market* launch campaign which targets all buyers. The payoff for the seller in this case is therefore given by $p \cdot \Pr(g)$, where the seller charges price p, that is accepted by all those with good signals.[12]

Proposition 2.3. *If $\alpha \leq \bar{\alpha}$ and the seller has the ability to set prices then a fully revealing launch campaign is always optimal.*

The insights of Lewis and Sappington (1994), Johnson and Myatt (2006) and Ottaviani and Prat (2001) are upheld when private information is weak and the seller is able to set their price prior to launch. As in Rayo and Segal (2010), who study a model of optimal information disclosure, allowing the seller to select optimal monetary transfers from buyers makes full revelation optimal. Intuitively, full revelation ensures that buyers gain no informational rents.

This complements the basic intuitions of Lewis and Sappington (1994) and Ottaviani and Prat (2001), who consider price discriminating monopolists that have the ability to offer a menu of contracts to buyers. My model partially extends the insights of Lewis and Sappington (1994) who place restrictions on the joint distribution of signals which the monopolist can provide, and Ottaviani and Prat (2001) who allow the seller to observe a public signal of quality and condition their contracts on the result. I show that even without offering a menu of contracts, or being able to observe the outcome of the signal, that a price setting monopolist will opt to allow buyers to discover their true valuation for the product when private information is weak.

When buyers have strong private signals ($\alpha > \bar{\alpha}$) then the seller either selects a niche campaign when $\mu^L\lambda < \mu^H\lambda < p$, giving payoff $p \cdot \Pr(g, H)$, or a segmented campaign when $\mu^L\lambda < p < \mu^H\lambda$, giving payoff $p \cdot (\Pr(b, H) + \Pr(g))$. It is straightforward to show that the seller's payoff is non-decreasing in p when they run a niche campaign, yet it is harder to show this for segmented campaigns. Although I have been unable to prove this result directly, numerical computations indicate that the same conclusion as in Proposition 2.3 holds in the case when private signals are strong.

To conclude this subsection, it is the inability to extract the surplus that buyers gain from fully learning their match type which drives sellers to obfuscate details of the

[12] Clearly, setting $p > \mu^L_g\lambda$ cannot be optimal.

product. It should be noted however, that when the seller can price their product freely and fully reveal the product details, then all surplus is extracted from buyers. When the setting of prices is restricted, high type buyers will be able to earn an informational rent. I now move on to examine how the behaviour of external reviewers influences the seller's campaign where the buyer's reservation utility is again exogenously given.

2.4 A Simple Model of Third Party Reviewers

Although the seller has private information about their quality, they are not able to disclose this credibly to buyers. For this reason, I assume that an independent third party (a reviewer) must evaluate the product's quality and produce a review which either passes or fails the product. I assume that the initial public belief of buyers is low enough that the firm must submit their product for review in order to attract buyers. Therefore I assume that $\lambda < \underline{u}$, so that even high type buyers will not wish to purchase unless the product is deemed high quality by the reviewer. This assumption fits the case where a seller is bringing a new product to market and must convince buyers to purchase over an incumbent product which has already been certified as good.

When evaluating a product, I assume that the reviewer receives an *impression* of the product which is a realisation of random variable $\tilde{X}$. The reviewer's impression of the product is distributed according to distribution function $G(x, q)$ where q is used to denote the quality of the good as either 1 or 0. Following the realisation $x \in X$, the reviewer then certifies the product as either passing or failing based on whether their impression of the product is sufficiently high (i.e. if $x \geq \bar{x}$). I assume throughout that $G(x \mid 1)$ and $G(x \mid 0)$ have full support on X so that noise is a necessary feature of the reviewer's technology and that all reviews are subject to some error.[13]

A higher threshold $\bar{x}$ for passing should be interpreted as a reviewer having higher standards which any product must meet before being certified as good quality. A reviewer therefore reveals the test result as either a pass or a fail according to the

[13] I assume also that $G(x \mid 1)$ and $G(x \mid 0)$ admit corresponding density functions $g(x \mid 1)$ and $g(x \mid 0)$.

function

$$T(x) = \begin{cases} \mathcal{P} & \text{if } x \geq \bar{x} \\ \\ \mathcal{F} & \text{if } x < \bar{x} \end{cases}$$

Following the result of the test, buyers will update their beliefs about quality to either $\lambda^{\mathcal{P}}$ or $\lambda^{\mathcal{F}}$. A condition which the density functions $g(x \mid 1)$ and $g(x \mid 0)$ could satisfy in order to satisfy this interpretation of the certification technology used by the reviewer is the strict monotone likelihood ratio property (MLRP) which was introduced by Milgrom (1981).[14] The standard intuition for this condition is that higher x are more likely to be observed when the quality of the good is 1 rather than 0. As shown by Eeckhoudt and Gollier (1995), whenever a family of densities satisfy MLRP, then the corresponding family of distribution functions satisfy what they call the monotone probability ratio property (MPRP):

Definition 2.1. A family of cumulative distribution functions $\{G(x \mid q)\}_{q \in \{0,1\}}$ has the strict monotone probability ratio property (MPRP) if and only if for every $\bar{x} > \bar{x}'$ and $q > q'$

$$\frac{G(\bar{x} \mid q)}{G(\bar{x} \mid q')} > \frac{G(\bar{x}' \mid q)}{G(\bar{x}' \mid q')}$$

The interpretation of this condition is similar to MLRP and can be most easily seen by using Bayes' rule to rewrite the condition as

$$\frac{\Pr(1 \mid x \leq \bar{x}) \Pr(x \leq \bar{x}) / \Pr(1)}{\Pr(0 \mid x \leq \bar{x}) \Pr(x \leq \bar{x}) / \Pr(0)} > \frac{\Pr(1 \mid x \leq \bar{x}') \Pr(x \leq \bar{x}') / \Pr(1)}{\Pr(0 \mid x \leq \bar{x}') \Pr(x \leq \bar{x}') / \Pr(0)}$$

and then cancelling to express the condition in terms of the posterior odds ratio

$$\frac{\Pr(1 \mid x \leq \bar{x})}{\Pr(0 \mid x \leq \bar{x})} > \frac{\Pr(1 \mid x \leq \bar{x}')}{\Pr(0 \mid x \leq \bar{x}')}$$

[14] A family of probability density functions $\{g(x \mid q)\}_{q \in \{0,1\}}$ has the strict monotone likelihood ratio property (MLRP) if and only if for every $\bar{x} > \bar{x}'$ and $q > q'$

$$\frac{g(\bar{x} \mid q)}{g(\bar{x} \mid q')} > \frac{g(\bar{x}' \mid q)}{g(\bar{x}' \mid q')}$$

As is well known, if the density functions satisfy MLRP then this implies first order stochastic dominance (FOSD).

The condition now states that the higher the threshold for passing the test, the more likely it is that the quality is 1 as opposed to 0 if the product fails. Intuitively, reviewers with higher thresholds fail more good products than reviewers with lower thresholds. For any two reviewers A and B, I will say that reviewer A is *harsher* than B if $\bar{x}_A > \bar{x}_B$. Since their threshold for passing the test is higher, they incorrectly fail a larger number of good quality products. I now show that if MPRP holds, then harsher reviewers always deliver higher posterior beliefs following any outcome of their test.

Proposition 2.4. *If Reviewer A is harsher than Reviewer B then* $\lambda_A^{\mathcal{F}} > \lambda_B^{\mathcal{F}}$ *and* $\lambda_A^{\mathcal{P}} > \lambda_B^{\mathcal{P}}$

If a reviewer has particularly high standards, then a pass provides a strong signal that the product quality is good. On the other hand, failing a harsh test is far less damning than failing a soft test and so $\lambda_A^{\mathcal{F}} > \lambda_B^{\mathcal{F}}$ and $\lambda_A^{\mathcal{P}} > \lambda_B^{\mathcal{P}}$ whenever A is harsher than B.

Taking the null hypothesis as $q = 0$, soft reviewers have a stronger tendency to commit 'Type I' errors by incorrectly passing poor quality products, whereas harsh reviewers have a stronger tendency to commit 'Type II' errors by incorrectly failing good quality products. Harsh reviewers have very little downside in terms of the beliefs which they generate post-test, as a pass is a strong signal of quality yet a fail carries little weight. The major detractor of having a harsh review is therefore the low pass rate associated with it. To summarise their relative merits, soft reviewers pass a larger fraction of products but provide very weak signals of quality following a pass, whereas harsh reviewers provide a stronger signal following a pass at the cost of having a lower pass rate.

A corollary of Proposition 2.4 is that if private information is weak ($\alpha \leq \bar{\alpha}$), then conditional on the result being either a pass or a fail, harsher reviewers will induce the seller to provide a less informative launch campaign. This arises due to the fact that when beliefs about quality are higher, the seller needs to do less to convince the buyers to purchase. In the case of weak private information, the seller will always run a mass market launch campaign, and since the beliefs about quality are higher, all cut-off

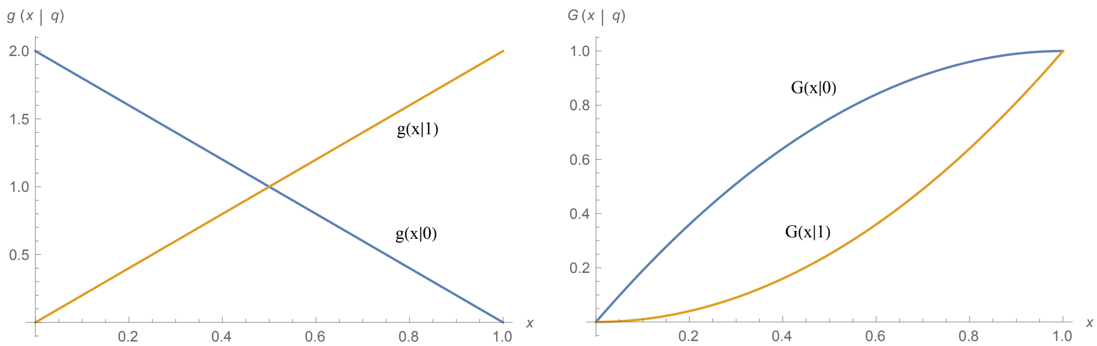


Figure 2.6: The PDF and CDF for $\tilde{X}$

beliefs for the seller to induce buyers to purchase are lower.

However, it may of course be the case that, on average, harsher reviewers induce the seller to undertake a more informative launch campaign. I now examine this issue in more detail.

2.4.1 The Seller's Optimal Reviewer

In order to proceed I will restrict attention to a particular functional form for the distribution of $\tilde{X}$. I assume that $X = [0,1]$ with $\{G(x \mid q)\}$ having full support on X. I assume that x is drawn from triangular distributions on X with densities given by $g(x \mid 0) = 2(1-x)$ and $g(x \mid 1) = 2x$ and corresponding distribution functions $G(x \mid 0) = 2x - x^2$ and $G(x \mid 1) = x^2$. These distributions satisfy MLRP and therefore $G(x \mid 1)$ first order stochastically dominates $G(x \mid 0)$.

From a buyer's perspective, all reviewers perform fair tests of the product and will induce the same posterior belief in expectation. However, the same is not true for the seller, since they have private information about their product's quality. Given that the high type seller knows that their product is of quality $q = 1$, their expectation of the reviewer's result will be more favourable. From the perspective of a high quality seller, different reviewer harshnesses will imply different expected posteriors. For example, a reviewer who is maximally soft or harsh (i.e. x equal to 0 or 1) will pass/fail every product and so will always induce posterior beliefs equal to the buyer's prior belief, even when the product is of high quality. However, any reviewer $\bar{x} \in (0,1)$ will provide an informative test of quality, which is more likely to certify high quality products as passes.

As Lemma 2.2 demonstrates, conditional on the seller being able to run an effective launch campaign, they will always benefit from having a test which maximises the posterior λ^T. However, some reviewers may have thresholds such that even if the product passes, the public belief will be so low that the seller is unable to use their launch campaign to manipulate the beliefs of buyers.

I consider now the possibility for the seller to select a reviewer from a continuum of reviewers $\bar{x} \in [0, 1]$. For example, a company may select amongst investment banks of differing standards to underwrite an IPO, or a record company may select amongst differing magazines to provide a preview of a new album. As mentioned previously, conditional on being able to run a launch campaign, the seller benefits from having the highest possible public belief about quality. However, there will be cases where some excessively soft reviewers will eliminate the possibility for the seller to run their launch campaign, and so the sellers choice $x^* \in [0, 1]$ must take account of this fact. Since it was assumed that $\lambda < \underline{u}$, only reviews which certify the product as passing will enable the seller to run their launch campaign. Reviewers who are too soft to induce a pass belief $\lambda^{\mathcal{P}} > \underline{u}$ will leave the seller with a payoff of 0, since the seller is unable to persuade buyers with high match types to purchase.

The choice of reviewer by a seller with private information also opens up the possibility of signalling via choice of test (as in the classic literature on signalling). I show below that all equilibria must be pooling equilibria, as here it is costless for the low type to mimic the high type's choice.

