Restoring Trust in Finance: 
from Principal-Agent to Principled Agent*

Gordon Menzies**, Thomas Simpson***, Donald Hay****, and David Vines*****

Abstract

We outline a narrative of how attempts to solve the principal-agent problem for financial managers have eroded moral restraint, leading to fewer principled agents. Bonus-based compensation inspired by Jensen and Meckling (1976) appears to have contributed to unfavourable attitudes, through motivational crowding out (Simpson, 2016). We classify the moral restraint of earlier times as either ‘moral optimization’—standard utility theory appended with other-regarding preferences (Becker, 1981)—or as ‘moral prioritization’—a commitment to not doing wrong (Sen, 1976). Disciplining unethical managers in a post deregulation world by competition policy runs into serious practical difficulties, rendering it necessary to address their moral motivations. In contrast, trustworthiness sustains trust.

Keywords: Bank Bonuses, Competition, Financial Crisis, Regulation

* The authors thank without implication the Oxford Martin School and the Political Economy of Financial Markets (PEFM) group at St Antony’s College, Oxford. Gordon Menzies acknowledges the financial support of both institutions with gratitude. We also thank without implication Peter Anstey, Adam Bennett, Geoff Brennan, Peter Docherty, Peter Eckley, Charles Enoch (and other PEFM seminar participants), Sam Filby, Natalie Gold, Ian Goldin, Colin Mayer, Nick Morris, Avner Offer, Paul Oslington and H Peyton Young.

All correspondence to be sent to gordon.menzies@uts.edu.au at the University of Technology Sydney, PO Box 123, Broadway, Sydney, Australia.

** University of Technology Sydney
*** Blavatnik School of Government, University of Oxford
**** University of Oxford
***** Economics Department, Balliol College, St Antony’s College, and Institute for New Economic Thinking (INET) at the Oxford Martin School, University of Oxford; Crawford School of Public Policy, Australian National University; and Centre for Economic Policy Research
1 Introduction

Since the 2008 financial crisis, there has been a renewed interest in the practical ethics of finance sector participants. In one diagnosis, regard for customers was crowded out by bonus-based incentivization (Simpson 2016). Bonus-based incentivization aligned the interests of shareholders as principals with the managers as agents (Jensen and Meckling 1976). The lack of moral restraint in market participants’ behaviour in the lead-up to the crisis is consistent with economic experiments on the erosion of social preferences by financial incentives (Bowles 2016). This has been a longstanding concern (e.g., Durkheim 1915, Titmuss 1970, Williams 1973, Sen 1976, Goodin 1982).

Motivation crowding out is dangerous in finance for three complementary reasons. First, finance as an industry particularly relies on trust, and so trustworthiness is correspondingly important. Second, many people in finance receive an economics training, and experiments show that other-regard is in short supply for those with this training (Frank et al 1993, Frank & Schulze 2000, Frey & Meier 2003, Rubinstein 2006, Bauman and Rose 2011, and Ruske 2015). Combining these two points, we might say there is excess demand for trustworthiness in the finance industry. Third, the remedy of competition policy, which ‘economizes on virtue’ by forcing firms to act for the benefit of customers (Brennan and Hamlin 1996), is relatively challenging to implement.

In this paper we represent moral restraint analytically to show what is lost during motivation crowding out. We develop the notion of principled agents who at times exhibit a high degree of other-regard in standard economic calculations, and at other times substitute a moral principle for an economic calculation.1 Solving Jensen and Meckling’s (1976) principal-agent problem is socially valuable. But putative solutions that drive out principled agents from the marketplace do not alleviate motivation crowding out. They worsen the problem.

We call the process of deliberating about how to act, when an individual has other-regarding preferences, ‘moral optimization’. The classic framework is found in the modelling of altruists in the Economics of the Family (Becker, 1981). As an illustration, we might consider the problem of how a professional determines a reasonable fee for their services. The egoist sets the fee at the level that will maximise their income, extracting the maximum possible fee from the client. The altruistic professional cares not only about what they receive, but also what the client receives. In practice, this takes the form of a discount deducted from the maximum fee. We call this a case of moral optimization, for the task for the altruistic professional is to set the fee at a level that optimizes preference satisfaction, where his or her preferences include regard for the client. Moral optimization is not a deep challenge to standard economic models of the agent. It is a moralized form of cost benefit analysis, where the components of the analysis include shared interests or empathy for others. Obeying the dictates of cost benefit analysis implies optimization.2

A deeper challenge to standard economic models of the agent derives from commitments not to do wrong. We call the process of deliberating how to act in accordance with such commitments, independently of whether the action is utility maximising, ‘moral prioritization’. Moral prioritization sometimes requires an agent to forego opportunities to increase his own welfare. It is not the same as moral optimization. Consider truth-telling. A moralized cost benefit analysis may conceivably recommend an optimal amount of deceit, just if the benefits to me, or those I love, are high enough.

---

1 Besely (2007) also canvasses agents acting in more than one mode, though with a focus on government (as a force for public good and as a slave to private interests).

2 For any stable neoclassical optimum that equates two terms at the margin, a perturbation away from equilibrium incurs a cost greater than the benefit, by the definition of an optimum. Thus any approach to equilibrium from disequilibrium obeys the dictates of cost-benefit analysis—that one ought to adjust the optimized quantities so that the benefits of any adjustment exceed the cost.
But generally, that is not how decisions about lying are made. Individuals act according to the principle: ‘You should not lie!’ The principle trumps any evaluation of costs and benefits. The fact that some people do not act according to the moral principle does not count against the phenomenon; the point is that many do. Moral prioritization is a principled eschewing of cost benefit analysis, even when its components include shared interests and empathy. Deceit is not the only thing for which the preamble ‘an optimal amount of…’ rings false. Principled people also reject, say, an optimal amount of workplace violence.

The paper proposes a way to model moral prioritization. The notion of a commitment has a long heritage, since Amartya Sen’s classic ‘Rational Fools’ (1977). After arguing for the reality of commitments, the task he identifies is how to model them. He makes provisional suggestions on how to introduce appropriate ‘structure’ in the model of an agent’s deliberations in that article, and has revised the proposal since (1977: 335-41; 1997; 2005; for interpretative overview of his work, see Cudd 2014). The debate on how to do so continues, albeit at the margins of mainstream economics (see, e.g., papers collected in Peter and Schmid 2007, Herfeld 2009, Menzies and Hay 2012). Rather than developing a model of the process by which moral prioritization happens, we instead propose a way of ‘pricing’ the welfarist consequences of its occurrence. This respects the phenomenon—which is not only a fact, but one we regard as socially valuable—while remaining agnostic about how it occurs. To that degree it is less committal than other approaches, and can be endorsed by a wider range of theorists. In addition to its substantive contribution, then, the paper also makes a methodological contribution.

Our notion of a principled agent overturns an orthodoxy about when to assume self-seeking behavior. Economists who see limits for utilitarian calculus have tended to demarcate certain types of human activities as domains where there is no other-regarding action, or only marginal amounts. Examples are market transactions (Mill 1843) or war (Edgeworth 1881). We claim that there can be limits to utilitarian economic analysis, allowing for other-regard, right at the heart of a money-making endeavor like banking.

