Restoring Trust in Finance: From Principal-Agent to Principled Agent[1]

Gordon Menzies, University of Technology Sydney
Donald Hay, University of Oxford
Thomas Simpson, Blavatnik School of Government, University of Oxford
David Vines, Department of Economics, 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

Abstract
Banking solves the Principal-Agent problem with incentive contracts which assume a Homo Economicus representative agent. However, the Principal-Agent problem can be solved by assuming a 'Reasonable Person' representative agent, from tort law, who reliably tells the truth about hidden action. We describe a non-virtuous circle in finance, which is a negative feedback loop between motivation crowding out arising from incentive contracts, and the subsequent need to offer incentive contacts because agents change from the Reasonable Person to Homo Economicus. We prescribe less reliance on incentive contracts in the practice of finance, and on Homo Economicus in the theory of finance.

Keywords: Bank Bonuses, Trust, Deregulation

JEL Codes: G21, G28, H12, E52

1 Introduction

Trust and trustworthiness are fundamental to economic welfare. They explain why most people are honest when making social security claims and why restaurants are happy to serve first and charge afterwards (Bacharach et al. 2007). It also explains why unmonitored efforts are rewarded on an hourly basis in so many workplaces, when neoclassical economic theory might suggest otherwise (Jensen and Meckling 1976). Furthermore, professionals of all kinds are sought out not just for their expertise, but for an assumed trustworthiness with respect to their clients (Downie 1990).

Banking is one industry where trust and trustworthiness are particularly important.[2] Banks collect detailed information on contracts and products as they interact with savers, debtors, investors and companies. Due to their expertise and access to private information, bank

---

1 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 to support his sabbatical visit to Oxford and thanks the Oxford Martin School for hosting his public lecture, from which this paper grew: https://www.oxfordmartin.ox.ac.uk/videos/restoring-trust-in-finance-moral-obligation-meets-econ-101-with-gordon-menzies/ We 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, H Peyton Young and Jingjing Zhang. Gordon Menzies thanks the Economic Humanist Group in Oxford. Correspondence to gordon.menzies@uts.edu.au at UTS, PO Box 123, Broadway, Sydney.

2 In this article 'banking' and 'manager' covers all kinds of financial intermediaries.

managers have power over shareholders and customers, and therefore a social responsibility to be trustworthy. As an indication of the centrality of trust to banking, the origin of the word 'credit' is the Latin *credere*: to believe, to trust.

Yet there is evidence that some bankers are relatively untrustworthy. A US Senate Inquiry into the 2008/9 Global Financial Crisis was critical of Goldman Sachs.

> 'You are taking a position against the very security that you are selling and you are not troubled? … And you want people to trust you? Why would people trust you?'

Senator Carl Levin, to Goldman Sachs CEO (quoted in US Senate 2010)

A decade later, the 2017-2019 Australian Royal Commission into Misconduct in the Banking, Superannuation and Financial Services Industry has also exposed many instances of untrustworthiness. One testimony relates how a client had paid fees to construct a financial plan not permissible under Australian law:

> 'I just feel now after all the time after all the fees and insurances … that all along they were just aiming for us to take out an investment property that you can't live in. I just felt after that, [pause] that we had been led up the garden path and had been lied to.'

Jacqueline McDowall (quoted in Danckert 2018)

Senate inquiries and royal commissions are subject to adverse selection, since only the most questionable behaviours are selected for media exposure and courtroom scrutiny. And it can be a mistake to generalize too widely based on particular business units. For example, the Salz review (2013) noted that many of Barclays' problems began in their investment arm.

However, even if the average level of misconduct is low there is experimental evidence that the *relative* incidence of untrustworthiness in banking is higher when compared with other professions. In Cohn et al. (2014), bank employees from a large international bank were rewarded in a coin flipping task, out of the view of the experimenter. When primed to think of their professional identity, the bankers as a group reported on average too many financially rewarding tosses, but they were generally honest when primed to think of their domestic identity. Yet when 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.[3]

Furthermore, banking as a profession seems to have a trustworthiness problem when compared to its past in the UK (Haldane et al. 2011, Jaffer et al. 2014, Martin 2016, Mayer 2013, Morris and Vines 2014, Offer 2014, The Economist 2017, Turner 2010) and the US (Fligstein and Roehrkasse 2016, Hill and Painter 2015). Deregulation, including extensive liberalization of derivatives, has been suggested as one possible factor allowing information asymmetries to be exploited (Noe and Peyton Young 2014, Wolf 2010) – what Berger et al. (2009) call competition fragility. In Australia, the ethical winds of change blew more than a decade after deregulation,[4] but there appears to have been a decline in trustworthiness nonetheless.

> 'From 1995 to 2010, there was an increase in the flow of foreign professionals into the big 4 banks either from overseas, or by Australians with international banking experience. … this lead to a cultural change in banking, where it [became] all about the money with a focus on short term profitability.' (Hogan 2018, slide 18)

The Salz review offers this appraisal for the UK.

> 'Over the last two or three decades, deregulation, globalisation, unprecedented prosperity and the availability of relatively cheap funding have been closely followed by an extended period of economic and financial turmoil. …[banks' alleged role in the turmoil] has been a cause of the loss of public confidence.' (Salz 2013, page 4)

In view of all this, it is a natural policy prescription to encourage errant bankers to act in a more trustworthy manner, setting high standards to meet the valid expectations of politicians and the public who 'have the right to insist on changes' (op. cit. page 5).