Lemma 2.3. *The low type seller always pools on any test chosen by the high type*

To support this equilibrium I assume that beliefs are such that if buyers observe an action off the equilibrium path, then they place probability 1 on the seller being a low quality type. Gill and Sgroi (2012) also place this restriction on off path beliefs and note that this is a variant of the 'undefeated' equilibrium refinement of Mailath et al. (1993). I also appeal to a version of this refinement used by Perez-Richet (2014) in his study of models of Bayesian persuasion where the seller has private information about their type.[15] Intuitively, this refinement states that for any off path deviation which is observed, the buyers must put positive weight only to those types who would benefit from such a deviation if it were an equilibrium. Since the high types are strictly worse off in any other pooling equilibrium, buyers must assume that any such deviation is by a low type.

Separation by the low type is never beneficial, even weakly, so long as tests are noisy to some small degree. Any type who separates reveals their quality as $q = 0$ but since I have assumed full support for $\{G(x \mid v)\}$, no review could ever perfectly reveal product quality and so they always do better by pooling. Consequently, a seller who observes their type as low prefers to mimic the high type.

Focusing therefore on the high type seller's choice of reviewer, I define X' as the set of reviewer thresholds which generate a pass belief sufficient to allow the seller to run an effective launch campaign (i.e. $X' \equiv \left\{\bar{x} \in X \mid \lambda_{\bar{x}}^{\mathcal{P}} \geq \underline{u}\right\}$). Proposition 2.4 tells us that $\lambda_{\bar{x}}^{\mathcal{P}}$ is increasing in $\bar{x}$ and so $X' = [x', 1]$, where x' solves $\lambda_{x'}^{\mathcal{P}} = \underline{u}$. Recalling the result of Lemma 2.2, the high type seller's problem is

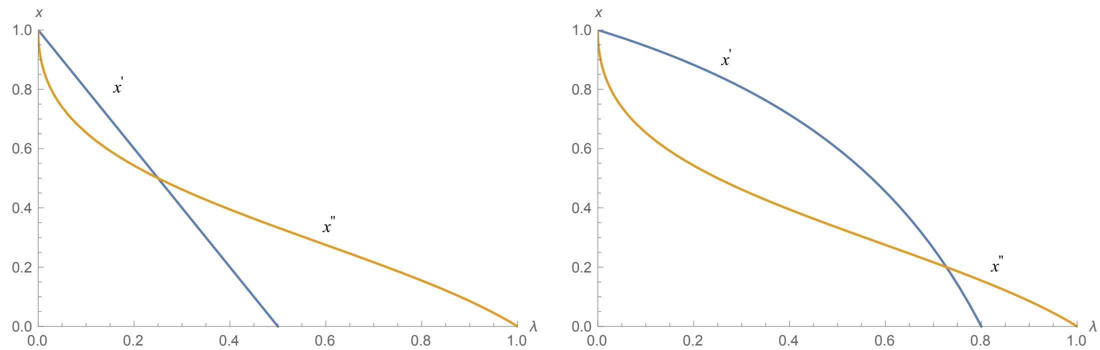
$$\max_{\bar{x} \in X'} \Pr(\mathcal{P} \mid q = 1)\, \lambda_{\bar{x}}^{\mathcal{P}} \tag{2.5}$$

Using x^* to denote the solution to (2.5) I now show that softer tests will be desired when public beliefs about quality are higher

Proposition 2.5. *The threshold of the seller's optimal reviewer x^* is weakly decreasing in λ*

The seller's optimal reviewer x^* is strictly decreasing in λ until x^* is such that lowering it further would imply that $\lambda_{\bar{x}}^{\mathcal{P}} < \underline{u}$, in which case the optimal reviewer has threshold x'. Harsh reviewers will generally be preferable to the seller when λ is low for two reasons. Firstly, when the public has low prior beliefs about quality, then harsh reviews are expected to make fewer 'Type II' errors from a buyer's perspective. The major detractor of harsh tests is that they fail too many good products, yet when the seller knows her quality is high but the public believe that the product is very likely

[15] See the definition of $\mathcal{R}_1$ in the working paper version of Perez-Richet (2014) for more details.



(a) $\underline{u} = 0.5$

(b) $\underline{u} = 0.8$

Figure 2.7: Effect of λ on the Seller's Choice of Reviewer

to be of low quality, then a pass from a harsh reviewer provides a very strong signal. For this reason, harsh reviews are more attractive to the seller when the public belief λ is low, since the buyers will view a fail as having eliminated the vast majority of bad quality products.

Harsh reviewers have a second feature which benefits the seller when the prior belief is low. Since the seller must generate a posterior belief $\lambda^{\mathcal{P}}$ which at least matches the reservation utility $\underline{u}$, the sellers are drawn to the harsh end of the reviewer spectrum, since only the high posterior induced by harsh reviewers will give them a platform to promote their product. Solving $\lambda^{\mathcal{P}}_{x'} = \underline{u}$ reveals that

$$x' = \frac{\underline{u} - \lambda}{\underline{u}(1-\lambda) + \lambda(1-\underline{u})}$$

which is also a decreasing function of λ, meaning that as the prior gets lower, the seller's choice is restricted to only the harshest reviewers. Therefore, as λ decreases, the high quality seller (who wants to reveal their quality) may be forced to pick a reviewer who is excessively harsh. This effect is more pronounced for higher $\underline{u}$, as shown in Figure 2.7, where I use x'' to denote the optimal reviewer in the unconstrained problem (i.e. where $\underline{u} = 0$). The seller's optimal choice of reviewer x^* is effectively the upper envelope of these two functions.

This highlights the difference between my results and those of Gill and Sgroi (2012), who show that when the seller can price their product conditional on the result of the review that the seller may desire the softest possible test available. In contrast to my model, Gill and Sgroi (2012) contend there can be demand for ultra-soft reviewers who effectively rubber stamp every product, even when the seller has private information that they are high quality. In my case, because the seller must run a launch campaign to persuade sceptical buyers to endorse their product, a sufficient level of credibility must be met before buyers will be open to persuasion.

This also has a consequence for the launch campaign that the seller will eventually run. When the seller is forced to pick a reviewer x' who will deliver posterior $\lambda_{x'}^{\mathcal{P}} = \underline{u}$, Corollary 2.1 tells us that the seller will opt for a fully revealing launch campaign. If we view the value of the outside option as a proxy for the saturation of the product market, then when the market is more saturated, the seller will be driven to submit their product to overly harsh reviewers and reveal all product information at launch. Overly harsh reviewers in this case means reviewers who do not maximise the high type seller's expectation of the buyer's posterior.

As for the welfare of the buyers, these two effects go in opposite directions. Overly harsh tests will fail a larger number of good products, but more information about their match type will always be beneficial. I now move on to briefly examine how the choice of reviewer (and resulting launch campaign) impacts the buyers.

2.4.2 The Buyers' Optimal Reviewer

I now consider how the positioning of the reviewer affects the welfare of the buyers, taking in to account the seller's design of launch campaign. In this case, I assume that the reviewer is 'buyer focused' and wishes to maximise the buyer's welfare by their choice of $\bar{x}$. Alternatively, the reviewer could be selected by a social planner, who wishes to install a consumer watchdog or standards setting organisation to regulate product quality.

From the buyers' perspective, all reviewers deliver the same posterior belief in expectation, however not all reviewers are equally desirable. In the absence of a launch campaign run by the seller, the optimal reviewer for the buyers is one who will maximise the posterior belief that product is high quality, conditional on the buyers purchasing.

If, for example, a reviewer's threshold for passing were so low that no buyers will ever purchase, (even if the product passes the test), then this reviewer provides no service for the buyers (or the seller). A buyer will only find a review useful if the outcome of the review will be able to have an impact on their purchasing decision.[16]

If we take account of the action of the seller through their launch campaign, then the optimal reviewer for the buyers may differ, as the amount of 'match type' information released by the seller is contingent on $\lambda^{\mathcal{P}}$. Given the results in the previous section, low type buyers always earn $\underline{u}$ in expectation, for any reviewer $\bar{x}$. If the reviewer fails the product, then they take their outside option, yet if they pass the product the seller runs a launch campaign which either gives $\underline{u}$ in expectation or simply $\underline{u}$ if they run a niche campaign. Therefore, only high type buyers stand to benefit, conditional on the seller's launch campaign and I focus therefore on reviewers who aim to maximise the expected welfare of the high type buyers.[17]

The only possibility for the high type buyers to earn a payoff above $\underline{u}$ is when the seller runs a mass market or segmented campaign and the buyer receives a good signal. The ex-ante expected payoff for high type buyers is therefore $\Pr(g, \mathcal{P} \mid H)\, \lambda^{\mathcal{P}} \mu_g^H + (1 - \Pr(g, \mathcal{P} \mid H))\underline{u}$. To keep this section of the chapter tractable I restrict attention to the case of $\alpha \leq \bar{\alpha}$, focusing on the case where the seller always runs a mass market launch campaign. By doing so and noting that result of the review and the realisation of private signals about match types are independent events, $\Pr(g, \mathcal{P} \mid H)\, \lambda^{\mathcal{P}} \mu_g^H$ can be re-written as $\Pr(g \mid H, \mathcal{P}) \Pr(\mathcal{P})\, \lambda^{\mathcal{P}} \mu_g^H = \Pr(\mathcal{P} \mid q = 1)\, \lambda \mu^H$, meaning that the problem for a buyer focused reviewer is

$$\max_{\bar{x} \in X'} \Pr(\mathcal{P} \mid q = 1)\, \lambda \mu^H + (1 - \Pr(g, \mathcal{P} \mid H))\underline{u} \tag{2.6}$$

In addition to the usual trade off between harsh and soft reviewers, there is another factor which influences the decision. Although softer reviews generate lower post-review

[16] Similarly, if the prior were such that $\lambda > \underline{u}$, then excessively harsh reviewers who induce beliefs $\lambda^{\mathcal{F}} > \underline{u}$ following a fail would be of no value to the buyers.

[17] In an alternative version of the model, one could analyse the incentives of reviewers who cater to high type buyers, perhaps hoping to extract surplus from them via advertising or charging a fee to access the review. In that case, however, the signal is no longer public and so low and high type buyers would have different beliefs about q also.

beliefs, this induces the seller to reveal more information about buyer's match types. Therefore, soft reviewers can become more appealing to buyers when the risks associated with them become relatively insignificant.

Lemma 2.4. *If the buyers have weak private information then the buyers' optimal reviewer $\hat{x}$ is decreasing in λ and is given by*

$$\hat{x} = \frac{(1-\lambda)\underline{u}}{(1-\lambda)\underline{u} + \lambda(1-\underline{u})}$$

As in the seller's case, the optimal reviewer harshness is decreasing in λ. Note also that $\hat{x} > x'$, so that the buyers never wish to assign a reviewer who is so soft as to force the seller to reveal all the information about the quality of the good. That is not to say, however, that the buyers always desire a harsher reviewer than the seller, as the following result shows:

Proposition 2.6. *When $\alpha \leq \bar{\alpha}$ and $\underline{u} < \frac{1}{4}$, then for λ sufficiently high, the buyers would select a softer reviewer than the seller.*

Somewhat counter-intuitively, the buyers may opt for a softer reviewer than the high type seller would choose. This occurs when the outside option $\underline{u}$ is low (buyers have little to lose if the soft reviewer passes a bad product) and the prior λ is sufficiently high (fewer bad products exist in expectation). When these conditions hold, the high type buyers will choose a softer reviewer than the seller would, in order to induce the seller to provide more information about their match type. Learning more about their match type is always beneficial for the buyers, but when there is a larger share of bad products in the market, or when the opportunity cost of their outside option is high, this incentive to select soft reviewers is diminished.

2.5 Discussion and Conclusion

The contribution of this chapter to the literature on persuasion and the revelation of product information by a seller is threefold. Firstly, I add to the Bayesian persuasion literature by considering the case of multiple receivers with private signals and show that the framework can be extended to this setting. Secondly, I add to the literature on provision of private signals by a seller, by examining the structure of optimal marketing campaigns. Thirdly, I develop the literature on the optimal choice of pre-launch reviewer initiated by Gill and Sgroi (2008) and Gill and Sgroi (2012).

Although Kamenica and Gentzkow (2011) mention that their model can be extended to the case of private information and (separately) multiple receivers, I have shown that their framework can be applied more broadly. To my knowledge, this model is the first to examine the case where a sender must design a signal to persuade a group of buyers with differing beliefs. The Bayesian persuasion framework has previously been applied to analyse the control of public information, yet I have shown that it can also be readily applied to the case of private information. Therefore, the key difference between my approach and the standard one is that although the signal generating mechanism is static, it generates potentially different signals for each buyer, rather than being publicly announced. This is an important application because it allows the model to extend to a number of different settings where a single seller wishes to persuade a multitude of heterogeneous buyers to endorse their product.