Our paper is organized as follows. In section 2 we provide background into the public backlash against the finance industry. In section 3 we contrast this highly incentivized world of recent financial markets with the situation in the UK prior to deregulation. In section 4 we analytically represent moral restraint. Drawing inspiration from the phenomenon of motivation crowding out in finance, we take as a stylized example a monopoly bank that chooses to operate at the competitive equilibrium, either as a matter of regard for consumers (moral optimization) or as a principled stand against the exploitation of market power (moral prioritization). In section 5 we show that competition policy is problematic for finance, and so in section 6 we canvass paths to the professionalization of finance.

2 How Bonuses Undermine Moral Restraint

A striking experiment strongly indicates that bankers do take a permissive view of moral restraints, at least when thinking in terms of their professional identity. Cohn et al. (2014) gave over one hundred bankers a coin flipping task, and were rewarded for the toss outcomes they reported. Subjects were given $20 for each ‘correct’ toss out of ten tosses, giving a range of payoffs from zero (no correct tosses) to $200 (ten correct tosses). The subjects knew which tosses would be deemed correct in

---

3 An observation by Bernard Williams implies—if correct—that it may be impossible in principle to model moral prioritization. He proposes that ‘the unthinkable’ is itself a moral category. ‘Entertaining certain alternatives, regarding them indeed as alternatives, is itself something that [the moral individual] regards as dishonourable or morally absurd’ (Williams 1973, 92). If so, even a lexicographical ordering of preferences would fail accurately to model moral prioritization. At least, it would fail to do so in a way that represented normatively defensible reasoning, rather than mere behavioral conformity.
advance. In this set-up, the experimentalist is a principal who asks for a truthful reporting of the tosses, and the subject is an agent who, we may presume, has a moral obligation to tell the truth. As in a classic principal-agent setup, there is hidden action. The experimental subjects flip the coin out of sight. No individual’s deceit can be detected.

Prior to the coin task, a control group was asked questions about the use of their leisure time and their hobbies, priming them to think in terms of their domestic identity. The treatment group was asked about their work life as bankers, priming them with their professional identity. In their chart below ‘a’ is the control group and ‘b’ is the treatment group. The blue Binomial distribution bars represent the expected frequencies of payoffs if all tosses are reported truthfully, and the red bars are the findings. Although an individual’s deceit cannot be detected, deceit across a group can.

**Figure 1: The trustworthiness of bankers**

![Graph showing trustworthiness of bankers](image)

When primed to think of their professional identity, the bankers as a group reported on average too many financially rewarding tosses. They were honest when focused on their leisure time. The experiment was repeated with other employment categories, including manufacturing, pharmaceuticals, telecommunications and information technology. No significant increase in dishonesty in the professional identity treatment was identified.

What explains this finding? There are two lines of explanation.

The first is a selection effect. Training in economics correlates strongly with someone being likely to seek personal gain over cooperation. The classic discussion is by Robert Frank, Thomas Gilovich & Denis T. Regan (1993). They survey a series of experiments with economics and non-economic undergraduates: a public goods game; prisoners’ dilemma; Ultimatum game, and an honesty test. On each, economists are less likely than a general sample to interact cooperatively. Corroborating studies include Frank & Schulze (2000), Frey & Meier (2003), Rubinstein (2006), Bauman and Rose (2011), Ruske (2015). The finding is sufficiently robust that a subordinate literature addresses the question of the causal direction of the correlation: does economics training make people selfish, or do selfish people choose to train in economics? The verdict is: both (see Cipriani et al 2009, Haucap & Just 2010, Bauman & Rose 2011, Etzioni 2015). Indeed, the causal effects are likely to be mutually reinforcing. More generally, so long as people who go into banking are significantly more likely to have an economics training than the general population, so egoist preferences will be more prevalent in the sector.

---

4 The tail ends of the distribution in the left panel show a small amount of cheating. The mean of the distribution was not significantly different from 0.5, however.
While the selection effect certainly exists, it is likely to explain the lack of moral restraint in banking only in part. Selection is unlikely to explain why bankers take a permissive view of moral restraints when primed to think in professional terms, but not when primed to think in domestic terms. A simple correlation between economics training and selfish preferences does not account for this, because egoist preferences are likely to be stable across domain. An explanation that is specific to the culture of banking is also needed.

A second explanation finds this in the direct and indirect effects of bonuses. We include in this category all forms of performance-related pay—including cash, stock, restricted stock, or options on the bank’s shares. Bonuses are significant because they are a direct source of ‘motivational crowding out’ for those receiving them, who may in turn indirectly create a moral culture for other employees.5

While the phenomenon of motivation crowding out is a contingent psychological fact, it is a fact nonetheless. The evidence for the effect is well established and we survey it briefly. The classic illustration is the study of six day-care centres in Haifa. On the introduction of a fine for parents who were late in picking up their children, the surprising result was that lateness increased, more than doubling. The effect remained after the fine was withdrawn (Gneezy and Rustichini 2000). A large-scale study played a variant on the Dictator game in fifteen societies. In the standard Dictator game, A decides how much, if any, of an initial endowment to transfer to B. Contrary to the prediction if homo economicus were to play, non-trivial offers are often made. The variant establishes a third-party, C, who may punish A by imposing a fine, say, if she decides the transfer is too low. On a simple view, the variant should only increase the mean level of offers by the dictator A. But this is not what is observed. Across the fifteen societies, there were increases in only two. Nine were unaffected, and in four the offers were significantly lower (Barr et al. 2009).

The experiments show that introducing a financial incentive does not have a predictable, linear effect on behaviour. On a simple view, those who were inclined to act fairly will do so regardless, and some of those who were not so inclined should be motivated by the new incentive to do so. This is not what is observed. Some of those who would previously act for broadly moralised reasons now do so for self-interested reasons. The gross effect may then be reduction in cooperative behaviour. Moral reasons for action are ‘crowded out’. The broad explanation is that the introduction of the incentive re-frames the interaction, for participants, from one structured by mutual moral expectation, to a transactional exchange, governed by self-interest. Policies intended to increase the rates of cooperative behaviour by those who are self-interested then often have an unintended, perverse effect. One condition under which crowding out occurs is when incentives signal distrust (Fehr and Rockenbach 2003, Sliwka 2007). Another is when they frame an activity as not subject to moral norms (Hoffman et al. 1994; Irlenbusch and Sliwka 2005; Cardenas et al. 2000; Gneezy and Rustichini 2000).


Motivation crowding-out is a highly plausible explanation of why bankers show a permissive attitude to moral requirements like telling the truth and avoiding deceit, and in the context of their professional

5 We later describe how ‘markets divisions’ in post-deregulation UK banks, managed by people who had begun their careers as traders, came to dominate the boards, management committees and culture of their banks.
but not domestic identity. High-powered incentives have undermined bankers’ moral motivations for acting, leading them to act on the basis of material self-interest, constrained by the law at best. Bonuses frame banking as conducted only for the money. They signal to bankers that their job is to maximise profit. Moral considerations on how to act, notably around obligations to the client, are implicitly downgraded as secondary or irrelevant. Bonuses also signal distrust. They carry the information that employers do not expect bankers to maximise the bank’s profit unless doing so results in their individual profit too. For if they did so trust bankers, then employers would not feel the need to offer large bonus packages. But they do offer them, so the relevant expectation must be missing. In both cases, the greater the bonus-to-salary ratio, the stronger the signal. The response from bankers is to view their work in transactional terms only, where the objective is to maximize income.