The contribution of this paper is to make 'trustworthy' meaningful within the analytic framework of the Principal-Agent model and to draw on that framework to discuss more broadly how to restore trustworthiness. We first make the observation that the kind of representative agent that inhabits the Principal-Agent model determines the kind of solution that is feasible. Banking solves the Principal-Agent problem with an incentive contract, which

[3] Many people would be curious to see some other employment categories included in this experiment. Peters (1987) discusses the use of lies in negotiations by lawyers 'The fact that all lawyers have deception available to them as a tactic creates a specific sort of uncertainty in negotiations.'(un-paginated). MacIntosh (2007) discusses similar issues related to accounting.

[4] 1984 saw a policy decision to allow foreign bank entry. In 1985, sixteen foreign banks were invited to liaise with the RBA to determine eligibility. Of these, fifteen went on to begin banking operations in Australia through locally incorporated subsidiaries (Parliamentary library 1995).

in turn assumes a Homo Economicus representative agent. However, the Principal-Agent problem can be solved by assuming a 'Reasonable Person' representative agent.

The Reasonable Person first made an appearance in Vaughan v. Menlove (1837), predating Homo Economicus by several decades. He or she is defined in tort law 'in terms of the interests of oneself in relationship to society's interests and the interests of others in that society' (Lydenberg 2014, pg. 288) and so we assume the Reasonable Person is someone who consistently tells the truth about their effort levels. This makes hidden action discoverable and reliability makes the hidden action contractible, so the first-best solution to the Principal-Agent problem is attainable. The first-best contract stipulates a fixed wage for optimal effort which the agent truthfully discloses, and this approximates the standard practice in many jobs.

Using both sorts of agents, we outline a mechanism – a non-virtuous circle – which helps explain the loss of trustworthiness in finance. The prominence in finance of incentive contracts in its practice, and Homo Economicus in its conceptualization, are part of finance's trustworthiness problem. We argue that their attenuation would help restore trust.

The paper is structured as follows. In section 2 we define moral optimization and moral prioritization, and we use the latter to model trustworthiness in economics. We also show how our distinction between moral optimization and moral prioritization applies to the Principal-Agent framework (Jensen and Meckling 1976) and allows for solutions to be tailored to the relevant representative agent – Homo Economicus or the Reasonable Person. In section 3 we outline our narrative for the demise of finance, called the non-virtuous circle. In section 4 we argue that a reduced reliance on incentive contracts and an altered economics training would be helpful to restore trust in finance.

2 Moral Optimization and Moral Prioritization Change the Principal-Agent Solution

2.1 Principled Truth-telling is Moral Prioritization

In this section we focus on the nature of trustworthiness required to restore finance. Trustworthiness involves many things: competence, reliability, promise keeping and truth-telling, to name a few. Arguably a lapse in any of these might be important, but we focus on truth-telling because two economic effects flow from lying in finance: It exposes customers to fraud by bankers who understand financial products better than they do, and it forces shareholders to solve the Principal-Agent model under the presumption that bank managers cannot be trusted to give a reliable account of their activities.

Central to this idea of trustworthiness is the notion that someone's truth-telling can be relied upon, even if it ceases to be in their interests, or the interests of those they care about – that is, trustworthy people don't lie even if lying passes a cost benefit analysis. Ordinary common-sense morality recognizes this kind of trustworthiness. It treats the mandate to speak truthfully as an obligation which is qualified or weakened only in unusual and exceptional circumstances. The strength of the obligation to speak truthfully is not sensitive to the standard costs and benefits that lying or deceit may bring; indeed, it is precisely because there are such benefits, that they are costly to others to bear, and that moral disapproval is our main defence against opportunistic liars, that the moral disapprobation that falls on the liar is so strong.

There is good experimental warrant to believe this kind of common-sense morality is empirically important. Erat and Gneezy (2010) provide a clean test for 'lie aversion'. Their experiment is noteworthy because they allowed subjects to improve everyone's financial payoff by telling an untruth, which they called a Pareto White Lie. The mainstream economic solution is straightforwardly determined in this situation because:

> '… a person should weight benefits against harm and happiness against unhappiness. The act of lying in itself carries no bad consequences.' (2010, pg. 724)

In their experiment around one third (36/102) of subjects refrained from lying when given the opportunity to pull off a substantially and mutually-rewarding Pareto White Lie.