This model is also amongst the first to embed the Bayesian persuasion framework in a setting of heterogeneous types. One similar contribution is the recent paper of Perez-Richet (2014) who looks at a model of interim Bayesian persuasion whereby the seller of the good (not the buyer) has private information about their type before designing the signal. Although both the seller and the buyers have private information in my model, it is the private information of the buyers which plays a key role in determining the seller's optimal information structure. Another recent paper which allows a sender to condition their design of signal on their private information is Kolotilin (2014). The key differences between my model and that of Kolotilin (2014) are firstly that I focus on the case of multiple buyers, secondly, he examines signals which are ex-ante optimal for the seller, rather than focusing on high type preferred signals, and lastly the receivers learn about the sender's type, rather than their own type.

In this model, firms send private signals about product features in the form of a launch campaign, whereas reviewers release a public signal about the underlying quality of the good. As mentioned by Johnson and Myatt (2006) and Ottaviani and Prat (2001), although the literature on the value of public information for a monopolist is well developed, the literature on the value of private information is still lacking. This chapter can therefore be viewed as a step in this direction which takes advantage of some of the recent developments in the field of information control and the design of optimal signals.

The results presented in Section 2.3 stand in partial contrast to some of those in the existing literature (e.g. Johnson and Myatt (2006), Lewis and Sappington (1994), Ottaviani and Prat (2001)). The main result of Lewis and Sappington (1994) is that the seller prefers either no information release or a signal which allows buyers to fully reveal their type. I have shown that fully revealing signals are only optimal for one special case, and sellers will generally wish to obfuscate the product details in order to increase the chance of a successful product launch. A key difference between my paper and both Lewis and Sappington (1994) and Ottaviani and Prat (2001) is that they make the assumption that the seller is a price discriminating monopolist and is able to offer a menu of contracts to extract surplus from both buyer types. Indeed, when the seller has full monopoly power and private signals are weak, then fully revealing tests are also optimal in my set-up, under weaker assumptions about the ability to price discriminate than Ottaviani and Prat (2001) and under more general signals than Lewis and Sappington (1994).

My model therefore shows that restricting the ability of firms to set prices can induce intentional obfuscation of product information by the seller. As is the case in all Bayesian persuasion models, the sender of a signal finds it difficult to induce posteriors for the receiver which are very far away from their prior belief. For that reason, launch campaigns which aim to fully reveal types have a larger risk of failing as they must reveal information which is 'surprising' to buyers and viewed (correctly) as ex-ante very unlikely.

Finally, I also bring out several novel points about the choice of optimal pre-launch reviewers. The main result is Proposition 2.6 which demonstrates cases where the buyers of a good wish to commission reviewers who are softer than those who would be commissioned by the seller. The optimal reviewer will be a decreasing function of the public's belief about quality, and unlike Gill and Sgroi (2012) where only the harshest or softest reviewers are used, my model accounts for the existence of reviewers of varying standard, who cater to products with different priors.

The other contribution which I make in this area is the result of Proposition 2.5, which is more in line with the paper of Gill and Sgroi (2008), who find that 'mildly tough' tests (i.e. just tough enough to induce informational cascades on buying) are optimal. Sellers will often be forced to submit their product for overly harsh tests, as only these test will generate sufficient credibility for the sellers to run their launch campaigns effectively. We should therefore expect to see the thresholds of reviewers being skewed more toward the harsher side, since soft reviewers will only be useful when the outside option of buyers is very low.

Recent work by Ely et al. (2014) has examined dynamic information revelation in the Bayesian persuasion framework with applications to the design of sports games, mystery novels and auctions. The 'seller' in this case has preferences over the entire sequence of the buyer's beliefs and wishes to provide information to maximise the amount of 'suspense' and 'surprise' experienced by the buyer. The task of selecting the optimal sequence of signals from the space of all possible belief paths can be simplified drastically by applying the insights of KG's original model. In their paper they focus on Markov processes, utilising the martingale property of beliefs which characterises feasible signals in KG to select the optimal policy amongst all Markov belief martingales.

Although their approach opens up the possibility of studying multi-period launch campaigns, such campaigns seem unlikely to be beneficial for the seller in my particular framework. Match types are independent across buyers and the seller learns nothing about the quality of their good which they do not already know, so they cannot use information gained in the first period to benefit them. Moreover, observational learning by consumers cannot take place since they only have a public signal of quality. Introducing other aspects such of word of mouth communication or observational learning may have an impact on the optimal dynamic campaign, but I leave this avenue for future work.

Appendix - Proofs

Proof of Lemma 2.1. Consider a fully revealing campaign run by the seller. If $\bar{\mu}^H < \bar{\mu}^L < 1$ then by (2.2) the seller can increase the probability of a good signal for all types by lowering μ_g to $\mu_g = \bar{\mu}^L < 1$. If $\underline{u} = \lambda$ then $\bar{\mu}^H = \bar{\mu}^L = 1$ and only buyers who are revealed to have match type $\theta = 1$ with certainty will purchase. In this case $\mu_g < 1$ yields zero payoff so $\mu_g = 1$ is optimal. To see that $\mu_b = 0$ is optimal in this case note again by (2.2) that the seller can increase the probability of a good signal by lowering μ_b as much as possible.

If $1 < \bar{\mu}^H < \bar{\mu}^L$ then buyers cannot be persuaded to purchase and all launch campaigns yield the same payoff. $\square$

Proof of Proposition 2.1. First I show that an optimal launch campaign satisfies $\mu_g^L \lambda = \underline{u}$. If $\mu_g^L \lambda < \underline{u}$ then low type buyers will never purchase and so the seller cannot benefit from such a signal. If $\mu_g^L \lambda > \underline{u}$ then low types buy following a good signal, yet (2.2) implies that reducing μ_g^L to satisfy $\mu_g^L = \underline{u}/\lambda$ increases τ^L, and hence τ.

To show that either $\mu_b^H \lambda = 0$ or $\mu_b^H \lambda = \underline{u}$ for any optimal launch signal, note that the martingale condition implies that $\mu_b^H \in [0, \mu_0^H)$. Since only high types can ever buy following a bad signal, if the posterior is such that $\mu_b^H \in [\underline{u}/\lambda, \mu_0^H)$ then only they will purchase. Yet for $\mu_b^H > \underline{u}/\lambda$, (2.2) implies that we strictly benefit from lowering μ_b^H to satisfy $\mu_b^H \lambda = \underline{u}$. If the bad posterior is such that $\mu_b^H \in [0, \underline{u}/\lambda)$ then high types never buy, and again (2.2) implies that we would strictly benefit from lowering μ_b^H to $\mu_b^H = 0$ to maximise the probability of a good signal.

Finally, to show that there exists a cut-off $\bar{\alpha}$, using (2.2) I write the payoff of the second campaign ($\mu_g^L \lambda = \underline{u}$;$\mu_b^H \lambda = 0$) in terms of the reference belief as

$$\Pr\left(s_i = g\right) = \frac{\mu}{\bar{\mu}^L} \tag{2.7}$$

And the payoff for the second campaign ($\mu_g^L \lambda = \underline{u}$;$\mu_b^H \lambda = \underline{u}$) as

$$\Pr(g) + \Pr(b)\Pr(H \mid b) = \frac{\mu - \bar{\mu}^H}{\bar{\mu}^L - \bar{\mu}^H} + \frac{\bar{\mu}^L - \mu}{\bar{\mu}^L - \bar{\mu}^H}\left(\alpha\bar{\mu}^H + (1-\alpha)(1-\bar{\mu}^H)\right) \tag{2.8}$$

Where the expression for $\Pr(H \mid b)$ is arrived by first rewriting it as $\Pr(H \mid b) = \frac{\Pr(H,b)}{\Pr(b)}$, then conditioning on θ_i to give

$$\frac{\Pr(H, b \mid \theta_i = 1)}{\Pr(b)}\Pr(\theta_i = 1) + \frac{\Pr(H, b \mid \theta_i = 0)}{\Pr(b)}\Pr(\theta_i = 0)$$

Finally, noting the conditional independence of σ_i and s_i gives

$$\Pr(H \mid \theta_i = 1)\frac{\Pr(b \mid \theta_i = 1)\Pr(\theta_i = 1)}{\Pr(b)} + \Pr(H \mid \theta_i = 0)\frac{\Pr(b \mid \theta_i = 0)\Pr(\theta_i = 0)}{\Pr(b)}$$

Using (2.7) and (2.8) we see that the first campaign is preferred to the second when

$$\frac{\mu_0}{\bar{\mu}^L} \geq \frac{\mu_0 - \bar{\mu}^H}{\bar{\mu}^L - \bar{\mu}^H} + \frac{\bar{\mu}^L - \mu_0}{\bar{\mu}^L - \bar{\mu}^H}\left(\alpha\bar{\mu}^H + (1-\alpha)(1-\bar{\mu}^H)\right)$$

Which simplifies to

$$\bar{\mu}^H \geq \bar{\mu}^L\left(\alpha\bar{\mu}^H + (1-\alpha)(1-\bar{\mu}^H)\right)$$

Substituting in from (2.4) reveals after some manipulation that this condition holds if and only if

$$\alpha\left(\underline{u}/\lambda\right) + (1-\alpha)\left(1 - \left(\underline{u}/\lambda\right)\right) \geq \alpha^2$$

Since by assumption $\underline{u} < \lambda$, when $\alpha = \frac{1}{2}$ this holds, yet when $\alpha = 1$ this is violated. The left hand side is bounded between $\frac{1}{2}$ and $(\underline{u}/\lambda)$ whilst the right hand side is bounded between $\frac{1}{4}$ and 1. Both sides are continuous monotone functions of α, therefore the existence of a unique cut-off is established by the intermediate value theorem. This cut-off can be found using the quadratic formula to yield

$$\bar{\alpha} = \frac{1}{2}\left(\left(2\left(\underline{u}/\lambda\right) - 1\right) + \sqrt{5 - 8\left(\underline{u}/\lambda\right) + 4\left(\underline{u}/\lambda\right)^2}\right)$$

□

Proof of Proposition 2.2. By a similar argument to Proposition 2.1, the seller will only ever wish to pick $\mu_b^H = 0$, since no buyers can ever purchase after a bad signal, and setting $\mu_b^H = 0$ maximises the probability of a good signal being drawn. As in the proof of Proposition 2.1, the good signal targets either all buyers or high type buyers only, so either $\mu_g^L \lambda = \underline{u}$ or $\mu_g^H \lambda = \underline{u}$.

When comparing the payoff of the two campaigns, targeting all buyers with $\mu_g^L \lambda = \underline{u}$ is preferred to targeting only high types with $\mu_g^H \lambda = \underline{u}$ if

$$\Pr\left[s_i = g\right] \geq \Pr\left[s_i = g'\right] \Pr\left[\sigma_i = H \mid s_i = g'\right]$$

Or

$$\frac{\mu}{\bar{\mu}^L} \geq \frac{\mu}{\bar{\mu}^H} \left(\alpha\bar{\mu}^H + (1-\alpha)(1-\bar{\mu}^H)\right)$$

Yet, this is an identical condition to the one in Proposition 2.1, and so the same cut-off applies. □

Proof of Lemma 2.2. If $\alpha \leq \bar{\alpha}$ then the seller's expected payoff is $\Pr(g) = \frac{\mu}{\mu_g}$. The threshold belief $\bar{\mu}^L = \mu_g$ is strictly decreasing in λ, therefore the seller's expected payoff strictly increases in λ. If $\alpha > \bar{\alpha}$ and $\mu^L\lambda < \mu^H\lambda < \underline{u}$, the seller runs a niche campaign which gives payoff $\Pr(g, H) = \Pr(g)\Pr(H \mid g) = \frac{\mu}{\mu_g}\left(\alpha\mu_g + (1-\alpha)(1-\mu_g)\right)$. Since $\bar{\mu}^H = \mu_g$, a similar argument shows that the payoff is strictly increasing in λ.