3 The Club and its Demise

The explanation above is primarily psychological. It is supported by a corroborating body of evidence from history. While longitudinal data is not available, changes in the culture of the City of London before and after the ‘Big Bang’ are widely attested to and consistent with the above.

For more than a century after the ‘panic of 1866’ (Bagehot, 1866), the stability of the banking system had never been in question (Offer, 2014) and British banking for most of the 20th Century was not marked by adventurous attitudes to risk and truthfulness. During the post-war construction of the British welfare state, financial markets were strictly regulated and international movements of financial capital were limited. The financial sector was highly fragmented, with participants being vetted to ensure they were deemed ‘fit and proper’ to carry out their functions. Individuals, firms, and partnerships not so deemed were dealt with by their peers and in extreme cases were excluded from the markets and from the social and professional networks of the Club.

The banking community at the time operated largely by self-regulatory agreement, but with some legal underpinning. The only institutions which engaged in complex or risky transactions were the merchant/investment banks and other specialist brokers and traders. They too were careful as they were taking risks mostly with their own funds given the partnership arrangements. Investment bankers depended very much on their reputation, which was developed through long-term relationships with clients and other counterparties within the City (Armstrong 2012).

Professional standards were maintained partly by the firms themselves and partly by the training and qualification of various types of functionary. Industry bodies, such as the Chartered Institute of Bankers in Scotland, established in 1875, maintained a variety of qualifications for bankers. The Bank of Scotland, for example, had a ‘Superintendent of Branches’ and employed inspectors who were responsible for control of the Bank’s 256 branches and who ensured that rules and procedures were strictly followed and clients were served well (Cameron 1995). Most banks had similar inspection regimes. This is the origin of the term ‘gentlemen bankers’, which we are referring to as the Club.

The reputation of members of the Club was protected by ‘recognition’ by the Bank of England. The Bank expected self-discipline and mutual support for Club members who got into difficulties, yet as lender of last resort it played a leadership role in this system, and could withdraw access to essential facilities or to government business. The trustworthiness of this regulatory structure was thus based on a combination of moral suasion and penalties. Remuneration was relatively low with little or no element of performance-related pay. Managers generally avoided conflicts of interest as a matter of

---

6 Bankers in the experiment not subject to bonuses may still be subject to motivation crowding out indirectly. As documented in the next section, the domination of bank boards by those incentivized in that way profoundly changed the culture of British banks post ‘Big Bang’.

7 This section draws extensively from Jaffer et al. (2014). See also Martin (2016) and Offer (2014).
personal integrity. Competition on loan and deposit rates was weak, and rationing (often at government behest) led to large unsatisfied demands for loans from creditworthy customers.

Growth and consolidation in British banking occurred over the latter half of the twentieth century. A larger banking sector was seen as a healthy part of the structural adjustment from manufacturing, which was in decline, towards a high-skill service-based sector. Consolidation was said to permit the British banking industry to compete internationally. However, an unintended consequence was the concentration of power in the large banks, which in 2008 were deemed too big to fail (TBF) or too important to fail (TIF). The leaders of the banks aggressively pursued ever-higher returns, and acquisitions provided additional capital to underpin an increasingly wide range of activities. Banks acquired major institutions, such as building societies, which were less highly leveraged, in order to take advantage of their assets and customer base.

Growth and consolidation was spurred by general financial deregulation, though the de-regulation of credit began in earnest after ‘Competition and Credit Control’ in 1971-73 and accelerated with: the lifting of the ‘corset’ in 1980; the entry of banks into the mortgage market; and finally the permission for building societies to convert into banks (Offer, 2014). The necessity of deregulation arose out of the economic growth of the mid-twentieth century with the attendant unsatisfied demand for loans. A dramatic uptake of home ownership and superannuation increased the volume, value and sophistication of financial transactions. The mood of the times paved the way for change too. Ideological forces were at play during the high-water mark of the Reagan/Thatcher era.

At the so called ‘Big Bang’ in 1986 fixed commission charges were abolished, as was the distinction between stock-jobbers and stockbrokers, and the Stock Exchange changed from open outcry to electronic trading. The impact went far beyond these particular reforms, however. Previously separate financial organisations began to merge, and capital markets became dominated by global investment banks with large capital bases. Firms that had previously been partnerships or operated under extended liability became large limited liability companies, creating incentives for greater risk taking. There was a wave of interest expressed by foreign owned banks in establishing a presence in the City, often through acquisition.

The arrival of overseas banks transformed the culture of the City of London. It was often accompanied by the formation of a ‘markets division’, managed by people who had begun their careers as traders. These individuals came to dominate the boards, management committees and culture of their banks. Their high levels of pay led to a compensating surge in the pay of other bank board members, which could be justified only by raising shareholder expectations of returns. These higher returns were achieved by increasing the levels of leverage and risk assumed by the bank and its shareholders.

Thus, to make the link to section 2, even those who did not receive bonus-based pay packages began to suffer motivation crowding out. Conservative virtues of probity and integrity which had been the hallmark of banking culture were replaced by the aggressive pursuit of profit.

The Club had been based on delivering a service, but deregulation created financiers who were all about delivering very high returns through the assumption of risk. As they say, the rest is history. Turner (2010) catalogues the unravelling of the system in 2008, which will be familiar to our readers. Instead, we now turn to the question of how to represent analytically what was lost with the demise of the Club.

---

8 There was also a trend consolidation among the building societies. The over 700 building societies which existed in 1960 were reduced to under 100 by 1990 (Jaffer et al., 2014)
4 Representing Moral Restraint

In this section we represent moral restraint analytically as either moral optimization or moral prioritization. Our stylized example is a monopolist bank that shows restraint by moving towards the competitive equilibrium instead of maximizing profits. In choosing this example we are drawing on one aspect of the history of the Club, namely that did not fully exploit its monopoly power.\(^9\)

**Figure 2: Banking Monopolist**

In the model outlined in the appendix, there is a downward sloping demand for loans in a loanable funds framework for the determination of the lending rate. The consolidated balance sheet has a binding reserve requirement, with deposits as the only liability and reserves and loans as the only assets. Deposits are available at an exogenous rate in infinite supply, leading to a flat supply (marginal cost) curve. Consumer surplus (CS) is the area below the demand curve and above the lending rate, and producer surplus (PS) is the rectangle between the price and the (flat) marginal cost curve. We assume no fixed costs, so PS is bank profit. The lending rate is the ‘price’ in this market and loans are the ‘output’.

As drawn in figure 2, a monopolist restricts loans to the profit (PS) maximizing point A so that CS is the small triangle with apex at the intercept of the demand curve, PS is the rectangle below it, and welfare is the sum of the two. If the monopolist moves from A towards B, the competitive equilibrium, CS grows and PS collapses to zero. The deadweight loss incurred if the monopolist remains at A is marked in figure 2.

In order to discuss the preferences the monopolist has over their own profits and the welfare of their customers, we need to transform the PS and CS duples in figure 2 for every chosen level of output into a function in PS×CS space. This will give us the set of feasible trade-offs between PS and CS allowing us to overlay the indifference curves for the monopolist.

This function is derived analytically in the appendix, but we can intuit it here. First, consider the relationship between PS and CS as the monopolist expands output from A to B. Point A gives the maximum of PS by definition while point B gives the minimum of zero, and CS is increasing in output. Thus, in moving from A to B, CS is increasing and PS is decreasing to zero. Second, consider the relationship between PS and CS as the monopolist restricts output, raising price from A. Since

\(^9\) More realistically, the appendix shows the oligopoly situation. We did not use this for the main text because, for the points we need to make, it adds complexity without insight.
point A gives the maximum of PS any reduction of output will cause PS to fall, and CS always falls as the price rises.