With this in mind we distinguish between what we call 'moral optimization' and 'moral prioritization'. Moral optimization is the modelling of ethical behaviour using standard preference-satisfaction techniques, where the content of preferences may include a regard for others (Collard 1978, Becker 1981, Hausman 2012). Moral optimization would always enjoin a Pareto White Lie because it would always pass a cost benefit analysis under any preference weighting of self versus others.[5]

[5] We use 'cost benefit analysis' and 'optimization' interchangeably in this paper because they are equivalent for a wide class of standard problems. Consider the maximization of f under a binding constraint.

$$\max_{x,y} f(x,y) \quad s.t. \quad g(x,y)=c; \quad f_x, f_y, g_x, g_y > 0.$$

We move towards optimality when $df>0$ by, without loss of generality, $dx>0$ (and $dy<0$ since a binding constraint implies $dy/dx=-g_x/g_y<0$). We denote the effect on f of positive dx as marginal benefit (MB), and the effect on f of negative dy as marginal cost (MC), giving the cost-benefit $MB>MC$ criterion to attain optimality.

$$df = f_x dx + f_y dy = f_x dx - f_y \frac{g_x}{g_y} dx > 0 \quad \rightarrow \quad f_x > f_y \frac{g_x}{g_y} \quad \text{or} \quad MB > MC$$

In contrast, moral prioritization rejects the framework of preference satisfaction when modelling some ethical acts, allowing utility-maximizing to be overridden (Williams 1973, Sen 1977). This appears to be a sensible way to explain a reluctance to commit a Pareto White Lie.

It is beyond the scope of this paper to adjudicate whether moral prioritization should be represented in non-standard preferences, or appear as a constraint.[6] However, as a practical matter, since the inception of economics (Smith 1759 and 1776) it has been recognized that certain ethical norms are necessary for a well-functioning market system (Wight 2015), and so it has proven uncontroversial to make them implicit in modelling constraints. That is, some cost benefit analyses of an ethical nature are actually ruled out in the way the models have been set up.

2.2 When is Moral Prioritization Important in a Model?

Hausmann (2012) provides one example of ruling out a cost benefit analysis, namely elementary utility maximization (figure 1A). As shown by the arrow, any consumer that has income taken away from her will accept a decline of utility as a result. What is missing in 1A, but is included in 1B, is a parameterization of theft. The optimal amount of theft could be calculated such that it is decreasing with utility. That is, when utility gets very low theft rises, expanding the budget set.[7]

---

[6] Some features of the debate can be summarized briefly: Hausman (2012) assumes that all ethical behaviour can be modelled by manipulating preferences (by including regard for others in the 'utility function') or by manipulating constraints, which he takes to be Amartya Sen's approach. Hausman himself opts for the former, and therefore believes that ethical behavior currently modelled in constraints should instead be modelled in preferences. For example, absolute prohibitions should be modelled as the satisfaction of lexicographic preferences (vertical or horizontal indifference curves). Having said that, he acknowledges that ethical considerations are sometime built into constraints, as we shall see below. In contrast to Hausman, Williams (1973) proposes that 'the unthinkable' is itself a moral category, thus denying that all morality can be built into preferences or constraints in a rational choice model. 'Entertaining certain alternatives, regarding them indeed as *alternatives,* is itself something that [the moral individual] regards as dishonourable or morally absurd' (Williams 1973, pg. 92). In de-emphasizing rational choice, Williams affirms a legitimate place for emotion in ethics (Augustine 426, Haidt 2013) in contrast to Kant's (1785) highly influential rejection of emotion.

[7] Mathematically, let preferences for X and Y be standard and let theft augment the budget set, albeit at an additively separable utility cost. We choose X, Y and T (theft) in Lagrangian $U(X,Y)-T^2/2+\lambda(m+T-PxX-PyY)$. The first order conditions give $Ux/Px=Uy/Py=T$ which implies that when X and Y are very low, their real marginal utilities are very high, leading to high level of theft. When X and Y are high, theft is low.

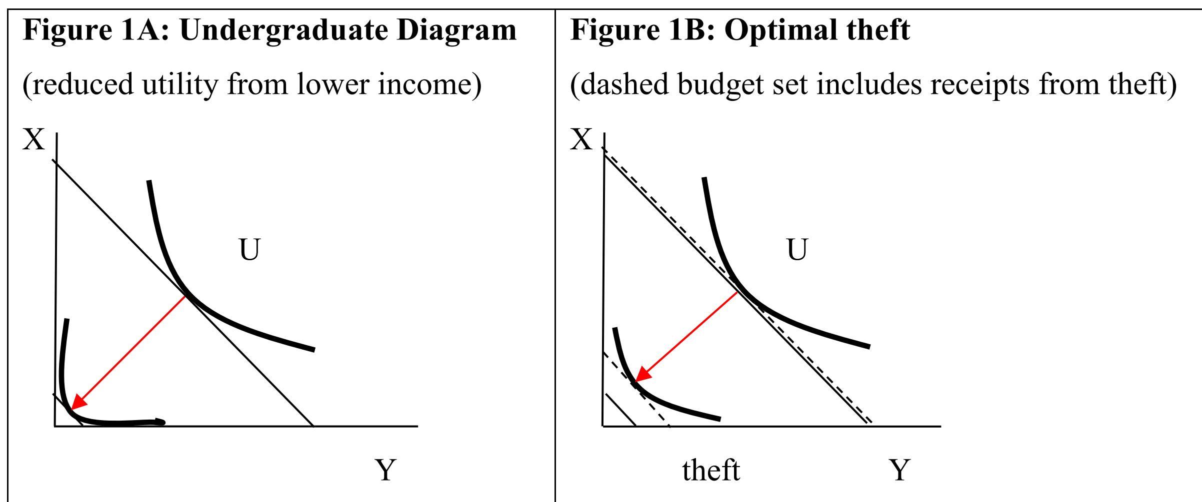


Figure 1A: Undergraduate Diagram (reduced utility from lower income)

Figure 1B: Optimal theft (dashed budget set includes receipts from theft)

We argue that this ruling out of theft is a desirable modelling convention, and we conjecture Smith (1759) might agree. As a *positive* matter, it is in fact the case that the vast majority of transactions in the economy occur without consumers contemplating theft. As a *normative* matter, those being trained in economics are not repeatedly being asked to calculate an optimal amount of theft, as they would be if figure 1B were used. Therefore any tendency for Homo Economicus to function as a moral norm, such that 'optimal' becomes confused with 'ethically good', is rendered harmless by the use of figure 1A.