Finally, if $\alpha > \bar{\alpha}$ and $\mu_0^L\lambda < \underline{u} < \mu_0^H\lambda$ then the seller's payoff from a segmented campaign is $\Pr(g) + \Pr(b, H)$ which can be written as

$$\Pr(g) + (1 - \Pr(g \mid H))\Pr(H) = \Pr(H) + \Pr(L \mid g)\Pr(g) \tag{2.9}$$

Using (2.4), $\Pr(L \mid g)$ becomes

$$\Pr(L \mid g) = (1-\alpha)\frac{\alpha\underline{u}}{\alpha\underline{u} + (1-\alpha)(\lambda - \underline{u})} + \alpha\frac{(1-\alpha)(\lambda-\underline{u})}{\alpha\underline{u} + (1-\alpha)(\lambda - \underline{u})} = \frac{\alpha(1-\alpha)\lambda}{\alpha\underline{u} + (1-\alpha)(\lambda - \underline{u})}$$

then multiplying by $\Pr(g)$ means that $\Pr(L \mid g)\Pr(g)$ is given by

$$\frac{\alpha(1-\alpha)\lambda}{\alpha\underline{u}+(1-\alpha)(\lambda-\underline{u})} \cdot \frac{\mu - \frac{(1-\alpha)\underline{u}}{(1-\alpha)\underline{u}+\alpha(\lambda-\underline{u})}}{\frac{\alpha\underline{u}}{\alpha\underline{u}+(1-\alpha)(\lambda-\underline{u})} - \frac{(1-\alpha)\underline{u}}{(1-\alpha)\underline{u}+\alpha(\lambda-\underline{u})}}$$

Multiplying the numerator and denominator of the second term by $(1-\alpha)\underline{u}+\alpha(\lambda-\underline{u})$, then simplifying eventually leads to

$$\Pr(L \mid g)\Pr(g) = \alpha(1-\alpha)\lambda\left[\frac{\alpha\mu}{(2\alpha-1)\underline{u}} - \frac{(1-\alpha)(1-\mu)}{(2\alpha-1)(\lambda-\underline{u})}\right] \tag{2.10}$$

The term in square brackets is positive, as this is simply $\frac{\Pr(g)}{\alpha\underline{u}+(1-\alpha)(\lambda-\underline{u})}$. Then we see that differentiating (2.10) with respect to λ proves that $\Pr(L \mid g)\Pr(g)$ is strictly increasing in λ for $\lambda > \underline{u}$. □

Proof of Proposition 2.3. If $\alpha \leq \bar{\alpha}$ the seller wishes to conduct a mass market campaign such that $\mu_g^L\lambda = p$ and $\mu_b^L\lambda = 0$. Since the threshold belief $\bar{\mu}^L$ is a function of p, we write the payoff from a mass market campaign as

$$p\frac{\mu}{\bar{\mu}^L} = p\mu\frac{\alpha p + (1-\alpha)(\lambda - p)}{\alpha p}$$

which is a linear function of p. Simplifying this expression gives $\frac{\mu}{\alpha}(\alpha p + (1-\alpha)(\lambda-p))$, which is increasing in p since $\alpha \in (0.5, 1)$. The constraint that $p \leq \lambda$ ensures that $p = \lambda$ is the optimal price for any mass market campaign. Given the fact that $\mu_g^L\lambda = p$ for any mass market campaign, this shows that $\mu_g^L = 1$. □

Proof of Proposition 2.4. We want to show firstly that if $\bar{x}_A > \bar{x}_B$ then $\Pr(1 \mid x < \bar{x}_A) > \Pr(1 \mid x < \bar{x}_B)$. Or, since the state space is binary, we want to show that $\frac{\Pr(1|x<\bar{x}_A)}{\Pr(0|x<\bar{x}_A)} > \frac{\Pr(1|x<\bar{x}_B)}{\Pr(0|x<\bar{x}_B)}$. Using Bayes' rule this condition reduces to $\frac{G(\bar{x}_A|1)}{G(\bar{x}_A|0)} > \frac{G(\bar{x}_B|1)}{G(\bar{x}_B|0)}$, which is satisfied due to the monotone probability ratio property, proving that $\lambda_A^F > \lambda_B^F$.

Next, to show that $\frac{\Pr(1|x\geq\bar{x}_A)}{\Pr(0|x\geq\bar{x}_A)} > \frac{\Pr(1|x\geq\bar{x}_B)}{\Pr(0|x\geq\bar{x}_B)}$, Bayes' rule gives $\frac{1-G(\bar{x}_A|1)}{1-G(\bar{x}_A|0)} > \frac{1-G(\bar{x}_B|1)}{1-G(\bar{x}_B|0)}$ which can be re-expressed as

$$G(\bar{x}_A \mid 1)\, G(\bar{x}_B \mid 0) + G(\bar{x}_A \mid 0) + G(\bar{x}_B \mid 1) > G(\bar{x}_A \mid 0)\, G(\bar{x}_B \mid 1) + G(\bar{x}_A \mid 1) + G(\bar{x}_B \mid 0)$$

Due to FOSD we can replace $G(\bar{x}_B \mid 0)$ with $G(\bar{x}_B \mid 1)$ and cancel to give

$$G(\bar{x}_A \mid 1)\, G(\bar{x}_B \mid 0) - G(\bar{x}_A \mid 0)\, G(\bar{x}_B \mid 1) > G(\bar{x}_A \mid 1) - G(\bar{x}_A \mid 0)$$

Where $G(\bar{x}_A \mid 1) - G(\bar{x}_A \mid 0) < 0$ again from FOSD.

The previous inequality holds if $G(\bar{x}_A \mid 1)\, G(\bar{x}_B \mid 0) - G(\bar{x}_A \mid 0)\, G(\bar{x}_B \mid 1) > 0$, which is satisfied since $\frac{G(\bar{x}_A|1)}{G(\bar{x}_A|0)} > \frac{G(\bar{x}_B|1)}{G(\bar{x}_B|0)}$ by assumption, proving that $\lambda_A^P > \lambda_B^P$. □

Proof of Lemma 2.3. Assume that a separating equilibrium exists, since the choice of reviewer is costless, the low type seller could mimic the decision of the high type seller and increase their payoff. To show that pooling is an equilibrium I repeat the argument in the text. Off path beliefs ensure that neither high type or low type seller can deviate to increase their payoff. □

Proof of Proposition 2.5. Using the definition of $G(\cdot \mid q)$, the seller's objective function can be written as

$$\Pr(\mathcal{P} \mid q = 1) \frac{\Pr(\mathcal{P} \mid q = 1)\, \lambda}{\Pr(\mathcal{P})} = (1 - x^2) \frac{(1 - x^2)\lambda}{\Pr(\mathcal{P})}$$

This function is strictly concave in x for $x \in [0, 1]$, so first order conditions imply that

$$2(1 - x^2)(-2x)\lambda \Pr(\mathcal{P}) - (1 - x^2)^2 \lambda \frac{d \Pr(\mathcal{P})}{dx} = 0$$

Rearranging and cancelling terms gives the condition $-4x \Pr(\mathcal{P}) = (1 - x^2) \frac{d \Pr(\mathcal{P})}{dx}$, then substituting in for $\Pr(\mathcal{P})$ this becomes

$$-4x\left((1-x^2)\lambda+(1-x)^2(1-\lambda)\right)=(1-x^2)\left(-2x\lambda-(2-2x)(1-\lambda)\right)$$

After rearranging we can express this as

$$2x(1-x)(1-\lambda)=(1+x)(1-\lambda-x)$$

Finally, expanding and collecting all terms on the left hand side we arrive at

$$x^2\left(2\lambda-1\right)+x(2-\lambda)+(\lambda-1)=0 \tag{2.11}$$

Implicit differentiation of (2.11) with respect to λ reveals that if $x\in[0,1]$ and $\lambda\in[0,1]$ then

$$\frac{\partial x^*}{\partial\lambda}=\frac{-2x^2-(1-x)}{\lambda(4x-1)+2\left(1-x\right)}$$

It is easily verified that the denominator is >0, proving that x^* is decreasing in λ. To find the optimal x we solve the quadratic in (2.11), taking only the solution such that x lies in the interval $[0,1]$. This assumes that x in (2.11) satisfies the condition that $\lambda_x^{\mathcal{P}}\geq\underline{u}$. If this does not hold then x^* is given by a boundary case where it solves the equation

$$\frac{(1-x^2)\lambda}{(1-x^2)\lambda+(1-(2-x^2))(1-\lambda)}=\underline{u}$$

□

Proof of Lemma 2.4. Rewriting (2.6) by again exploiting the independence of T and σ we can express the objective function as

$$\Pr\left(\mathcal{P}\mid q=1\right)\lambda\mu^H+\left(1-\Pr\left(g\mid\mathcal{P},H\right)\Pr\left(\mathcal{P}\right)\right)\underline{u}$$

The results of Proposition 2.1 and 2.2 show that $\Pr\left(g\mid\mathcal{P},H\right)=\frac{\mu^H}{\mu_g^H}$, substituting in for

$\mu_g^H = \frac{\alpha^2\underline{u}}{\alpha^2\underline{u}+(1-\alpha)^2(\lambda^{\mathcal{P}}-\underline{u})}$ we get

$$\Pr(\mathcal{P} \mid q=1)\,\lambda\mu^H + \left(1-\mu^H\left(1+\frac{(1-\alpha)^2(\lambda^{\mathcal{P}}-\underline{u})}{\alpha^2\underline{u}}\right)\Pr(\mathcal{P})\right)\underline{u}$$

Or

$$\Pr(\mathcal{P} \mid q=1)\,\lambda\mu^H + \left(1-\mu^H\left(\Pr(\mathcal{P})\left(1-\frac{(1-\alpha)^2}{\alpha^2}\right)+\frac{(1-\alpha)^2}{\alpha^2}\frac{\Pr(\mathcal{P}\mid q=1)\,\lambda}{\underline{u}}\right)\right)\underline{u}$$

Further manipulation yields

$$\Pr(\mathcal{P} \mid q=1)\,\lambda\mu^H\left(1-\frac{(1-\alpha)^2}{\alpha^2}\right)+\underline{u}-\underline{u}\mu^H\left(\Pr(\mathcal{P})\left(1-\frac{(1-\alpha)^2}{\alpha^2}\right)\right)$$

and finally

$$\mu^H\frac{(2\alpha-1)}{\alpha^2}\left(\Pr(\mathcal{P} \mid q=1)\,\lambda\,(1-\underline{u})-\Pr(\mathcal{P} \mid q=0)\,(1-\lambda)\,\underline{u}\right)+\underline{u}$$

Substituting in for $\Pr(\mathcal{P} \mid q=1)$ and $\Pr(\mathcal{P} \mid q=0)$ and then maximising with respect to x gives the expression for $\hat{x}$. □

Proof of Proposition 2.6. The buyer's optimal reviewer is given by

$$\hat{x} = \frac{(1-\lambda)\underline{u}}{(1-\lambda)\underline{u}+\lambda(1-\underline{u})}$$

To find the seller's optimal (interior) reviewer we solve (2.11) and take the only solution which lies in interval $[0,1]$ to get

$$x^* = \frac{2-\lambda-\sqrt{(8-7\lambda)\lambda}}{2(1-2\lambda)}$$

By directly substituting in the values $\underline{u}=\lambda=\frac{1}{4}$ we see that $\hat{x}=\frac{1}{2}$ and

$$x^* = 2-\frac{1}{4}-\sqrt{\left(8-\frac{7}{4}\right)\frac{1}{4}}=\frac{1}{2}$$

Therefore, at values $\underline{u} = \lambda = \frac{1}{4}$, the seller and the buyer pick an identical reviewer of toughness $\frac{1}{2}$. Note however that $\hat{x}$ is strictly increasing in $\underline{u}$, yet x^* is constant in $\underline{u}$. Therefore, for $\underline{u} < \frac{1}{4}$, there exist λ sufficiently close to $\underline{u}$ such that $\hat{x} < x^*$, as $\lim_{\lambda \to \underline{u}} \hat{x} = \frac{1}{2}$, but as shown in Proposition 2.5, x^* is strictly decreasing in λ. $\square$

Chapter 3

Costly Communication and Organisational Attention

3.1 Introduction

As we move further into the age of ubiquitous information gathering and data collection, it is clear that while information is abundant, not all of it is worthy of our attention. Indeed, as was highlighted by Simon (1973): "In a world of this kind, the scarce resource is not information; it is the processing capacity to attend to this information". Not only what we listen to, but who we listen to (and who listens to us), can greatly affect the decisions we make. The same is true inside of organisations, where the need to coordinate with others meets our capacity (or desire) to diligently process and interpret what they have to say. In this chapter, I examine how the allocation of attention within organisations impacts their ability to react in a coordinated manner to the ever-changing external environment.

This chapter follows the team theory approach of Marschak and Radner (1972), Prat (2002) and Dessein and Santos (2006), which abstracts from the incentive issues traditionally studied in the literature and views firms primarily as information processing entities. In this framework, organisations must process information about their competitive environment and then take actions which strike some balance between adaptation to external conditions and coordination of internal activities. Knowledge is dispersed throughout the organisation and agents must communicate with others in order to learn about local conditions and plans of action.

As absorbing information about the activities of those around us is costly (in that it consumes our attention), agents in the model allocate more attention to those who require it most. I extend the model of Dessein et al. (2013), who examine attention allocation in organisations, to the case of non-uniform task structures. In my model, the tasks of some agents will be broadly defined and will require coordination with many other individuals in the firm, whereas other tasks will be more specialised and insular. I look at how the scope of a given agent's task influences the amount of attention paid to them by others in the organisation.

I also examine how the span of control for the CEO of an organisation can influence their allocation of attention amongst members of the top management team. The idea that limited attention creates a trade off between information processing and coordination, particularly at higher levels of the organisation, was also highlighted by Simon (1973):

> "Attention is the chief bottleneck in organizational activity, and the bottleneck becomes narrower as we move to the tops of organizations, where parallel processing capacity becomes less easy to provide without damaging the coordinating function that is a prime responsibility of these levels."