Finally, from figure 2 it is clear in moving from A to B that the gain in CS is offset by a corresponding loss is the sum of PS and the deadweight loss. These intuitions are all combined to draw figure 3. We then consider a monopolist with Cobb-Douglas preferences \(CS^\alpha PS^{1-\alpha}\) who either maximizes profit (\(\alpha=0\)), or balances interests of consumers and producers (\(0<\alpha<1\)) to arrive at a point of moral optimization (O).

**Figure 3: Moral Optimization (O) and Profit Maximization (A)**

Figure 3 shows a monopolist with other-regard choosing O as the outcome of moral optimization, and a monopolist with lexicographical preferences for profits choosing A. We relate the production points (O and A) to a stylized description of history as follows: the Club first failed to exploit monopoly power (at O), and then those who followed exploited monopoly power (at A).

A monopolist who approaches B (in the limit, B would require \(\alpha=1\) for it to be optimal) trades off PS and CS one-for-one at the margin. We might say monopolists with \(\alpha>0\) have ‘stakeholder’ preferences because they approach a point (B) where profits and the CS of customers are traded equally at the margin. That this is so at B is proven by using the result that the competitive equilibrium maximizes the sum of PS and CS (which is obviously true in figure 2).

\[
\frac{d(PS + CS)}{dCS} = \frac{dPS}{dCS} + \frac{dCS}{dCS} = \frac{dPS}{dCS} + 1 = 0 \quad \Rightarrow \quad \frac{dPS}{dCS} = -1.
\]

In our stylized historical narrative, a post deregulation monopolist for which point A is optimal acquires indifference curves where \(\alpha=0\), where the direction of the arrow on these lexicographical preferences shows increasing utility. With these preferences, point A becomes the optimum. This post deregulation monopolist only cares about customers because they are a means to profit. We might call these ‘shareholder value’ preferences (assuming for simplicity that all profits go to shareholders who are assumed to be the same people as managers).

Returning to the pre-deregulation world of the Club, we are not required to model the eschewing of monopoly power by the moral optimization shown in figure 3. We can also use the concept of moral prioritization, and we do so in figure 4.
Suppose that the monopolist decides as a matter of principle that monopoly power is not to be fully exploited, despite having lexicographical preferences with respect to profits. Their decision is then to produce at O, the point we said was optimal with other-regarding preferences in figure 3. The dashed arrow indicates a ‘commitment’ (Sen, 1977) the cost of which can be measured in terms of the utility difference between the two lexicographical indifference curves or by the loss in profits marked as PS* in figure 4.

**Figure 4: Moral Prioritization, and Competition Policy**

In showing how O can be chosen as a matter of moral optimization or as a matter of moral prioritization, we are making an important point that there are some moral decisions which defy a full explanation without access to the mental states or reasoning of the agent. We know that moral prioritization is real because of the unease that accompanies utterances such as ‘optimal amount of deceit’ or, even more so, the ‘optimal amount of workplace violence’. However, this example shows that some choices can be motivated either by moral optimization or by moral prioritization.

Figure 4 also shows how a policymaker can avoid any consideration of motivation at all, using the tried and true remedy of competition policy. Faced with uncertainty about the motives of a monopoly bankers—both in terms of what quantity of loans they offer (A or O or B) and why they might not fully exploit market power (moral optimization or prioritization)—it is straightforward to add more firms into the market for loans rendering this discussion irrelevant. Thus the dashed line in figure 4 does double duty in showing how competition policy could drive the market for loans from a monopolist who chooses A towards a competitive equilibrium. With enough firms, the government could push the market all the way to B.

Competition policy is particularly germane for making motives about exploiting monopoly power irrelevant, and this has been our focus to this point. However, it is also relevant for moral lapses like deceit. If there is an abundance of firms to choose from, a reputation for dishonesty could spell disaster for the morally unrestrained firm. Society as a whole would benefit if these firms were then plucked from the market by the invisible hand.

For these, and other moral lapses, completion policy thus has an important part to play in disciplining firms and thereby ‘economizing on virtue’ (Brennan and Hamlin, 1995). Can it fulfill this role for the finance industry?
5 The practical limits to the promotion of competition in banking

We make three claims in this section, and in so doing raise doubts that competition can discipline financial firms and individuals.

Our first claim is that survival of poorly performing firms in the finance industry is more likely than for other industries. This is in part due to the special importance of the finance industry as a source of basic economic infrastructure, such as the payments mechanism, but it is also arises from the relative severity of financial sector recessions. Our second claim is the finance industry is much harder to understand, and therefore to monitor, than many other industries, and this is a problem as much for consumers as for regulators. Our third claim is that the effectiveness of competition policy in the financial sector is problematic. These claims are relative. We are not claiming that the promotion of competition in finance is doomed to failure. Rather our claim is that competition is relatively difficult to achieve, making our concerns about the practical ethics of the finance industry relatively important vis a vis other regulated industries.

The desirability of competition has been a longstanding area of discussion within the economics discipline, starting with Adam Smith’s observation that self-interest guides society to good outcomes via the operation of the so called invisible hand. The high water mark of this program was the Nobel-Prize-winning work of Arrow and Debreu (1954), which Bowles (2016) persuasively suggests should be re-named the Invisible Hand Theorem.

This is an ongoing conversation (Bowles, 2011). Yet whatever the general benefits of competition, the workings of the invisible hand depend upon the elimination of firms and techniques which do not find support in the marketplace – the rough and ready testing ground for social value in our economic system. It is precisely here that the first difficulty appears.

5(i) Impediments to the elimination of poorly performing banks

Financial institutions, and the credit contracts they write, generate very special risks for the economy (Turner, 2010). As was amply demonstrated in 2008, these risks can prove too great in a crisis, so that the authorities cannot afford to let poorly performing firms go bankrupt. This includes financial firms deemed Too Big to Fail (TBF), although size is not the only factor involved in any bailout decision. Not only have bailed-out institutions been able to access public money, but some unethical managers may walk away with substantial bonuses, leaving them free to mismanage their institutions again. Quite apart from the moral problems outlined in Simpson (2016), Wolf (2010) argues that bonus pay schemes, particularly those based on share options, were and are too successful in aligning these managers’ interests with shareholders. The difficulty is that both groups benefit from risky strategies which, in the event of a bad outcome, can be underwritten with public money. A good outcome, on the other hand, rewards both for a heads-I-win-tails-you-lose investment style.

Since the crisis measures to ring fence banking, so that ordinary depositors are not financing risky investment banking, are being applied in many jurisdictions (CMA, 2016). There are also procedures for ‘bailing in’ failing banks, requiring private bond holders to suffer losses to avert a public bailout, and requirements for increased capital. Both are discussed in the Vickers Report on the UK banking

---

10 “[An investor] intends only his own gain, and he is in this, as in many other cases, led by an invisible hand to promote an end which was no part of his intention. Nor is it always worse for the society that he was no part of it. By pursuing his own interest he frequently promotes that of the society more effectually than when he really intends to promote it.” (Smith, 1999/1776, Book IV: 32)

11 In considering the situation of Lehman Brothers and Bear Stearns in 2008, each carried counter-party risk and more general contagion risk. We might say they were Too Important to Fail (TIF).
system (Edmonds, 2013) and the intention is to hold high risk takers to account. Time will tell if these measures are sufficient to avert the pilfering of public money in a serious crisis. The difficulty is that the interconnections that remain in the financial system are such that the authorities might yet choose to assist an ailing investment bank because of systemic risk or due to the adverse confidence effects on the whole economy.