We now come to a key point in our paper. We would argue that the classic incentive contract of Jensen and Meckling (1976) contains an undesirable modelling convention, judged by the same positive and normative criteria we have just applied. As is well-known, the agent in a Principal-Agent model acts on behalf of the principal, but their action is hidden. The standard analysis then asserts that because their actions are hidden they cannot be contracted on.

Importantly, this is making an unwarranted ethical assumption that reliable truth-telling is impossible. For only in that case does hidden action necessarily imply that hidden action is both undiscoverable and uncontractible. If instead an agent tells the truth, hidden action is discoverable. Furthermore, if they do so reliably hidden action can be contracted on.

We think this is an undesirable convention. As a *positive* matter, we have already noted Erat and Gneezy's (2010) evidence that reliable – in the sense of being Pareto pessimal[8] – truthful communication occurs too frequently to ignore. We could add that some situations in everyday life seem to bear this out.[9] As a *normative* matter, those being trained in economics are

[8] A Pareto pessimal outcome is one which makes all parties worse off. Refraining from a Pareto White Lie makes all parties worse off.

[9] Many people will answer an everyday question truthfully before, or without ever, justifying this with cost-benefit considerations. Trust enables us to know far more than we can independently verify.

repeatedly being presented with the claim that when it is impossible to directly confirm a fact (here, the effort level in Jensen and Meckling's model) no one is trustworthy enough to ask.

Interestingly, despite this implication that truth-telling is impossible in the setup of Jensen and Meckling (1976) – Jensen himself now believes that truth-telling can and must be taught, and that teaching the universal applicability of cost benefit analysis stands in the way of this.

> 'This is a great failure of the curriculum of every business school I know: we teach our students the importance of conducting a cost/benefit analysis in everything they do. In most cases, this is useful – but not when it comes to behaving with integrity [including truth-telling[10]]. In fact, treating integrity …as a matter of cost/benefit analysis virtually guarantees that you will not be a person of integrity.' (Jensen, 2014, page 18)

All this is ethically problematic to the extent that Homo Economicus functions as a moral norm. This potentially negative educative role of Homo Economicus is one possible explanation for the experimental evidence that economists as a group are relatively lacking in pro-social preferences (Bauman and Rose 2011, Cipriani et al. 2009, Frank et al. 1993, Frank and Schultze 2000, Frey and Meier 2003). Some work in this literature controls for problems of adverse selection, whereby certain types choose to study economics, but it still finds a training effect.[11]

2.3 How does the Solution to the Principal-Agent Model Change Under Moral Prioritization?

The kind of representative agent that inhabits the Principal-Agent setup determines the kind of solution that is feasible. Banking has solved the Principal-Agent problem with Jensen and Meckling (1976) which assumes a Homo Economicus representative agent. As we just noted, agents like these are untrustworthy because they cannot be relied upon to tell the truth about hidden action.

Homo Economicus cannot be offered the first-best contract, which is a fixed wage for the optimal effort. If they were offered such a contract, Homo Economicus would subvert it by shirking, all the while claiming to have put in optimal effort. In the event of poor performance they would collect the wage and blame bad luck. Thus it is not optimal for the Principal to offer such a contract, so they must instead offer the agent a contract where good outcomes result in more remuneration than poor outcomes. This is the theoretical justification for the widespread

---

[10] For Jensen behaving with integrity includes truth-telling ('Word-4' in Erhard and Jensen 1998, page 17).

[11] Naturally we accept that the free market liberalism system can function with a modicum of good motives both in the domains of optimal allocation (Arrow and Debreu 1954, Smith 1759 and 1776) and information processing (Hayek 1945), so we accept 'ethically problematic' is a matter of degree. However, recent events in the finance industry suggest that the system was crippled by bad motives that were significant in degree.

practice of incentive contracts in finance, but they are second-best because the risk-averse agent must be compensated for being loaded up with risk.

Other professions solve the Principal-Agent problem using moral prioritization. Rather than *incentivize* good behaviour, the solutions *enjoin* or *command* the agent to act well, by appealing to fiduciary duty.[12] This framework of professionalization makes the most sense if the representative agent is more like the Reasonable Person from tort law, whom we met earlier. In terms of the current discussion we may say the Reasonable Person is someone who exhibits moral prioritization with respect to truth telling. This makes hidden action discoverable and reliability makes the hidden action contractible, so the first-best solution to the Principal-Agent problem is attainable. They are paid a fixed wage for optimal effort which they truthfully disclose, and this approximates the standard practice in many situations.

3 The Non-virtuous Circle

We now describe the decline in trustworthiness in finance with a mixed population containing people who act as either Homo Economicus or Reasonable People. The framework is inspired by Bowles's (2011) stylized evolutionary-game theory of institutional and cultural change.