Recent empirical findings by Bandiera et al. (2014), using a unique data set on CEO time use, have suggested some counter intuitive facts about the allocation of attention for CEOs in large firms. Their main finding is that CEOs with larger executive teams (more subordinates) tend to spend more time and attention interacting with others in the firm and focus more on coordinating activities. This suggests that the documented increase in the span of control of CEOs over the last two decades (Rajan and Wulf, 2006) has not been driven by a desire to free up attention for CEOs by delegating decision making.

This chapter therefore applies the insights of the model to examine how the composition of top management teams can affect the CEOs attention allocation. I proceed by first examining the existing literature on information processing and communication in organisations, in addition to more recent work on information transmission in social networks. Following this I present the main results of the model, before discussing the relationship to observations in the empirical literature.

3.2 Literature

This chapter spans two main areas of the literature: organisational economics and the economics of social networks. A number of papers also explore the interplay between organisations and networks, many of which have emerged due to recent developments in the latter literature. There are also some earlier contributions in the intersection of these two areas, going back at least to Marschak and Radner (1972), who analyse optimal information processing networks in their seminal work on ‘team theory’. These ideas are developed much later in Radner (1993), who examines why information processing and computation can best be done in hierarchies. This is an important contribution, but has become somewhat removed from the current literature on information processing in firms, as it is mainly concerned with centralised decision making (although decentralised information processing) and optimal networks for computation. Another early paper which is similar in spirit is Bolton and Dewatripont (1994), who also analyse the structure of optimal information processing and communication networks in firms. Again, decision making is centralised, and the organisation wishes to maximise the flow of information which can be processed from the external environment. Bolton and Dewatripont (1994) introduce increasing returns to scale, both in information processing and communication, showing again that hierarchical processing has an inherent benefit in permitting specialisation at lower levels, allowing for more information processing, which can then synthesised for use by agents at the top. The difference between my model and Bolton and Dewatripont (1994) lies in the reason for communication within the firm, as they assume it is to aggregate information, rather than coordinate actions. In their model a central agent makes all decisions and so coordination is not the main issue of concern.

A vast number of papers in organisational economics have examined the role which incomplete information plays in determining the organisation's balance between specialisation and coordination. Indeed, this has been seen as one of *the* central problems of organisations and has been discussed in the literature for several decades (e.g. Chandler (1962), Arrow (1974)). The trade off which incomplete information imposes on the organisation is the primary focus of Marschak and Radner (1972) and the voluminous literature which has arisen as a result of their work.

A recent contribution in that vein which is particularly relevant to the model at hand is that of Dessein and Santos (2006), which addresses the argument, originally put forward by Becker and Murphy (1992), that the division of labour is limited not by the extent of the market but primarily by the cost of coordination. They analyse the impact of task bundling on the adaptiveness of an organisation to the external environment, where specialisation increases the need to coordinate activities across complementary tasks. Since communication is noisy, centralised decision making in their model reduces the need for information sharing and increases the adaptability of the firm to the changing environment. This increase in coordination is then traded off against an exogenously given benefit from task specialisation. They point out that the trade off between coordination and specialisation is in fact more complicated than Becker and Murphy (1992) suggest, since an increased need for coordination may lead to less specialisation, but also to tasks becoming more routinised, and therefore less variable (and hence less communication is needed).

Closer still to the approach of this chapter is Dessein et al. (2013), who examine the role which attention constraints play in the design of organisational communication channels. In order to do so, the authors combine ideas present in the classical literature on bounded rationality in organisations with the approach developed in Dessein and Santos (2006). Limited attention is introduced in the form of a constraint on the time which agents can spend coordinating activities. This paper also appeals to ideas from finance and macroeconomics on 'rational inattention' and information theory (Sims (2003), Veldkamp (2011)). They take the task structure of the organisation as given and ask how the organisation should best devote its time to strike a balance between adaptability and coordination. They find that organisational focus arises endogenously as agents who seek to balance attention costs with benefits from coordination look to only a subset of 'leaders' in order to coordinate action. Too many focal hubs in their firm results in a lack coordination and reduces the ability of the organisation to jointly adapt their actions to changing local conditions.

This team theory literature abstracts from incentive problems within the firm, as do I, however much focus has been placed on these issues by others. An important contribution in the area of strategic communication in organisations is Dessein (2002) who investigates the role of communication in determining issues of integration and bundling of tasks when preferences are not aligned. Another example of work in this area is Alonso et al. (2008) who consider a situation where two divisions with differing objectives communicate strategically about locally observed states of the world. Aghion and Tirole (1997) discuss the impact of the allocation of decision making authority by a principal, where this allocation may influence the agent's incentives to collect relevant information. They show that delegation of decision making authority may encourage biased agents to collect more information than they would if authority lies with the principal.

My model aligns itself more with the costly communication side of the literature, such as Dewatripont and Tirole (2005) who analyse a model where agents choose communication efforts (at a cost) and also must exert effort in listening to messages sent to them. This chapter therefore develops the idea first present in Chapter 6 of March and Simon (1958), but later in Arrow (1974) and more prominently in Crémer et al. (2007), that the capacity for organisations to process and communicate information about the environment can be constrained by a lack of attention, effort or expertise.

The second strand of the literature to which this paper contributes is the recent work on information transmission in social networks. Examples of such work includes models of word of mouth communication and information aggregation in networks (Golub and Jackson, 2010), communication games with endogenous link strength (Bloch and Dutta, 2009), information acquisition and dissemination (Galeotti and Goyal, 2010), and the literature on cheap talk in networks, exemplified by Galeotti et al. (2013) and Hagenbach and Koessler (2010).

A paper which deals directly with the link between organisational communication and networks is Calvó-Armengol et al. (2014). This paper extends the ideas present in Crémer et al. (2007) and Dewatripont and Tirole (2005) to a network setting in which agents may invest in active and passive communication and must match their actions with their local information and the actions of others. They show that equilibrium actions and communication efforts are found to depend on a version of the individual's Bonacich centrality, building on earlier results regarding games of strategic complements in network settings due to Ballester et al. (2006). Another closely related model is that of Calvó-Armengol and Beltran (2009) who derive similar results but also use a communication protocol akin to the consensus building process studied in Golub and Jackson (2010) to characterise the properties of optimal firm communication networks.

Finally, this chapter relates to the models of endogenous information acquisition which have been studied by Hellwig and Veldkamp (2009) and Myatt and Wallace (2012). The central insight from Hellwig and Veldkamp (2009) is that when agents acquire information endogenously, complementarity in action leads to a complementarity in information acquisition. Knowing what others know becomes valuable as it helps to coordinate actions in the face of uncertainty. When actions are substitutes the opposite incentive arises, since we wish to have actions which are negatively correlated with the actions of others. A similar model is analysed in Myatt and Wallace (2012) where agents play a linear quadratic 'beauty contest' game with endogenous information acquisition. In their model, attention to different sources of information is costly and must be allocated to sources with higher intrinsic precisions. As in Hellwig and Veldkamp (2009), complementarity in action drives agents to select similar information sources. I now lay out the details of my model in the following section.

3.3 Model

The modelling approach taken in this chapter is inspired by the normal-quadratic team theory framework of Marschak and Radner (1972), which is predominant in the literature on organisational communication where the incentives of all players are aligned.[1] In particular, I extend the model of Dessein and Santos (2006) and Dessein et al. (2013) who consider a set of agents $N = \{1, \ldots, n\}$ who must individually carry out a task based on information which they process from the environment. In contrast to them, I assume that the production externalities between tasks are represented by an exogenous undirected network g, where g is as set of links or edges such that $(i, j) \in g$ implies that tasks i and j are linked in the network.

This network of interactions is given exogenously and is determined by which tasks exhibit synergies or externalities in the production function of the organisation. The realisation of this network can be viewed as the result of an earlier job design choice which the firm has made, so that in principle, tasks could be made more or less interdependent by altering who does what in the organisation. I choose to interpret tasks with many interactions as being very broadly defined and encompassing many separate activities and responsibilities (e.g. management positions), whereas tasks which have very few links should be viewed as being highly specialised.

Each agent $i \in N$ perfectly observes their local state of nature θ_i selecting an action $a_{ii} \in \mathbb{R}$ to minimise the loss function $(a_{ii} - \theta_i)^2$. Local states should be viewed as idiosyncratic conditions which affect the optimal action for agent $i \in N$. For example, the agent may be a marketing director who is deciding how much to spend for the coming month, and needs to tailor this decision to the behaviour of competitors or demand conditions.[2]

In addition to the local action a_i, each agent interacts with a subset of other agents $N_i \subseteq N \setminus \{i\}$. I define the neighbourhood of agent $i \in N$ to be $N_i = \{j \in N \mid (i, j) \in g\}$ and their degree to be $d_i(g) = |N_i|$. Production externalities within the firm mean that each agent must also select a coordinating action $a_{ij} \in \mathbb{R}$ for each agent $j \in N_i$ to minimise the loss function $(a_{ij} - \theta_j)^2$ for each $j \in N_i$ with whom he interacts. Continuing the previous example, if the marketing director decides to spend more than

[1] That is, in contrast to the case where different agents may have biases for certain actions to be taken by managers and communicate via cheap talk. For example, Alonso et al. (2008).

[2] For simplicity I assume that local states are independent, as is commonly done in the literature, e.g. Alonso et al. (2008), Calvó-Armengol et al. (2014), Dessein and Santos (2006)

usual this coming month then the purchasing manager must adjust his plans to take into account the anticipated change in output.

As pointed out by Dessein and Santos (2006) and Dessein et al. (2013), although other approaches in the team theory tradition and elsewhere (e.g. Morris and Shin (2002),Hellwig and Veldkamp (2009)) have used a single action rather than a vector of actions, this introduces a necessary trade off between coordination and adaptation. Even with perfect communication and unlimited attention, the organisation cannot be perfectly adaptable and perfectly coordinated unless all realisations of the local states are equal. Since I am interested in how the endogenous allocation of attention affects the adaptability and coordination of the organisation, this particular set up will provide a clearer picture of its influence.

Although each agent $i \in N$ perfectly observes their own local state θ_i , they only receive noisy signals about other local states. I denote the vector of local states by $\theta = (\theta_1, \ldots, \theta_n)$ and assume that agents have a common prior distribution over θ such that $\theta \sim \mathcal{N}(0, \Sigma_\theta)$ where Σ_θ is a diagonal covariance matrix with identical diagonal entries σ_θ^2. An agent i can receive a message m_{ij} from agent j, informing them about agent j's observation of their local state. This message takes the form $m_{ij} = \theta_j + \varepsilon_{ij}$ where the receiver's errors ε_{ij} are independent across messages and distributed normally with mean 0 and variance σ_{ij}^2. The correlation between m_{ij} and θ_j is denoted by ρ_{ij}.

As the earlier quote by Simon (1973) highlights, "attention is the chief bottleneck in organizational activity". Allocation of attention in this model will reduce the variance of the messages received by agents by lowering σ_{ij}^2. I assume that attention is costly and the agent must always trade off the benefits from listening closely to a given message versus its cost. We could also view message noise as being partially a result of problems in codifying information sent between individuals who have differing expertise, as discussed in Crémer et al. (2007). As the model of Crémer et al. (2007) argues, the ability to communicate accurate information about the state of the world must rely on a shared language between the two parties. When a sender has more expertise, then his understanding of the state of the world is more precise than a receiver who is unfamiliar with important terminology or concepts involved in another employee's role. Therefore, the assumption that the message is never perfectly understood can be thought of as arising from a combination of agents having differing areas of expertise while also being attention constrained.

For now, I will ignore the attention allocation decision of the agents and focus on their optimal actions a_i, contingent on the signals they receive. The team payoff function which agents maximise is given by

$$\pi\left(a_i, a_{-i}\right) = \bar{\pi} - \sum_{i=1}^{n}\left(\left(a_{ii} - \theta_i\right)^2 - \gamma \sum_{j \in N_i}\left(a_{ii} - a_{ji}\right)^2\right)$$

where the coefficient $\gamma > 0$ measures the degree to which coordination is important to the firm relative to adaptation. Letting $d_i\left(g\right)$ or simply d_i denote the degree of the agent (i.e.$d_i \equiv |N_i|$), best responses are then given by

$$a_{ii}^{*} = \frac{1}{1+\gamma d_i}\theta_i + \sum_{j \in N_i} \frac{\gamma}{1+\gamma d_i} \mathrm{E}\left[a_{ji} \mid m_i\right] \tag{3.1}$$

$$a_{ij}^{*} = \mathrm{E}\left[a_{jj} \mid m_i\right] \tag{3.2}$$

The best response conditions for this normal-quadratic set-up take the usual form, where a_{ii}^{*} is a weighted function of the local state and agent i's expectation of $j's$ coordinating action. The optimal coordinating action a_{ij}^{*} is simply agent i's best guess of agent j's local action, given the information they have available. Ideally, each agent wishes to match their action to their observed local state, yet we can see from (3.1) that when they have more ties to others in the organisation, then their action will have to take more of an account of what they think *other's believe* their state to be.