On the latter, there is compelling evidence, if that were needed after 2008, that recessions which include a financial sector crisis are deeper and longer than other crises (IMF, 2009). This means that even if ring-fencing, bail-ins and increased capital (CMA, 2016, Edmonds, 2013) can protect the public purse, an easing of monetary policy in and of itself during a financial crisis may protect firms that otherwise should perish. Figure 5 shows these features averaged over financial sector recessions, and we note particularly that easier monetary policy is generally pursued.

Thus, given the prudential protection of TBF and the macroeconomic protection arising from the threat of a protracted recession, there remains a danger that dysfunctional firms will survive and have a longevity that undercuts the invisible hand. To the extent that this is so, the culture of finance will be adversely affected by the survivors.

**Figure 5: Financial Sector Recessions are More Severe**

(IMF calculations, quarters since peak in real output)

![Financial Sector Recessions are More Severe](image)

(IMF, 2009, Figure 3.8, pg. 118)

5 (ii) Difficulties for customers and regulators in monitoring bank performance

We now come to the second claim: that it is relatively difficult to understand and monitor the financial system. The fundamental problem here lies at the conceptual level, rather than the pragmatic. Whatever accounting conventions are used, there is conceptual uncertainty about the measurement of risk, and so any accounting profits cannot be effectively adjusted for risk. Haldane et al. (2011) propose that this should be a priority in any reform of the measurement regime.

‘As it is rudimentary to its activities, finding a more sophisticated approach to measuring risk, as well as return, within the financial sector would seem to be a priority. The conflation of the two can lead to an overstatement of banks’ contribution to the economy and an understatement of the true risk facing banks and the economy at large.’

(Op. cit. pg. 106)

Financial firms are able to conduct trades with a zero, or even negative expected return which are nonetheless extremely profitable in the short term (Wolf, 2010). They may undertake large-volume trades each with a high probability of a small gain and a small probability of a huge loss. A good
example of this is the so-called carry trade. A bank may purchase a relatively high yielding currency asset which incurs an expected cost of a subsequent depreciation which is rarely realized. The resultant risks are not easily provisioned, and mistakes can occur. Although isolated ‘tail events’ (in this example, large currency depreciation) can be absorbed relatively easily, correlated losses are another matter. No one could have imagined the confluence of events which set off the 2008 financial crisis.

These kinds of trades are further explored in Noe and Peyton Young (2014). They show how a manager can use derivatives to increase investors’ normal rewards, and their own reward, at the same time as creating tail risks for the investor. They show that it is easy for managers to do this by purchasing a derivative which regularly pays out a fee, in exchange for the occasional right to seize the entire asset—the ‘tail risk’ feature of such derivatives. Compensation schemes with bonuses usually have the feature that in the periods when the client makes a good return, the manager gets a good bonus. However, when the asset is seized by the person who issued the derivative, the investor loses everything, but the fund manager merely fails to get his or her bonus.

Performance contracts would guard against such untrustworthy behaviour only if they enforced very large losses on managers in bad times—for example, bankruptcy or imprisonment—but managers do not often face that possibility. Since such a setup will normally yield good returns for the investor, it will take many years before investors can determine with any degree of confidence whether a fund manager who is generating good returns is actually trustworthy, or is instead acting in a dishonest way by inserting tail risks. Of course, this is made more difficult if the manager actually believes that he is acting in a trustworthy manner, but is unwittingly relying on tail risks to deliver the promised returns (Morris and Vines, 2014).

Another measurement problem concerns marginal cost. In our stylized model in the appendix, this is just the deposit rate adjusted for the reserve requirement. However in reality many bank inputs are difficult to measure and some, such as wages, do not seem to obey ordinary market forces. In figure 6, from Philippon and Resheff (2009), we have evidence of extraordinary departures from a benchmark wage in the US. The benchmark ostensibly takes into account the higher education of finance workers and the risks of unemployment, relative to average wages (unity) but it comes nowhere near the actual wages paid. The period in the mid-twentieth century is generally regarded as one of financial repression, where a range of government restrictions prevented the expansion of finance, but it did at least ground wages in some understandable way.

**Figure 6: Inexplicable Wages**
The seemingly ungrounded cost base of some banks makes evidence of relatively low margins, seen in figure 7 (reproduced from Coccorese (2014, pg. 37, figure 2)), less impressive than it might otherwise be. The so called Lerner index is a standard measure of the proximity to perfect competition. It is the proportional reduction from current prices required to reach marginal cost. This is also called the mark-up. In a perfectly competitive environment price is equated to marginal cost, and so the value of the index ought to be zero.

Mark-ups are generally 10 to 15 per cent, though margins can be compressed during a downturn, as they were in some jurisdictions over 2008. After the crisis there were some forced mergers, and naturally these can increase banking concentration.12

The US was at the centre of the 2008 financial crisis, and so it holds a special place of interest in the diagram. Prior to 2008, the US financial system was believed to be an engine room of socially valuable innovation, and the downward trend evident from 2004 to 2008 might easily have been interpreted as indicating economic value creation. However, the most destructive financial crisis since the Great Depression and the catalogue of dubious practices which subsequently came to light (Turner, 2010) overturns this claim about the benefits of competition in the US.

**Figure 7: Financial Markup in Different Regions**

(competitive Lerner index benchmark of zero)

In evaluating the state of financial institutions a great deal depends on how the assets of the institution are valued, and these are overwhelming made up of loans. Writing contracts over far-flung contingencies takes the risk bearers into the realm of Knightian uncertainty, rather than the stochastic uncertainty amenable to constructing probabilistic models. So it becomes problematic for regulators to gain a deep understanding of financial institutions and their accounting, let alone the consumers who are trying to make judgments about the performance of their own institution (Wolf, 2010). On the

---

12 The compression of margins was also an outcome of the crisis and recession, but at least in the UK banking became more concentrated (Edmonds, 2013).
latter point, the Vickers Report (Edmonds, 2013) described a generally disengaged UK consumer base which is reluctant to apply competitive pressures by switching deposit accounts.\textsuperscript{13}

The upshot of all these information problems is that there is a significant degree of unquantifiable risk built into the financial system, and the sign of the partial derivative with respect to increased deregulation may be more uncertain than it is for other industries. Indeed, recent US history is rather supportive of the competition fragility story (Berger et al., 2009).\textsuperscript{14} In this account, it is competition itself which tempts banks to pursue zero or negative expected excess return strategies which in the short run look very profitable.

5 (iii) The probable ineffectiveness of competition policy

Competition policy has three main concerns in dealing with concentrated industries: the scope for reducing market power by restructuring leading firms, preventing the formation of explicit or implicit cartels, and reducing barriers to new entrants. In services, a further concern is the inability of consumers to understand the offerings available in the market place due to the complexity of the service provision and/or the pricing of alternatives. Competition enquiries in the UK provide an example of the scope and potential efficacy of policy in the banking industry.