A bank contains two types of managers Homo Economicus and Reasonable People whose type at any moment is assumed to be known to shareholders, perhaps by their actions or by their words. Shareholders choose whether to motivate the managers by incentive contracts or by appealing to fiduciary duty. Since they know the type, the optimal procedure is to offer an incentive contract to Homo Economicus and pay a fixed wage for optimal effort to a Reasonable Person.

Following Bowles we assume that exposure to the incentive contracts leads to motivation crowding out, which we model as a Reasonable Person turning into Homo Economicus. A classic motivation crowding out example is the study of six day-care centres in Haifa (Gneezy

[12] The fiduciary relationship is a concept of legal theory. The fiduciary is a person who holds a legal or ethical relationship of trust with one or more other parties (person or group of persons). A core legal duty that arises from the fiduciary relationship is the duty of loyalty (Gold and Miller 2016). The fiduciary has to be trusted because they have greater expertise, or have special powers to act on behalf of the person who trusts them. In the case of banking, ordinary customers or shareholders are generally unable to understand the complexity of the financial system, so managers often have more expertise and are given more powers to act, making them fiduciaries. But managers actions are arguably 'hidden' by this complexity too, making them 'agents' in the Principal-Agent framework. This justifies the identification of the fiduciary in legal theory with the agent in Principal-Agent theory, as we have done in the text. This identification makes it clear that the requirement of trustworthiness is met in legal theory by enjoining responsibility (buttressed by legal penalties) whereas it is essentially abandoned in economics. Instead, the principal manipulates incentives to cajole the agent, who is assume to be egotistical, to act as well as possible.

and Rustichini 2000). On the introduction of a fine for parents who were late in picking up their children, the surprising result was that the incidence of lateness increased, more than doubling. Financial incentives apparently had the effect of transforming late-arrival from a morally reprehensible violation of a principle to a problem to be solved with cost-benefit analysis.

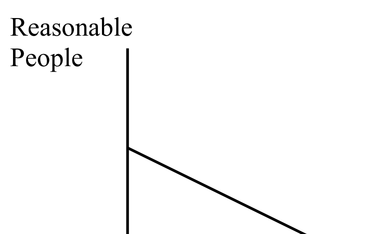
There is now a significant body of international evidence supporting motivation crowding out. An early contribution was due to Titmuss (1970), who argued that financial incentives can evacuate intrinsic motivation, as they did in the day-care example.[13] For a long time, these effects were regarded as curiosities by economists, and with the exception of Collard (1978) there was little attempt to incorporate them into mainstream theory. However, by the close of the last century a substantial body of experimental evidence pointed to the fact that financial incentives could crowd out, and sometimes crowd in, good motivations (Frey 1997).[14]

When incentive contracts change Reasonable People into Homo Economicus, the firm moves down the Motivation Crowding Out (M) schedule in figure 2. We identify the *mentoring power of bonus culture* as one possible crowding out mechanism, whereby employees in ethically negative environments 'will work out for themselves what is valued by the leaders to whom they report' (Salz 2013, page 7) via bonuses. Hogan (2018) puts it this way:

> 'Bonuses not tied to formal outcomes but to the approval of the manager charged with dispersing them, have a largely unrecognised power to change culture quickly. … this was a factor in changing the culture of bankers from 1995 to 2010' (Hogan 2018, slide 18)

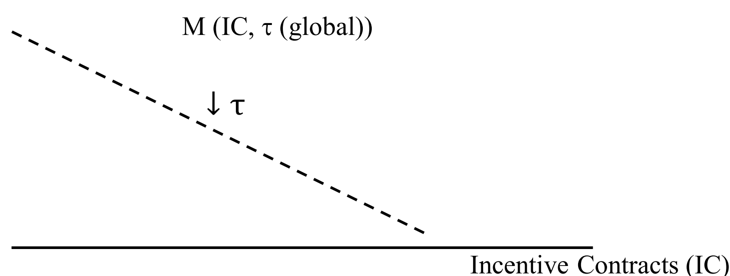
The more prevalent incentive contracts are, the more a culture which favours transactions over relationships, and financial purposes over business purposes (Salz 2013) can spread. We also assume the culture of the bank τ can deteriorate independently with the uptake of global banking culture ('global') resulting in a shift down to the dashed line.

Figure 2: Motivation Crowding Out



[13] Titmuss compared the UK blood donation system, which relied on voluntary contributions, with the US for-profit system, and showed how a non-market system based on altruism can be more effective.

[14] An example of crowding in is given by Bowles (2016). In a Public Goods Game with punishments, agents have to reveal their contributions. In some cultures, low contributions invoke punishment by fellow players, and this holds up the contributions of everyone. On balance, however, crowding out is more common. See Bowles and Polania-Reyes 2012 for an extensive review, and Bowles 2016 for a popular discussion.