As in Calvó-Armengol et al. (2014), Dessein et al. (2013) and Myatt and Wallace (2012), I will restrict attention to equilibrium strategies which are linear in the signals received. If other players also use linear strategies, then due to the linearity of conditional expectations when states and signals are normally distributed, we can express the player's equilibrium strategy as linear functions of their signals.[3]

[3]This assumption is frequently made in normal quadratic team-theoretic models and while non-linear equilibria could exist there are currently no known counter-examples. This is discussed further in Myatt and Wallace (2008) and Myatt and Wallace (2012).

Agent i knows that j's coordinating action a_{ji} can only depend on the message m_{ji} which they receive, since all local states are independent. Moreover, agent i's expectation of the signal which j receives is simply θ_i, since the errors are mean zero and independent. Given this, equilibrium strategies therefore take the form

$$a_{ii}^* = w_{ii}\theta_i \tag{3.3}$$

$$a_{ij}^* = w_{ij}m_{ij} \tag{3.4}$$

The weights w_{ii} and w_{ij} can be seen as the responsiveness of the agent to the information they receive. Substituting the actions in (3.3) and (3.4) into the team payoff function gives the following expected payoff:

$$\bar{\pi} - \sum_{i=1}^{N} \underbrace{((1-w_{ii})^2\,\sigma_i^2}_{\text{Adaptation Cost}} + \gamma \sum_{j\in N_i} \underbrace{[(w_{jj}-w_{ij})^2\,\sigma_j^2}_{\text{Coordination Cost}} + \underbrace{w_{ij}^2\sigma_{ij}^2}_{\text{Communication Cost}}]) \tag{3.5}$$

The total cost of noisy communication for the organisation can be decomposed into three components. The first component is a cost of adaptation, where if w_{ii} is low then agent i does not respond to his signal of θ_i and task i is poorly aligned with the environment. The second cost results from poor coordination, such that if agent i is very responsive to his signal about θ_i but j does not respond to i's message about their signal (or vice versa), then their actions are uncoordinated. The final term is the cost incurred from the noisy messages sent from each i to their neighbours j. If the error variance of the signal is high, then when agent j puts more emphasis on i's signal, (i.e. responds more elastically to it), the costlier this becomes for the firm, since it is not an accurate signal of i's local state.

Attempting to solve for these coefficients involves substitution from (3.3) and (3.4) into (3.1) and (3.2), and are given in the following lemma

Lemma 3.1. *Optimal actions are given by*

$$a_{ii}^* = \frac{1}{1+\gamma\sum_{j\in N_i}(1-\rho_{ji}^2)}\theta_i \tag{3.6}$$

$$a_{ij}^* = \frac{\rho_{ij}^2}{1 + \gamma \sum_{k \in N_j} (1 - \rho_{kj}^2)} m_{ij} \tag{3.7}$$

We see that each agent is less adaptive to their own local state when the correlation between their state and the message received by others is low. Noisy communication induces inertia in the actions of each agent, as the uncertainty introduced will encourage them to select actions which are more closely aligned with the common prior of $\theta_i = 0$. As pointed out by Dessein et al. (2013), this resonates with the ideas present in March and Simon (1958), who noted two distinct forms of coordination within organisations: coordination by plan and coordination by feedback. Actions can be coordinated by having a commonly understood notion of what the usual state of affairs is (i.e. a common prior on θ_i), or conversely by constant communication and feedback.

Equation (3.6) also shows that if communication is imperfect, then an increase in d_i will lower the responsiveness of agent i to their private information. This occurs because a higher degree in the network (broader task assignment) means that we have a greater incentive to ensure that our action is more predictable. When d_i is high, the coordination cost associated with task i will increase, meaning that deviating too far from our prior expected action of 0 will lead to miscoordination. With more accurate messages (i.e. ρ_{ji}^2 close to 1) this effect is diminished, however, the fact that $w_{ji} \in (0, 1)$ means that other agents $j \in N_i$ always put some positive weight on their prior belief of a_i, which is $\mathrm{E}[\theta_i] = 0$.

As one might expect, (3.7) shows that agents are less responsive to the messages sent by others if these messages are noisier. However, (3.7) also demonstrates that when the sender of the message has more interactions (i.e. d_j is higher), then their message is more likely to be ignored. This again arises due to the noise in the communication process, since agent i who receives message m_{ij} from j will not be aware of the messages received by other agents $k \in N_j \setminus \{i\}$. Because k will not be fully responsive to their message m_{jk}, agent j's action will be less responsive to θ_j and so this renders the message m_{ji} less useful.

This highlights a central property of the model, that agents with broadly defined roles (i.e. many connections in the network of interactions) will find their private information less useful due to the uncertainty introduced by noisy communication. This means that agents with broader roles must select their actions more on the basis of the prior expectations of *others*, rather than their own observation of θ_i.

As is clear from (3.5), reducing the noise inherent in the communication process will benefit the organisation. When communication is less noisy, then each agent i can be more responsive to his local information, as he knows that others can understand his message clearly. This increases w_{ii} towards 1, thereby reducing the adaptation cost. In addition to lowering the adaptation and communication costs, more accurate communication lowers the coordination cost, as now agents will have the confidence to act more elastically to the messages they receive. Since the organisation would benefit from minimising the message variances as much as possible, I now consider the endogenous allocation of attention in the organisation as a means of achieving this.

3.3.1 Endogenous Attention

As discussed earlier, owing to the difference in expertise between employees, all messages will be subject to some misinterpretation and therefore noise. I now allow for agents to reduce the level of noise through allocation of effort or *attention*, as in Dewatripont and Tirole (2005) and Calvó-Armengol et al. (2014). In particular, I assume a simple linear cost of attention given by $c\left(\sum_{j\in N_i}\tau_{ij}\right)$, where τ_{ij} is the precision of the error ε_{ij} and therefore $\varepsilon_{ij} \sim \mathcal{N}\left(0, \frac{1}{\tau_{ij}}\right)$. The parameter c can be seen as representing the opportunity cost of attention spent on listening to messages, as opposed to devoting it to other responsibilities. Alternatively, we could suppose that each individual is attention constrained such that $\sum_{j\in N_i}\tau_{ij} \leq \kappa$ and that c is the marginal cost of hiring support staff to increase their capacity.

To avoid uninteresting cases where agents pay no attention to signals because it is too costly to do so, I assume that $c < \frac{\sigma_\theta^4}{1+\gamma d_{max}}$ where $d_{max} \equiv \max_{i\in N} d_i$. Each agent therefore selects a level of attention $\tau_{ij} > 0$ for each neighbour in order to maximise the payoff $\Pi = \pi - \sum_{i\in N} c\left(\sum_{j\in N_i}\tau_{ij}\right)$. To more easily work with the term π, I rewrite

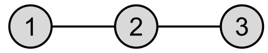


Figure 3.1: A Three Person Organisation

(3.5) first as

$$\bar{\pi} - \sum_{i=1}^{n} \left[\sigma_\theta^2 (1 - w_{ii})^2 + \gamma \sum_{j \in N_i} \left(w_{ii}^2 \, \mathrm{E} \left[(\theta_i - \mathrm{E}[\theta_i \mid m_{ji}])^2 \right] \right) \right]$$

Then, since $\mathrm{E}\left[(\mathrm{E}[\theta_i \mid m_{ji}] - \theta_i)^2 \right]$ is given by $\mathrm{Var}[\theta_i \mid m_{ji}] = \sigma_\theta^2 \left(1 - \rho_{ji}^2 \right)$, we can rewrite this as

$$\bar{\pi} - \sum_{i=1}^{n} \left[\sigma_\theta^2 (1 - w_{ii})^2 + \gamma \sum_{j \in N_i} w_{ii}^2 \sigma_\theta^2 \left(1 - \rho_{ji}^2 \right) \right]$$

Substituting in from (3.6) and (3.7), it is possible to express the total payoff in the following convenient form:

$$\pi = \bar{\pi} - \sum_{i=1}^{n} \sigma_\theta^2 (1 - w_{ii}) \tag{3.8}$$

The rewritten payoff above is similar to that found in Dessein et al. (2013) and shows that by allowing for the endogeneity of actions, the total payoff now depends only on the responsiveness of each task. I now consider the attention allocation decision in two archetypal organisational forms, the star and the tree. I concentrate on these forms as firstly it most easily highlights the key mechanism of the model, and secondly as it allows me to relate the results to some recent empirical work by Bandiera et al. (2014).

3.3.2 Endogenous Attention in the Star Network

I show in Figure 3.1 an example of a simple organisation in the case of $n = 3$ and $g = \{(1,2), (2,3)\}$. This could correspond to a situation where 1 and 3 are two separate divisions or local offices and 2 is a central office. Each observe separate local states and play actions in accordance with (3.6) and (3.7) in order to maximise Π.

Whether peripheral agents 1 and 3 should allocate more costly attention to the central agent 2, or vice versa, is not initially clear. The actions of 2 will incur the largest coordination cost for the organisation when they take actions very far from their prior of 0. Therefore, attention allocated to this agent will help to minimise the coordination cost. However, as is suggested from (3.6) and (3.7), the central agent's task should be less adaptable to the environment and so attention is better allocated elsewhere.

Proposition 3.1. *If the network of interactions is a star, then peripheral tasks receive the most attention per task and central tasks the least.*

Corollary 3.1. *If* $n = 3$ *and* $g = \{(1,2), (2,3)\}$ *then* $\tau_{12} < \tau_{21}$ *and* $\tau_{32} < \tau_{23}$.

Taking the example of the simple firm with three divisions, the centralised offices must allocate more attention to the local offices than vice versa, since their task will be more adaptable.

In the model of Dessein et al. (2013), ex-ante identical tasks can lead to organisational focus, where attention is concentrated on a select few 'leaders'. In their model, 'leaders' in the firm will be allowed to adapt more elastically to their local state, since more attention is paid to their messages and hence the rest of the organisation is better coordinated with their action. As Dessein et al. (2013) point out, this provides a justification for why firms may wish to focus on performing a small number of tasks well, without appealing to increasing returns to scale arguments. However, since the task structure is symmetric in their model, this means that their network of connections is a complete network, and any node can be selected as a 'leader'.

However, when the tasks are not ex-ante identical, then as shown in (3.6) and (3.7), agents with broader overarching roles in the organisation will exhibit more inertia in their actions than those with more narrowly defined roles (fewer links). The asymmetry in responsiveness to local information therefore provides an anchor for the attention allocation decision within the firm. Even though broader tasks have a greater need for coordination, it is precisely this which causes them to be made routine and therefore relatively unresponsive to changing external conditions. The inertia of central tasks has the effect of lowering the amount of attention paid to them, which then further lowers their responsiveness.

I now turn to how the optimal allocation of attention depends on the parameters of the problem. The effect of γ on the attention allocation decision is again not obvious. On the one hand, when coordination costs are high relative to adaptation, then we should expect relatively more intensive communication, since this enables actions to be better aligned. On the other hand, when coordination becomes relatively more important, it becomes optimal to make all tasks more routine and so communication is less valuable. As I show, the effect of γ on the optimal τ_{ij} is shown to be non-monotonic in γ.

Corollary 3.2. *Increasing γ causes the optimal τ_{ij} to increase if and only if*

$$\gamma < \left(\frac{\sigma_\theta}{2d_j\sqrt{c}}\right)^2$$

If γ is low, then higher coordination costs lead to more demand for communication. If γ is high then the second effect dominates, and an increased demand for routinisation leads to less attention being devoted to messages. As can also be seen, a higher variance of θ_i will *ceterus paribus* lead to more intensive communication, since the prior belief is now less dependable.

Attention Constraints

A difference between my approach and Dessein et al. (2013) is that they assume that each agent's total allocation of attention is constrained, as opposed to assuming a cost function.[4] Assuming briefly that each agent has a strict attention constraint such that $\sum_{j\in N_i}\tau_{ij} \leq \kappa$, I now show that again the central agent of the star is least adaptive to the local environment. Since the attention constraint always binds at any optimum, the central agent must receive the full attention of all peripheral nodes. Moreover, since there are decreasing returns in allocating attention, the central agent must divide his attention equally amongst all peripheral nodes. Letting the central node be denoted by

[4] Different approaches have been seen in the literature, for example Maćkowiak and Wiederholt (2009) and Myatt and Wallace (2012) assume costly attention, yet Sims (2003) and Van Nieuwerburgh and Veldkamp (2010) assume an attention constraint.