Competition policy has a particular reason for wanting to understand financial institutions in depth. If the entry of extra firms is not straightforward, it might be that greater efficiency can be achieved by restructuring existing firms to curtail monopoly power. In the model in the appendix, this would mean splitting up existing firms rather than relying on the entry of new ones. To implement this strategy in the real world the competition authorities need to confirm the existence or otherwise of monopoly rents. Yet informational asymmetries may mean that enough information ‘leaks’ to other competitors to make implicit collusion possible, but not enough information to support the prosecution of any resultant cartel. Where to draw the line for a sustainable legal case is a difficult matter. Any general trends in evidential standards are particularly relevant for complex court cases about financial firms, and if authorities are averse to legal risk, there may come about a de facto higher burden of proof for finance competition policy than for competition policy in other industries. In the absence of ethical standards for conduct, a point to which we will return, this may lead to implicit collusion and rent seeking which cannot be curtailed by the authorities.

The Independent Commission on Banking in the UK (chaired by Sir John Vickers) published its final report in September 2011. It noted the very high concentration in the UK retail banking sector, with a four firm concentration ratio of 77\% for personal accounts and 84\% for SME accounts. These ratios had increased markedly as a result of the financial crisis, as the major banks had been encouraged by the financial authorities to take over failing smaller banks and building societies. The Commission’s focus was on measures to promote competition in the market, by the encouragement of challenger banks, greater transparency and ease of switching for consumers, and giving the Financial Conduct Authority (FCA) a ‘clear primary duty to promote effective competition’. The Commission’s immediate proposals involved ensuring that a strong challenger be created by the divestiture of a major portion of the branches and current accounts by Lloyds Bank. In respect of ease of switching,

\textsuperscript{13} Part of this is doubtless due to the relatively small size of bank fees and charges in a household budget, rather than an information problem. As in the case of bottled water, which can be thousands of times more expensive than tap water, price differences can remain un-noticed when an item has a trivial expenditure burden.

\textsuperscript{14} As explained by Berger et al. (2009), this contrasts with the competition stability story where monopoly power raises interest rates to a point where bad borrowers and bad managers respond to adverse selection and moral hazard.
the proposal was that it should be made possible for an account holder to switch accounts within seven working days, with safeguards for customers with respect to redirection of credits for up to a year.

Retail banking in the UK was the subject of a competition investigation by the Competition and Markets Authority (CMA) which reported in August 2016. It noted that all the key competition recommendations of the Commission had either been implemented or were in the pipeline for implementation. It noted the continuing high concentration in the sector, but on the basis of evidence available (in particular the integration of banking and investment activities in the major banks made it problematic to explore the profitability of retail banking alone), it concluded that it could not be demonstrated that monopoly profits were being earned. Instead the investigation focused on the lack of consumer engagement and the apparent reluctance of account holders to switch from their current banks, even if better deals were available. This reluctance was strongest where the account holder had an existing overdraft. The problem was partly that consumers were ignorant of potential gains from switching, and partly that they were afraid that switching might be complicated, time consuming, and even costly if something went wrong. This fear was despite institutional innovation that had simplified the process of switching and provided strong safeguards (the ‘current account switching service’). This lack of consumer engagement remained a major stumbling block to the emergence of new competitors: while several new entrants had established themselves in the market since the financial crisis, they had found it costly to attract new account holders. The most successful were those that linked their new banking operations to an existing customer base in retailing. The main recommendations of the CMA were for measures to increase consumer awareness, and in particular the development of online tools to enable a bank’s offerings to be measured against a common standard for price and quality. The CMA also noted that innovations in internet banking and the use of mobile technologies had the potential to weaken the dominance of the major banks in retail banking sustained their extensive network of branches on high streets. New entry is made much easier if a challenger bank does not need to develop a branch network, and can instead rely on digital networks, though these too will involve substantial up-front costs to develop robust and secure systems.

To conclude, there is much more that could be said about the feasibility and desirability of pursuing competitive financial markets, but we have achieved our purpose by laying out some of the unique challenges. Interestingly, the number of papers measuring the extent of competition has dropped off dramatically since the crisis, though this could be due to a concern about the stability of the stochastic processes underlying the models, as much as disillusionment with the promotion of competition. Given our interest in the relative difficulty of promoting competition, we may also note the sentiment of Casu and Girardone (2009), that the view that competition is unambiguously good is less likely to be true in banking than it is in other industries. As we flagged earlier, we do not by this imply that competition policy is useless. We claim only that it faces special problems in finance compared with other industries.

6 Banking as a Profession

We have argued that competition policy is at best partially effective in pushing banks to operate at the competitive equilibrium, or undertaking other desirable actions such as not misleading clients or

regulators. To the extent that competition policy fails, and moral restraint succeeds, there is a corresponding imperative to foster moral restraint among individual participants in the marketplace. Principled agents solve the principal-agent problem because trustworthiness is the basis for trust. This conclusion poses a question of feasibility. Is this possible? We address this in closing. Our aim is not to identify who is responsible for fostering moral restraint in finance. Nor is it to provide a definitive set of instructions on how to do so. It is merely to show that, plausibly, it is possible.

The idea of fostering a culture of moral restraint may seem quaint or romantic to those who work in banking. But a broader perspective shows it to be entirely feasible. Sustaining such a culture is a central task of any organised body that oversees and maintains a profession. Fostering that culture where it does not presently exist is no more demanding—and no less—than turning banking from a business to a profession. More accurately, the task is of returning it to a profession.

Professionalization is a characteristic response to markets where there is significant asymmetry of information; where there is reliance on judgment, which in the short term can be opportunistically exploited by a professional with detection by the non-professional difficult; and where what is offered in the transaction has a critical practical value, not being easily replaceable. In law, it is, at the limit, one’s freedom; at school, education; in medicine, health; for the military and police, our collective security and safety. In banking, it is one’s critical capital: pension and housing. As the examples show, there are numerous other workplaces that maintain standards of professionalism, and where it is expected that practitioners do not exploit informational or monopoly power at the cost of those whom they serve. Moral restraint is there expected and provided.

The asymmetry of information implies that it is easy for the professional to look out for their own interests, either by engaging in undetected (or undetectable) negligence, or by charging exorbitant fees. Thus, while it is true that any job should ultimately be about the service of another person, it is especially important to emphasise this service ethic in the case of the professions. Robert Downie’s classic (1990) article outlines a number of features of professionals, and we highlight three.

1. The professional has specialist skills and a sophisticated knowledge base.
2. Partly because of a knowledge advantage, and partly because the professional acts on behalf of the client, there is a power imbalance in the professional’s favour. The professional needs therefore to express beneficence tempered with honesty and justice (what Downie calls ‘integrity’) so that beneficence doesn’t degenerate into paternalism.
3. The professional’s fundamental telos is the good of the client rather than sectional or other interests.

The contrast between this understanding of professionals and the modelled agents in a principal-agent setup is striking. Both understandings acknowledge an informational advantage, due to hidden action in the principal-agent model and to expertise in the case of a professional (point 1). However, in the principal-agent model anyone possessing an advantage exploits the situation egoistically. In contrast, Downie flatly contradicts this in point 3. The end-goal of professionalization is to develop a group of people who choose to seek the good of the client.