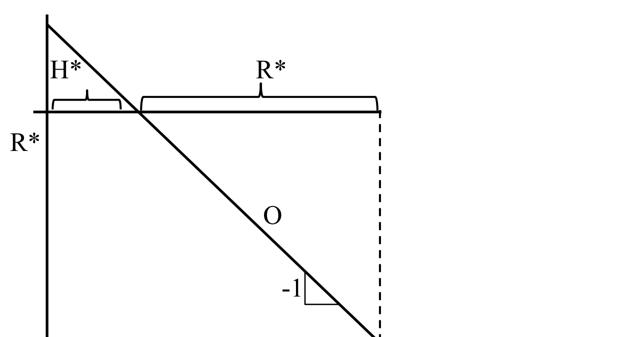


Over the time relevant for the recent deterioration in finance, Hogan describes global banking culture as being, 'all about the money with a focus on short term profitability' (2018, op. cit.). The Salz review describes how Barclays imported it via the investment arm of their business:

> 'Most but not all of the pay issues [inappropriate remuneration] concern the investment bank. To some extent, they reflect the inevitable consequences of determinedly building that business – by hiring the best talent in a highly competitive international market (and during a bubble period) – into one of the leading investment banks in the world. … Nevertheless, based on our interviews, we could not avoid concluding that pay contributed significantly to a sense among a few that they were somehow unaffected by the ordinary rules. A few investment bankers seemed to lose a sense of proportion and humility.' (Salz 2013, page 9)

We now show in figure 3 the optimal contract offers curve (O) which describes how many incentive contracts shareholders should offer. If shareholders are not trained to treat everyone as Homo Economicus, by articles like Jensen and Meckling (1976), then the number of incentive contracts offered could be low, even in an ethically compromised bank, like Barclays was.[15]

Figure 3: Optimal Contract Offers



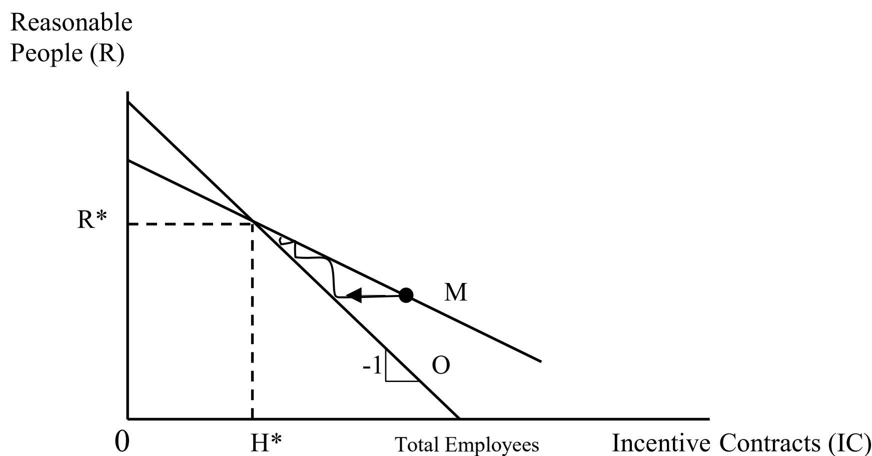
[15] We have no reason to doubt that Salz's observation about Barclays '… the overwhelming majority [of employees] are focused on the bank's customers and doing their best for them' (Salz 2013, page 5) is applicable elsewhere. That is, we might suppose that most bankers are Reasonable People.

0 Total Employees Incentive Contracts (IC)

If R* of total employees on the vertical axis are Reasonable People, they exhibit moral prioritization with respect to truth telling. Thus, they can be offered a first-best fixed wage contract and it is optimal for the shareholders to do so. This leaves a balance of employees H* (total employees minus R*) who act like Homo Economicus and must therefore be offered an incentive contract. There is no alternative to this because Homo Economicus will subvert the first-best solution, as described above. The slope of the optimal offers curve is minus one, because, for every Reasonable Person who turns into Homo Economicus, an extra incentive contract must be offered.

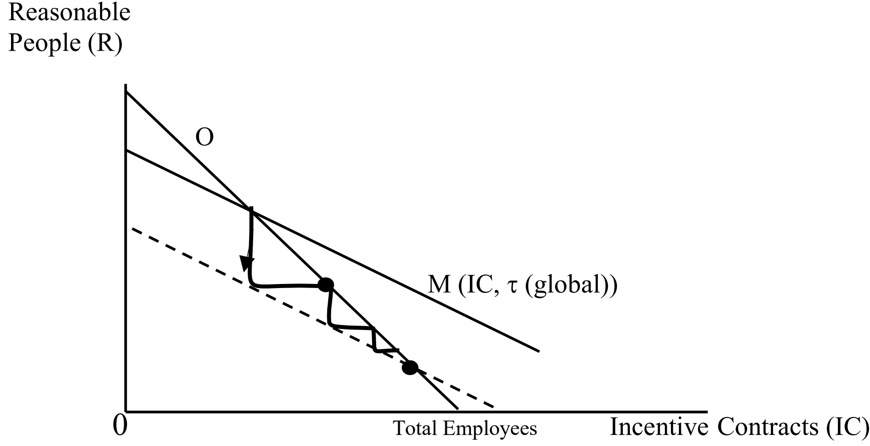
The crossing of the two curves creates a temporary cultural-institutional equilibrium (Bowles 2010). This is stable since a perturbation in incentive contracts (shareholders mistakenly offering too many) will result in a move down the M curve as some Reasonable People turn into Homo Economicus, as shown at the dot in figure 4. However, because motivation crowding out is not complete (that is, not everyone who is offered a contract becomes untrustworthy), when the shareholders realize their mistake they can offer a fixed wage contract again to those who have not been transformed into Homo Economicus, which in turn undoes some motivation crowding out, which allows the bank to cancel some more incentive contracts, and so on, taking us back to the original equilibrium.