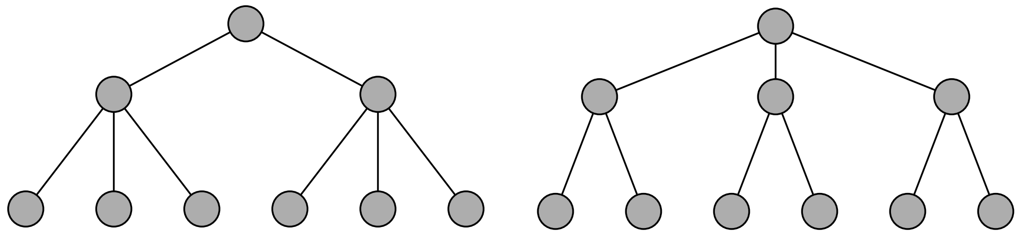


Figure 3.2: Tree Networks with Differing CEO Spans of Control

i and any peripheral node by j, this gives

$$w_{ii} = \frac{1}{1 + \gamma d_i \left(\frac{\tau_\theta}{\tau_\theta + \kappa} \right)} \qquad w_{jj} = \frac{1}{1 + \gamma \left(\frac{\tau_\theta}{\tau_\theta + \kappa / d_i} \right)}$$

It is easy to verify that by assuming $w_{ii} > w_{jj}$ we reach a contradiction. Therefore, the central agent, although by default receiving the most attention, has the least adaptable task and the insights of the previous subsection hold. The main point which is made by Dessein et al. (2013) is that attention and adaptability are complements. Indeed, this same mechanism drives the main result of this section, namely, that adaptable tasks command more attention and that tasks can be more adaptable if they have fewer associated costs of miscoordination.

3.3.3 Endogenous Attention and Span of Control

It is well documented in the literature that the composition of top management teams of leading firms has changed significantly in the period since the early 1980s, with the average CEO's span of control (number of managers who report directly to them) increasing dramatically. For example, Rajan and Wulf (2006) study the managerial structure of over 300 large US firms between 1986 and 1999, showing that the average CEO's number of direct reports increased from 4.4 to 7.2 in this period. A more recent study by Guadalupe et al. (2014) has shown that this trend has continued in subsequent years with the average number of direct reports increasing to 9.8 for a large sample Fortune 500 companies.

I now focus on recent empirical findings by Bandiera et al. (2014) in their study of

CEO time use in a sample of 65 large US firms. One explanation for the increasing size of the average executive team is that it results from delegation of responsibilities and a move towards decentralised decision making. However, the main conclusion of the authors is that the larger the executive team, the more time and attention CEOs tend to spend interacting with them. This suggests that the documented increase in the span of control of CEOs has not led to more delegation of tasks to members of the executive team (freeing up attention of the CEO) as might be assumed. On the contrary, they show that, particularly for large firms, CEOs spend more time dealing with internal activities and less time working alone or with outsiders.

In Figure 3.2, I give two examples where the network of interactions within the organisation takes the form of a tree. I refer to the individual at the top of the tree as the CEO of the organisation and the individuals at the second level as the directors. The leaf nodes which comprise the final level will be referred to as subordinates. In Organisation A, the CEO interacts with fewer directors, yet these directors have more broadly defined roles and therefore more subordinates. In Organisation B the CEO has a larger span of control, yet these directors are more specialised in their roles, hence having to coordinate their actions with fewer subordinates.

Treating all agents at the same level symmetrically and again assuming that the marginal cost of attention is sufficiently low, I find the following result:

Proposition 3.2. *The CEO devotes more attention per director in Organisation B than Organisation A*

The intuition for this result is similar to the previous case of the star network. By creating more directors, each individual director's role is more specialised and hence has less interactions with other tasks in the organisation. As a consequence, each director can afford to be more responsive to her local information, since her associated coordination cost is lower. In turn this leads the CEO to devote *more* attention to each director's messages, since their activities are less routinised.

This result therefore provides one possible explanation for the recent empirical findings of Bandiera et al. (2014) relating the CEOs span of control to their allocation of attention within the firm. The recent study of Guadalupe et al. (2014) also finds that the documented increase in the average CEO's span of control has resulted primarily from the growth in functional managers, who are often highly specialised (e.g. Directors of R&D, CMOs, CIOs) as opposed to more broadly interacting general managers. In my model, specialised directors are more adaptable to their local conditions and so require more attention from the CEO.

3.4 Conclusion

In this chapter I have examined how the allocation of costly attention within organisations influences their ability to adapt to the changing external environment and maintain internal coordination. I bring into focus the relationship between specialisation, coordination and attention, which was first highlighted by March and Simon (1958) over half a century ago.

I show that the actions of those in broadly defined roles which encompass many separate activities and responsibilities (such as managerial positions) are subject to inertia. This inertia arises due to the need for coordination, which forces these central agents to be more predictable in their actions. On the other hand, tasks which are more isolated and narrowly defined are concerned less with coordination and can afford to respond more elastically to their own local shocks. The endogenous allocation of costly *attention*, which lowers the cost of communication, compounds this effect. Individuals who are more generalised will be focused on less by other members of the organisation, since they have tasks which are more routinised. Consequently they must conform to the prior expectations of others by limiting their adaptiveness.

The model presented here is based on an extension of Dessein et al. (2013) to the case of non-uniform task structures. Although their model is similar in set-up to the one presented in this chapter, the questions which are asked differ. Their paper focuses on the question of why leadership emerges in firms, what determines how many leaders an organisation has, and what this says about the optimal size of the firm. To state that leaders are focused on due to their wide influence is almost tautological, so one of the key insights of their paper is to show that leadership can emerge, even with a completely uniform task structure and ex-ante identical agents.

I take a different approach and have assumed that division of labour within the firm has endowed some individuals with more interconnected tasks than others. This is represented in my model by an exogenous network of production externalities, which is predetermined by the job design decision of the organisation. Whilst Dessein et al. (2013) restrict attention to ex-ante identical tasks and allow the communication network to vary, I consider more general patterns of interaction while restricting communication to occur only along these channels. The definition of what constitutes a 'leader' differs in the two cases differs and so different conclusions are arrived at. On the one hand, Dessein et al. (2013) consider leaders to be focal points of the organisation whose messages are closely listened to and will be well coordinated with. On the other hand, I equate agents with high degrees in the network with individuals who have many separate activities and responsibilities, such as managers, directors or executives. If we take leaders to be those who interact most broadly with the organisation, then I show that the need for predictability limits the amount of attention placed on them by others. However the central insight of their paper (that attention and adaptiveness are complements) holds throughout.

With the work presented in this chapter, I have also highlighted a possible mechanism for the recent empirical observations of Bandiera et al. (2014). The main finding of Bandiera et al. (2014) is that CEOs with larger executive teams tend to spend more time and attention dealing with activities inside the firm. This effect is present in my model and results from the endogenous reallocation of attention following the growth of the executive team. When directions have fewer subordinates, their actions become more responsive to the outside environment, and so more attention must be placed on them.

An alternative explanation which is also in line with the insights of this model is that even if directors themselves are not more specialised, the process of delayering brings them 'closer to the product' and increases their focus on external conditions (and hence their adaptability).

As recently discussed by Garicano and Prat (2011), research by economists over the last several decades has placed a heavy focus on incentive problems within organisations. For that reason, there are several promising directions for research in organizational economics. One possible line of study which is directly related to this chapter, is the organisation's endogenous choice of job design within the firm and how this impacts its adaptability and internal allocation of attention. With the exception of Dessein and Santos (2006), this is an area which has been relatively under explored in the literature.

Appendix - Proofs

Proof of Lemma 3.1. Solving first for a_i^*, we must substitute from (3.3) in to (3.1), giving $w_{ii}\theta_i = \frac{1}{1+\gamma d_i}\theta_i + \sum_{j\neq i}\frac{\gamma}{1+\gamma d_i}\mathrm{E}\left[a_{ji} \mid m_i\right]$. Substituting in (3.4) we get

$$w_{ii}\theta_i = \frac{1}{1+\gamma d_i}\theta_i + \gamma\sum_{j\neq i}\frac{1}{1+\gamma d_i}w_{ji}\mathrm{E}\left[m_{ji} \mid m_i\right]$$

Since, $\mathrm{E}\left[m_{ji} \mid m_i\right] = \theta_i$, so this gives $w_{ii} = \frac{1}{1+\gamma d_i} + \gamma\sum_{j\neq i}\frac{1}{1+\gamma d_i}w_{ji}$. To find w_{ji} we use the standard results on conditional expectations of mean zero normally distributed random variables following normal signals to get

$$w_{ji}m_{ji} = w_{ii}\frac{\mathrm{Cov}\left[\theta_i, m_{ji}\right]}{\mathrm{Var}\left[m_{ji}\right]}m_{ji}$$

Substituting in to the above, therefore gives

$$w_{ii} = \frac{1}{1+\gamma\sum_{j\neq i}\left(1 - \frac{\mathrm{Cov}\left[\theta_i, m_{ji}\right]}{\mathrm{Var}\left[m_{ji}\right]}\right)}$$

Noting that in this case $\frac{\mathrm{Cov}\left[\theta_i, m_{ji}\right]}{\mathrm{Var}\left[m_{ji}\right]} = \rho_{ji}^2$ gives the equation in (3.6). Substituting (3.4) into (3.2) gives $w_{ij}m_{ij} = w_{jj}\rho_{ij}^2 m_{ij}$, yielding the equation in (3.7). □

Proof of Proposition 3.1. The payoff function which each individual maximises is given by $\pi_i = \bar{\pi} - \sum_{i=1}^{n}\sigma_\theta^2\left(1 - w_{ii}\right) - c\left(\sum_{j\in N_i}\tau_{ij}\right)$. Writing w_{jj} in terms of precisions we get

$$w_{jj} = \frac{1}{1+\gamma\sum_{k\in N_j}\frac{\tau_\theta}{\tau_\theta+\tau_{kj}}}$$

We can easily verify that this is increasing and strictly concave in τ_{ij} and that $\lim_{\tau_{ij}\to\infty}\frac{dw_{jj}}{d\tau_{ij}} = 0$. The linearity of the cost function then implies that the central node will allocate the same amount of attention to all peripheral nodes at any interior optimum. Peripheral nodes also choose identical attentions at any optimum since marginal benefit is strictly increasing in own attention τ_{ij} and marginal cost is constant.

First order conditions for central node i are found after differentiating π_i with respect to τ_{ij}:

$$\gamma \left(\tau_\theta(1+\gamma) + \tau_{ij}\right)^{-2} = c$$

Solving this for the central node's optimal attention τ_{ij} yields

$$\tau_{ij} = \sqrt{\frac{\gamma}{c}} - \tau_\theta(1+\gamma)$$

Similarly for the peripheral nodes, by imposing symmetry of actions we arrive at

$$\tau_{ji} = \sqrt{\frac{\gamma}{c}} - \tau_\theta(1+d_i\gamma) \tag{3.9}$$

Thus showing that attention allocated per task is decreasing in d_i, the broadness of that task. □

Proof of Proposition 3.2. For simplicity I normalise τ_θ to 1 without loss of generality and appeal to a version of Topkis's Theorem presented by Amir (2005). Let τ_{cd} denote the attention that the CEO pays to each director's message, and similarly for subordinates define τ_{sd}. Again assuming that c is sufficiently low to allow interior solutions, differentiating Π with respect to τ_{cd} gives

$$\frac{\partial \Pi}{\partial \tau_{cd}} = \gamma \left(1 + \tau_{cd} + \gamma\left(1 + |s_d|\right)\frac{1+\tau_{cd}}{1+\tau_{sd}}\right)^{-2} - c \tag{3.10}$$

It can be noted that $\frac{\partial^2 \Pi}{\partial \tau_{cd}\partial \tau_{sd}} > 0$ whereas $\frac{\partial^2 \Pi}{\partial \tau_{cd}\partial |s_d|} < 0$. Differentiating Π with respect to τ_{sd} gives

$$\frac{\partial \Pi}{\partial \tau_{sd}} = \gamma \left(1 + \tau_{sd} + \gamma\left|s_d\right| + \gamma\frac{1+\tau_{sd}}{1+\tau_{cd}}\right)^{-2} - c$$

It can be noted that $\frac{\partial^2 \Pi}{\partial \tau_{sd}\partial |s_d|} < 0$. This shows that Π is supermodular in the attentions of the CEO and subordinates, and moreover exhibits decreasing differences in $-\left|s_d\right|$. The statement then follows from Theorem 10 of Amir (2005) after applying appropriate bounds to the action space. This implies that the optimal τ_{cd} and τ_{sd} are strictly decreasing in $|s_d|$. $\square$

Bibliography

Aghion, P. and J. Tirole: 1997, 'Formal and Real Authority in Organizations'. *Journal of Political Economy* **105**(1), 1–29.

Alonso, R. and O. Câmara: 2014, 'Bayesian Persuasion with Heterogeneous Priors'.

Alonso, R., W. Dessein, and N. Matouschek: 2008, 'When Does Coordination Require Centralization?'. *American Economic Review* **98**(1), 145–179.

Amir, R.: 2005, 'Supermodularity and Complementarity in Economics: an Elementary Survey'. *Southern Economic Journal* **71**(3), 636–660.