There are a number of ways that professionalization happens. The most obvious is through a professional body’s self-certification of its members. In the UK, the General Medical Council enjoys a monopoly over the medical profession. In order to practice, you must be a member. The power to strike someone off the GMC is then one of the key ‘levers’ with which the standards of the medical profession are maintained. Because there is a process to adjudicate on unsatisfactory performance, a non-professional’s suspicion of opportunistic behaviour by a professional can be investigated and
charged. The process is swifter and more accessible than formal law. Other professional organisations do not enjoy the same monopoly power as the GMA, but the certification function they provide gives an assurance of quality to those using the service. Proposals that bankers should be required to take oaths of service on starting work should be understood as a strategy for introducing professionalization (de Bruin 2016). Part of professionalization will also involve reform of pay structures, insofar as performance-related pay has been a key contribution to the crowding out of the moral motivations in finance.

What does the work of motivating performance for a professional? It is sometimes said that the loss of incentives such as performance-related pay will lead to financiers becoming lazy. In the short term, perhaps this is so, but it is dubious in the long term. Instead, we conjecture that the claimed connection is a rationalisation for those who need to justify to themselves their high pay. Plenty of other workplaces are characterised by exceptionally hard work without exceptionally high income. Many become professionals because they endorse, for themselves, the ethic of service that it embodies, and they feel rewarded in the work because of the good they thereby enable for others. The older (theistic) term to describe this was a ‘vocation’, or calling. Others do so out of a desire for esteem. This may be the esteem of those outside the profession. Often it is the esteem of those inside too, with people working their way up the ladders of a distinction out of a desire for the approval of their peers. As Philip Pettit observes, however, because most people dislike it to be known that they desire others’ esteem, the esteem-based mechanism for motivating this work is ‘parasitic’ on the more avowable, service-related conception of the profession (1995: 212).

Professionalization is thus a strategy that seeks to recruit the disposition to cooperate in pursuit of socially beneficial ends. Promotion within the profession, for instance, becomes conditional on an individual having shown that they endorse the norm of service that the sector embodies. In Pettit’s terms, again, a screening process protects a culture of compliance. This ensures that morally restrained individuals advance within the institution (2002: 297ff.). With professionalization, further, the adverse selection problem of finance whereby selfish individuals are more likely to enter the sector will be, if not overcome, at least mitigated.
References


Casu, B. and Girardone, C.


IMF (2009), ‘From Recession to Recovery’, Chapter 3, World Economic Outlook, April

John Stuart Mill (1843), The Collected Works of John Stuart Mill, Volume VIII - A System of Logic Ratiocinative and Inductive Part II CHAPTER IX: Of the Physical, or Concrete Deductive Method.


Sen, A. (2005), ‘Why Exactly is Commitment Important for Rationality?’ *Economics and Philosophy*, 21(1), 5-14


Appendix Oligopoly Model

Consider a simplified loanable-funds-theory banking world where banks accept deposits at a rate \( r_z \), determined by the monetary authority, and loan out funds \( Q \) for consumer projects at an interest rate \( p \) determined by a decreasing demand for funds schedule. The size of the banking system is determined by a reserve \( R \) requirement, which is assumed to bind. In the absence of fixed costs profits are the same as producer surplus (PS), and we can write this as a function of loans.

The balance sheet balances, so the following table applies to each bank, and, to the consolidated banking sector.

<table>
<thead>
<tr>
<th>Bank Balance Sheet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assets</td>
</tr>
<tr>
<td>Q (loans)</td>
</tr>
<tr>
<td>R (reserves)</td>
</tr>
<tr>
<td>Liabilities</td>
</tr>
<tr>
<td>Z (deposits)</td>
</tr>
</tbody>
</table>

Since the balance sheet balances,

\[
Q + R = Z \tag{2}
\]

and we assume the reserve requirement \( R = \theta Z \) binds, we can substitute the reserve requirement into (2) and then derive a reduced form for profits.

\[
Z = \frac{Q}{1 - \theta} \tag{3}
\]

\[
PS = pQ - r_z Z = pQ - \frac{r_z}{1 - \theta} Q \equiv pQ - rQ \left( r = \frac{r_z}{1 - \theta} \right) \tag{4}
\]

We assume that the demand for funds has a choke price (intercept) of 100 per cent, and is linear.

The Demand for Funds

\[ P \text{ (lending rate)} \]

\[ P \text{ (lending rate)} \]

\[ r \]

\[ q \]

\[ Q \text{ (loans)} \]

\[ PS \]

\[ CS \]

Demand: \( P = 1 - \theta Q \)

Supply MC: \( P = r \)

\[ 16 \text{ This ignores capital adequacy requirements for simplicity. A premium on the deposit rate could be added for monitoring costs, which are largely marginal costs rather than fixed costs. Since the size of the balance sheet is determined by the demand for loans, money is endogenous.} \]
We outline a multi-firm model (with a monopolist as a special case) which leads to a market choice of $Q$ and therefore $PS$ and $CS$ as shown in the figure. We can make $PS$ an explicit function of $CS$ by substituting out $Q$. This shows the feasible trade-off between $PS$ and $QS$. We can then put preferences associated with moral optimization or moral prioritization on the same mapping. We confine ourselves to considering a situation where the consumers and banks interact in a market supporting a single lending rate, ruling out price discrimination and side payments. We also assume that the bank cannot make losses.

We write consumer surplus ($CS$) and producer surplus ($PS$) as a function of $Q$, and eliminate. We must distinguish $q$ as a particular value versus $Q$ as a variable over which integration occurs.

$$PS = \int_0^q \{1 - a q - r\} dQ = (1 - r) q - a q^2$$  \hspace{1cm} (5)

$$CS = \int_0^q \{1 - a Q - (1 - a q)\} dQ = \frac{a q^2}{2}$$  \hspace{1cm} (6)

$$\Rightarrow \quad q = \left(\frac{2}{a} CS\right)^{\frac{1}{2}}$$  \hspace{1cm} (7)

The profit maximizing monopoly solution maximizes (5) and is a useful benchmark.

$$q = \frac{1 - r}{2 a}$$  \hspace{1cm} (8)

A social planner who wanted to maximize total surplus would first sum (5) and (6),

$$PS + CS = (1 - r) q - a q^2 + \frac{a q^2}{2} = (1 - r) q - \frac{a q^2}{2}$$

and differentiate. The solution naturally implies more output than the monopoly solution.

$$q = \frac{1 - r}{a}$$  \hspace{1cm} (9)

We will come presently to how a set of oligopolistic firms come to choose the market outcome $q$, but (5), (6) and (7) will be valid irrespective of this, and so in (10) we functionally connect $PS$ and $CS$ by substituting (7) into (5).

$$PS = (1 - r) \sqrt{\frac{2}{a} \sqrt{CS} - 2 CS}$$  \hspace{1cm} (10)

We now turn to the task of showing which ($CS$, $PS$) duples will occur, with their attendant $q$ values, given the number of oligopolistic banks ($n$) and the degree of sympathy for consumers ($\alpha$) measured in a utility function (11).

$$U_i = CS[q_i]^{\alpha} PS[q_i]^{1-\alpha}$$  \hspace{1cm} (11)

We assume the market being served by $n$ identical banks. For loanable funds demand (12) we assume zero conjectural variation so bank $i$ treats bank $j$’s lending as (lower case) $q_i$. The individual supply curve remains at $r$ for all banks.
Individual bank producer surplus is derived from revenues (based on (12)) minus costs.

\[ PS_i = \int_0^{q_i} \left(1 - a \sum_{j \neq i} q_j - a q_i - r \right) dQ_i = \left(1 - a \sum_{j \neq i} q_j - r \right) q_i - a q_i^2 \]  