Figure 4: Temporary Cultural-Institutional Equilibrium



We now consider the situation where there is an exogenous shift downwards in the MCO curve due to a deterioration in ethical culture from global banking.

Figure 5: Deterioration in Ethical Culture from Global Banking



We take the total derivative at the equilibrium O=M; we define 'global' so that it enters τ linearly with coefficient of unity; and we note from before that the slope of O is minus one:

$$dO = dM$$

$$O_{IC}dIC = M_{IC}dIC + M_{\tau}d\tau$$

$$-dIC \quad = M_{IC}dIC + M_{\tau}dglobal$$

$$\frac{dIC}{dglobal} = \frac{-M_{\tau}}{1 + M_{IC}} < 0 \qquad (1)$$

$$M_{\tau} > 0 \text{ and } M_{IC} < 0\,. \qquad (2)$$

In the first instance, there is a direct effect of a negative shock to global banking culture (dglobal<0) which takes the firm to the first dot. Intuitively, if the global banking culture makes agents less trustworthy, the shareholders are forced to offer more incentive contracts for every additional agent who becomes untrustworthy as Homo Economicus.

The strength of the effect from there depends on the existence of motivation crowding out, which in our model is parameterized by the slope M_{IC}. If the M lines were flat, this would be the final equilibrium. However, since M_{IC} is non-zero, extra contracts that have been offered up to the first dot crowd out motivation and turn more Reasonable People into Homo Economicus, shown on the diagram as the decline from the first dot to the dashed line. Thus, more incentive contracts need to be offered, and so on, to the final dot.

We call this negative feedback between the initial offers of incentive contracts, on the one hand, and the demand for further contracts that this creates, on the other, a *non-virtuous circle*. If the denominator of (1) approaches zero, the non-virtuous circle could succeed in transforming all

Reasonable People into Homo Economicus. However, a crisis or government intervention or both (none of which are modelled) might interrupt the model before this happens.

We earlier noted that trustworthiness involves many dimensions such as truth-telling: competence, reliability, and promise keeping. While we have emphasised truth-telling about manager effort within the Principal-Agent model as a mechanism that propagates incentive contracts, we believe that a lack of trustworthiness has general effects.

The most direct effect of untrustworthiness involves the mistreatment of clients, who are often far less informed about financial products than bankers are. Increased competition and deregulation in the latter decades of the twentieth century tended to reduce the number of repeated interactions between bankers and clients as the markets swelled in size and, interestingly, as the cultural dominance of particular groups, like White Protestants in the US, waned (Hill and Painter 2015).[16]

We know from the Folk theorem that Homo Economicus agents require repeated interactions to support cooperation. So the joint impact of reduced interactions with particular clients, plus a growing proportion of agents becoming like Homo Economicus through the non-virtuous circle, was always likely to erode general integrity.

> 'In traditional investment banking … reputational capital was as important as financial capital. The people who ran investment banks understood that temptation to take advantage of customers for short-term gain was always present, but they made a concerted effort to control it.' Hill and Painter (2015, page 101)

In one interesting anecdote, a client of Salomons in the 1970s paid the equivalent of a hefty 100 basis points for a bond trade facilitated by the organization. In a development surprising by today's standards, the managing partner returned part of the money and castigated the young bond trader with the words 'Salomon Brothers does not make such a profit' (op. cit. page 101). Hill and Painter wryly note that the trader's actions were not even illegal whereas, in the ethically transformed decades that followed, the same firm under a different partner broke the law in the 1991 US Treasury bid scandal. During these turbulent times Goldman Sachs emails

[16] It is not for nothing that UK bankers prior to deregulation were known as 'gentlemen bankers'. For more than a century after the 'panic of 1866' (Bagehot, 1866), the stability of the British 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 City.

circulated by some managing directors referred to their clients as ‘muppets’.

In model language we might say that there are significant externalities associated with financial professionals changing from Reasonable People to Homo Economicus, especially if the newly transformed agents eventually come to occupy key positions.

4 Restoring Trust in Finance

In this section we make a number of proposals for the restoration of trust in the finance industry. We have chosen the word ‘restore’ carefully, so as to avoid being utopian. Perhaps there are ways to ‘transform’ finance so that it is more ethical that other professions, or more ethical than it ever was, but that is more than we know.

The model of section 3 is fully symmetric, implying that a *positive* shock to ethical culture could lead to a *virtuous* circle, reversing the damage of the last few decades. However, there are two matters of concern to be raised: First, we do not know very much about the timeframe for the reversibility of motivation crowding out. Interestingly, in the day-care example (Gneezy and Rustichini 2000) where fines perversely increased bad behavior, the incidence of late coming stayed high even after the fine was removed. That is, a cost benefit mindset seemed ‘sticky’ and we haven’t modelled this. Second, we cannot see anything comparable in magnitude to the 1980s change in global banking culture, but of the opposite sign, to set off such a virtuous circle. It is true that regulations have been tightened in many jurisdictions, but regulations may be ineffective at changing character, or at least their effectiveness may not be clear for some time.