Anderson, S. P., J. K. Goeree, and C. A. Holt: 2001, 'Minimum-Effort Coordination Games: Stochastic Potential and Logit Equilibrium'. *Games and Economic Behavior* **34**(2), 177–199.

Arrow, K. J.: 1974, *The Limits of Organization*. W.W. Norton & Co Inc.

Bader, G. and C. Hogue: 2003, 'An Automated Method for Finding Molecular Complexes in Large Protein Interaction Networks'. *BMC Bioinformatics* **27**, 1–27.

Bagwell, K.: 2007, 'The Economic Analysis of Advertising'. *Handbook of Industrial Organization* pp. 84–104.

Ballester, C. and A. Calvó-Armengol: 2010, 'Interactions with Hidden Complementarities'. *Regional Science and Urban Economics* **40**(6), 397–406.

Ballester, C., A. Calvó-Armengol, and Y. Zenou: 2006, 'Who's Who in Networks. Wanted : The Key Player'. *Econometrica* **74**(5), 1403–1417.

Bandiera, O., A. Prat, R. Sadun, and J. Wulf: 2014, 'Span of Control and Span of Attention'.

Baur, M., U. Brandes, M. Gaertler, and D. Wagner: 2004, 'Drawing the AS graph in 2.5 dimensions'. In: J. Pach (ed.): *Graph Drawing*, No. Lncs 3383. Springer, pp. 43–48.

Becker, G. and K. Murphy: 1992, 'The Division of Labor, Coordination Costs, and Knowledge'. *The Quarterly Journal of Economics* **117**(4).

Bergemann, D. and M. Pesendorfer: 2007, 'Information Structures in Optimal Auctions'. *Journal of Economic Theory* **137**(1), 580–609.

Bhawalkar, K., J. Kleinberg, K. Lewi, T. Roughgarden, and A. Sharma: 2012, 'Preventing Unraveling in Social Networks: The Anchored k-core Problem'. *39th International Colloquium, ICALP 2012* pp. 440–451.

Bloch, F. and B. Dutta: 2009, 'Communication Networks with Endogenous Link Strength'. *Games and Economic Behavior* **66**(1), 39–56.

Blume, L.: 1993, 'The Statistical Mechanics of Strategic Interaction'. *Games and Economic Behavior* **5**, 387–424.

Bolton, P. and M. Dewatripont: 1994, 'The Firm as a Communication Network'. *The Quarterly Journal of Economics* **109**(4), 809–839.

Bonacich, P.: 1987, 'Power and Centrality: A Family of Measures'. *American Journal of Sociology* **92**(5), 1170–1182.

Bramoullé, Y. and R. Kranton: 2007, 'Public Goods in Networks'. *Journal of Economic Theory* **135**, 478–494.

Brualdi, R. A.: 2011, *The Mutually Beneficial Relationship of Graphs and Matrices*. American Mathematical Society.

Calvó-Armengol, A. and J. d. M. Beltran: 2009, 'Information Gathering in Organizations: Equilibrium, Welfare, and Optimal Network Structure'. *Journal of the European Economic Association* **7**(1), 116–161.

Calvó-Armengol, A., J. D. Martí, and A. Prat: 2014, 'Communication and Influence'.

Chandler, A. D.: 1962, *Strategy and Structure*. Cambridge, MA: MIT Press.

Chen, R. and Y. Chen: 2011, 'The Potential of Social Identity for Equilibrium Selection'. *The American Economic Review* **101**(October), 2562–2589.

Chwe, M. S.-Y.: 2000, 'Communication and Coordination in Social Networks'. *Review of Economic Studies* **67**, 1–16.

Crawford, V. P.: 1991, 'An "Evolutionary" Interpretation of Van Huyck, Battalio, and Beil's Experimental Results on Coordination'. *Games and Economic Behavior* **3**(1), 25–59.

Crémer, J., L. Garicano, and A. Prat: 2007, 'Language and the Theory of the Firm'. *The Quarterly Journal of Economics* (February), 373–408.

Dessein, W.: 2002, 'Authority and Communication in Organizations'. *Review of Economic Studies* **69**(4), 811–838.

Dessein, W., A. Galeotti, and T. Santos: 2013, 'Rational Inattention and Organizational Focus'.

Dessein, W. and T. Santos: 2006, 'Adaptive Organizations'. *Journal of Political Economy* **114**(5), 956–995.

Dewatripont, M. and J. Tirole: 2005, 'Modes of Communication'. *Journal of Political Economy* **113**(6), 1217–1238.

Eeckhoudt, L. and C. Gollier: 1995, 'Demand for Risky Assets and the Monotone Probability Ratio Order'. *Journal of Risk and Uncertainty* **11**(2), 113–122.

Ellison, G.: 1993, 'Learning, Local Interaction and Coordination'. *Econometrica* **61**(5), 1047–1071.

Ely, J., A. Frankel, and E. Kamenica: 2014, 'Suspense and Surprise'.

Galeotti, A., C. Ghiglino, and F. Squintani: 2013, 'Strategic Information Transmission Networks'. *Journal of Economic Theory* **148**(5), 1751–1769.

Galeotti, A. and S. Goyal: 2010, 'The Law of the Few'. *American Economic Review* **100**(September), 1468–1492.

Galeotti, A., S. Goyal, M. O. Jackson, F. Vega-Redondo, and L. Yariv: 2010, 'Network Games'. *Review of Economic Studies* **77**(1), 218–244.

Garicano, L. and A. Prat: 2011, 'Organizational Economics with Cognitive Costs'. *CEPR Discussion Paper 8372*.

Gentzkow, B. M. and E. Kamenica: 2014, 'Costly Persuasion'. *American Economic Review* **104**(5), 457–462.

Gentzkow, M. and E. Kamenica: 2012, 'Competition in Persuasion'. (September), 1–36.

Gill, D. and D. Sgroi: 2008, 'Sequential Decisions with Tests'. *Games and Economic Behavior* **63**(2), 663–678.

Gill, D. and D. Sgroi: 2012, 'The Optimal Choice of Pre-Launch Reviewer'. *Journal of Economic Theory* **147**(3), 1247–1260.

Goeree, J. and C. Holt: 2005, 'An Experimental Study of Costly Coordination'. *Games and Economic Behavior* **51**(2), 349–364.

Goldberg, A.: 1984, 'Finding a Maximum Density Subgraph'.

Golub, B. and M. O. Jackson: 2010, 'Naïve Learning in Social Networks and the Wisdom of Crowds'. *American Economic Journal: Microeconomics* **2**(1), 112–149.

Granovetter, M.: 1978, 'Threshold Models of Collective Behavior'. *The American Journal of Sociology* **83**(6), 1420–1443.

Guadalupe, M., H. Li, and J. Wulf: 2014, 'Who Lives in the C-Suite ? Organizational Structure and the Division of Labor in Top Management'. *Management Science* **60**(4), 824–844.

Hagenbach, J. and F. Koessler: 2010, 'Strategic Communication Networks'. *Review of Economic Studies* **77**(3), 1072–1099.

Hellwig, C. and L. Veldkamp: 2009, 'Knowing What Others Know: Coordination Motives in Information Acquisition'. *Review of Economic Studies* **76**(1), 223–251.

Hvide, H.: 2009, 'Oligopolistic Certification'. *The BE Journal of Theoretical Economics* **9**(1), 1–19.

Jackson, M. O.: 2008, *Social and Economic Networks*. Princeton University Press.

Jackson, M. O. and A. Watts: 2002, 'On the Formation of Interaction Networks in Social Coordination Games'. *Games and Economic Behavior* **41**(2), 265–291.

Jackson, M. O. and Y. Zenou: 2012, 'Games on Networks'.

Johnson, J. P. and D. P. Myatt: 2006, 'On the Simple Economics of Advertising, Marketing, and Product Design'. *The American Economic Review* **96**(3), 756–784.

Kamenica, E. and M. Gentzkow: 2011, 'Bayesian Persuasion'. *American Economic Review* **101**(6), 2590–2615.

Kolotilin, A.: 2014, 'Optimal Information Disclosure: Quantity vs. Quality'.

Lerner, J. and J. Tirole: 2006, 'A Model of Forum Shopping'. *The American Economic Review* **96**(4), 1091–1113.

Lewis, T. R. and D. E. Sappington: 1994, 'Supplying Information to Facilitate Price Discrimination'. *International Economic Review* **35**(2), 309–327.

Maćkowiak, B. and M. Wiederholt: 2009, 'Optimal Sticky Prices Under Rational Inattention'. *American Economic Review* **99**(3), 769–803.

Mailath, G., M. Okuno-Fujiwara, and A. Postlewaite: 1993, 'Belief-Based Refinements in Signalling Games'. *Journal of Economic Theory* **60**, 241–276.

March, J. G. and H. A. Simon: 1958, *Organizations*. London: John Wiley & Sons, Inc.

Marschak, J. and R. Radner: 1972, *Economic Theory of Teams*. New Haven: Yale University Press.

Milgrom, P. and J. Roberts: 1990, 'Rationalizability, Learning and Equilibrium in Games with Strategic Complementarities'. *Econometrica* **58**(6), 1255–1277.

Milgrom, P. R.: 1981, 'Good News and Bad News: Representation Theorems and Applications'. *The Bell Journal of Economics* **12**(2), 380–391.

Milgrom, P. R. and J. Roberts: 1986, 'Relying on the Information of Interested Parties'. *The RAND Journal of Economics* **17**(1), 18–32.

Milgrom, P. R. and R. J. Weber: 1982, 'A Theory of Auctions and Competitive Bidding'. *Econometrica* **50**(5), 1089–1122.

Monderer, D. and L. S. Shapley: 1996, 'Potential Games'. *Games and Economic Behavior* **143**, 124–143.

Morris, S.: 2000, 'Contagion'. *Review of Economic Studies* **67**, 57–78.

Morris, S. and H. Shin: 2002, 'Social Value of Public Information'. *The American Economic Review* **92**(5), 1521–1534.

Myatt, D. and C. Wallace: 2008, 'On the Sources and Value of Information: Public Announcements and Macroeconomic Performance'.

Myatt, D. P. and C. Wallace: 2012, 'Endogenous Information Acquisition in Coordination Games'. *The Review of Economic Studies* **79**(1), 340–374.

Ottaviani, M. and A. Prat: 2001, 'The Value of Public Information in Monopoly'. *Econometrica* **69**(6), 1673–1683.

Perez-Richet, E.: 2014, 'Interim Bayesian Persuasion: First Steps'. *American Economic Review* **104**(5), 469–474.

Prat, A.: 2002, 'Should a Team be Homogeneous?'. *European Economic Review* **46**(7), 1187–1207.

Radner, R.: 1993, 'The Organization of Decentralized Information Processing'. *Econometrica* **61**(5), 1109–1146.

Rajan, R. G. and J. Wulf: 2006, 'The Flattening Firm: Evidence from Panel Data on the Changing Nature of Corporate Hierarchies'. *Review of Economics and Statistics* **88**(4), 759–773.

Rayo, L. and I. Segal: 2010, 'Optimal Information Disclosure'. *Journal of Political Economy* **118**(5), 949–987.

Sandholm, W. H.: 2010, *Population Games and Evolutionary Dyanmics*. MIT Press.

Schelling, T. C.: 1973, 'Hockey Helmets, Concealed Weapons, and Daylight Saving: A Study of Binary Choices with Externalities'. *The Journal of Conflict Resolution* **17**(3), 381–428.

Seidman, S. B.: 1983, 'Network Structure and Minimum Degree'. *Social Networks* **5**, 269–287.

Sgroi, D.: 2002, 'Optimizing Information in the Herd: Guinea pigs, Profits, and Welfare'. *Games and Economic Behavior* **39**, 137–166.

Simon, H. A.: 1973, 'Applying Information Technology to Organization Design'. *Public Administration Review* **33**(3), 268–278.

Sims, C. A.: 2003, 'Implications of Rational Inattention'. *Journal of Monetary Economics* **50**(3), 665–690.

Topkis, D. M.: 1998, *Supermodularity and Complementarity*. Princeton University Press.

Van Huyck, J. B., R. C. Battalio, and R. O. Beil: 1990, 'Tacit Coordination Games, Strategic Uncertainty, and Coordination Failure'. *The American Economic Review* **80**(1), 234–248.

Van Nieuwerburgh, S. and L. Veldkamp: 2010, 'Information Acquisition and Under-Diversification'. *Review of Economic Studies* **77**(2), 779–805.

Veldkamp, L. L.: 2011, *Information Choice in Macroeconomics and Finance*. Oxford, UK: Princeton University Press.

Wasserman, S. and K. Faust: 1994, *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.

Wuchty, S. and E. Almaas: 2005, 'Peeling the Yeast Protein Network'. *Proteomics* **5**(2), 444–9.

Young, H. P.: 1998, *Individual Strategy and Social Structure*. Princeton University Press.