And the derivative with respect to \( q_i \) follows immediately.

\[ \frac{dPS_i}{dq_i} = \left(1 - a \sum_{j \neq i} q_j - r \right) - 2aq_i \]  

For a symmetric solution all quantities are \( \frac{Q}{n} \), so we can re-write (13) and (14).

\[ PS_i = \left(1 - a \frac{(n-1)Q}{n} - r \right) \frac{Q}{n} - a \left(\frac{Q}{n}\right)^2 = \frac{1}{n} \left(1 - r\right)Q - a Q^2 \]  

\[ \frac{dPS_i}{dq_i} = 1 - r - a Q \left(1 + \frac{1}{n}\right) \]  

Unsurprisingly, (15) is just a fraction \( 1/n \) of total profits in (5), slightly rearranged. Equation (16) makes use of the fact that \( \sum_{j \neq i} q_j + q_i = Q \). Banks are assumed to care about aggregate consumer surplus, \( aQ^2/2 \) in (6), and when this is differentiated with respect to \( q_i \) (with zero conjectural variation) we obtain \( aQ \). We are now in a position to optimize utility (11) with respect to \( q_i \).

\[ U_i = CS[q_i]^\alpha PS_i[q_i]^{1-\alpha} \]  

\[ \frac{dU}{dq_i} = \alpha CS^{\alpha-1} PS_i^{1-\alpha} \frac{dCS}{dq_i} + CS^\alpha (1-\alpha)PS_i^{-\alpha} \frac{dPS_i}{dq_i} = 0 \]

\[ \Rightarrow \frac{PS_i}{CS} \frac{dCS}{dq_i} + (1-\alpha) \frac{dPS_i}{dq_i} = 0 \]  

From figure 3 in the main text, it is clear that if \( \alpha=1 \) in (17) banks maximize CS by setting \( PS_i=0 \) which gives the highest feasible value of CS. If on the other hand \( \alpha=0 \), and using instead (18), banks maximize \( PS_i \) by setting by setting its derivative with respect to \( q_i \) equal to zero. Intermediate values of \( \alpha \) lead to intermediate values of \( q_i \), and we showed such a point \( O \) in figures 3 and 4. We substitute (15), (6) and (16) for \( PS_i, CS, \) and \( dPS/dq_i \), set \( dCS/dq_i=aQ \) and solve (18) for \( Q \).

\[ Q = \frac{2(1-r) + (1-\alpha)(1-r)}{\alpha \left( \frac{2a}{n} \right) + (1-\alpha) \left( a \left(1 + \frac{1}{n}\right) \right)} \]
Table 2: The Range of Solutions for Q, for given $\alpha$ and n

<table>
<thead>
<tr>
<th>n general</th>
<th>$\alpha=0$</th>
<th>$\alpha=1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>n=1</td>
<td>$\frac{n}{n+1} \frac{1-r}{a}$</td>
<td>$\frac{1-r}{a}$</td>
</tr>
<tr>
<td>monopoly</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n=2</td>
<td>$\frac{2}{3} \frac{1-r}{a}$</td>
<td>$\frac{1-r}{a}$</td>
</tr>
<tr>
<td>duopoly</td>
<td></td>
<td></td>
</tr>
<tr>
<td>....</td>
<td>...........</td>
<td>...........</td>
</tr>
<tr>
<td>n=\infty</td>
<td>$\frac{1-r}{a}$</td>
<td>$\frac{1-r}{a}$</td>
</tr>
<tr>
<td>competitive</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In following figure we show how the solutions in the table are located on a plot of (10). The solutions shift rightwards down the LHS curve as $\alpha$ increases, for a given $n$. We show $n=1$ for an example. And for an infinite number of banks the solution converges to the competitive point B even if $\alpha$ is zero in the RHS diagram. At that point, banks are forced to trade off a unit of CS for a unit of PS, from (1).

![LHS: Increasing $\alpha$ for Monopoly](image1)

![RHS: Increasing $n$](image2)

It should now be clear that a representation of the Club, which was an oligopoly, would have $n>1$ and so it would be located between A and B on the RHS, but without the presumption that $\alpha=0$. It would then be possible to model a shift down towards B either as $\alpha_2 > \alpha_1$ (using Becker’s methodology) or as an exogenous shift for which the diagram could give us the utility cost (using Sen’s methodology). In both diagrams we assume the first point is a tangency – i.e. the oligopoly point is optimal. In Sen’s diagram (RHS Moral Prioritization) the utility difference could in principle be measured if it were old fashioned cardinal utility, or it could be measured in terms of PS, with some assumption made about the trade-off between PS and CS.
Although the above model contains no dynamics, it is possible to construct a narrative about increased competition. Firms could observe the positive profits, and then enter as new banks with the result that the positive profits are competed away.

A further point is that fixed costs would put a theoretical brake on how small firms could feasibly be. As highlighted in Draghi (2016), too much competition can force banks to operate at an inefficient scale. If the model were cast in terms of monopolistic competition with fixed costs, the solution would change to involve ‘many’ firms rather than an infinite number, provided that the fixed costs were not too large. A large number of monopolistically competitive firms also put a discipline on market participants as it does in perfect competition.

There is nothing about this model which captures the concerns raised in section 5. Indeed, as it stands it is very close to a market for any good, with the exception that the marginal cost curve might be expected to rise for a generic good. The purpose of this appendix has simply been to show that the case described in the text with just one firm – is easily generalizable to a group of firms which was of course the case with the Club. That is to say, the distinction between moral optimization and moral prohibition can still be made in an oligopolistic setting.
About PEFM

The Political Economy of Financial Markets programme (PEFM) aims to shed light on the way in which institutions, including macroeconomic policy frameworks, interact with financial markets. In the wake of the global and euro area crises, it seeks to promote a better understanding of financial markets and to contribute to improved policy formulation in the future.

Its main activities are to carry out research, hold seminars, and publish findings in outlets that range from academic articles and books to policy briefings and op-ed pieces in the international press. Three initial research groups were set up at the outset, bringing together academics, officials and market participants:

- The first research topic is **Financial Integration in Europe** – why this has not lived up to expectations, and the implications for banking and fiscal union.
- The second research topic is **Regulatory Capture**. This explores how relations between the financial sector and regulators interacted with political and ideological influences in the ‘regulatory space’, during the run-up to the crisis.
- The third research topic is **Macroeconomic Policies and Financial Stability** – asking how monetary and fiscal policy regimes can respond to instability in the private sector, without jeopardizing policy transparency.
- Several future research priorities have been identified. These include shadow banking, and also the impact of advanced economy financial policies on emerging market countries.

European Studies Centre

The European Studies Centre at St Antony’s College is dedicated to the interdisciplinary study of Europe. It has particular strengths in politics, political economy, history and international relations, and also brings together sociologists, social anthropologists and students of culture. The Centre is a meeting place and intellectual laboratory for the whole community of those interested in European Studies at Oxford. Beside its permanent Fellows, the Centre has Visiting Fellows from several European countries, as well as graduate students from around the world working on European affairs. The Centre also participates in several collaborative international research projects. Seminars and workshops on a wide range of topics are held regularly at the Centre. These involve Oxford scholars from all disciplines and their counterparts from abroad, often with the participation of students. A number of special lectures and international conferences, bringing both leading academics and distinguished practitioners to Oxford, are offered to a wider audience under the auspices of the Centre.