If it is granted that most bankers are well represented by Reasonable People, and that Homo Economicus tends to be concentrated in certain business units, one way to depopulate these business units of Homo Economicus is to be conservative about pay structures. If Gneezy and Rustichini (2000) is to be believed on this matter – that Homo Economicus does not quickly change back into a Reasonable Person – it may be possible to un-incentivise them so that they leave of their own accord, naturally after the appropriate cost benefit analysis.

Change could also be assisted by altering some norms in the workplace. Where there is a reasonable prospect that agents act like Reasonable People, professional codes of conduct can be encouraged through rules, though professionalism is about more than rules. Positively, the professional is enjoined to exhibit what Downie calls beneficence, which includes truth-telling and loyalty (Downie 1990, Gold and Miller 2016).

On a practical level there are a number of ways that professionalization functions. The most obvious is through a professional body's self-certification of its members. One of the tasks of a self-certified regime would be review performance pay, and we would recommend less reliance on it, based on the phenomenon of motivation crowding out. Self-certification has a long history in environments without external authorities. This is clearly not the case in the current environment, but it is equally true that internally generated self-denying ordinances are necessary. Hill and Painter (2015) describe some hopeful signs that banks are taking responsibility for ethical reform, and developing a more professional ethical culture.[17]

There is a lively debate on whether professionalism requires a degree of monopoly power to operate effectively. Certainly the system of 'gentleman bankers' in the UK up until the 1980s deregulation was, on the one hand, relatively uncompetitive, yet on the other it was considered to be more trustworthy than banking in recent decades. Taking a much longer view, Krause (1996) argues that professions – like medicine, law, engineering and the university professoriate – are essentially outgrowths of the medieval guilds by virtue of possessing the same 'guild powers' of association, workplace control and presenting a united front in dealing with the state, all made possible by a degree of market power.

Guilds are generally regarded negatively by economic historians, with Olgvie (2011) following in the footsteps of Smith (1776) who denounced them as a conspiracy against the public. But Epstein and Prak (2008) counter that there are many tasks for which it is helpful to specify rules for the ways in which work is done, and that creating communities of practice enable valuable tacit knowledge to be transferred from person to person. Sox (2007) critiques the contemporary US medical profession by drawing on this 'guild model'. He argues that professions act as a buffer for consumers against a controlling state or commercial interests more generally, echoing Krause's (1996) reading of Marx, Weber and Durkheim. Paradoxically, he says, some monopoly profits by professionals gives them the freedom to enshrine service rather than profit orientation in its code of ethics.

Some of Sox's critique sounds plausible, especially when we recognize the phenomenon of motivation crowding. Extreme focus on profitability is indeed likely to crowd out ethical concerns, and extreme focus on profitability is a natural consequence of operating in highly

[17] They list Deutsche Bank's promises of 'concrete cultural change' on its website; Goldman Sach's business standards committee which emphasises service to clients and reputational sensitivity; and the response by Barclays to the Salz review. Although it was birthed earlier (following the 1997 Asian Financial Crisis) the Toronto Centre (https://www.torontocentre.org) has had an active role in promoting financial sector integrity, though it operates primarily through regulators.

competitive environments. This is true of any workplace, but perhaps the finance profession is especially prone to such an extreme financial focus, given the nature of the day-to-day tasks are already ‘about money’. Such a focus may crowd out ethical motivations even more than in other industries.

Yet for all this we are not proposing a general clamp down on finance, greatly reducing its size and creating large monopoly rents for a favoured few firms. There is a great deal of policy middle ground to cover without returning to ‘gentlemen bankers’ era, and there is currently a lot of healthy pressure on banks to recover good reputations. Once earned, these may be more of a barrier to unethical behaviour than they have been in the recent past.

Finally, we argue that economics training could be altered so that students are exposed to representative agents that are different to Homo Economicus. For example, the utility maximization problem of figure 1B could be taught assuming the Reasonable Person who displays moral prioritization with respect to stealing, and Homo Economicus who optimizes by choosing the optimal amount of theft. Another step would be to teach agency theory differently, bringing the concerns from Jensen (2014), quoted above, into conversation with Jensen and Meckling (1976), much as we have done in this paper. On this note, Jensen is an instructive figure to study because he began as a canonical contributor to agency theory, became an instigator of some of the changes at Harvard Business School in the 1980s which Khurana (2007) criticizes, and, finally, has ended up as an advocate of moral prioritization for matters of integrity.

One thing we have reason to doubt is that the ethical challenges we have discussed can be solved by further general deregulation. It is an appealing idea that any kind of firm that behaves unethically will be driven out of a competitive market, in much the same way that a seller trying of blemished apples would be. However, this has not proven to be a reliable metaphor in finance. The fundamental informational asymmetries in finance make it a special case – its irreducible complexity requires that society asks for principled agents who act with the same integrity as other professions.

References

Anderson, E. (2001), ‘Symposium on Amartya Sen’s philosophy: 2 Unstrapping the straitjacket of ‘preference’: A comment on Amartya Sen’s contributions to philosophy and economics’, *Economics and Philosophy*, 17, 21-38.

Arrow, K., and Debreu, G. (1954), ‘Existence of an equilibrium for a competitive economy’, *Econometrica,* 22 (3): 265–290.

Augustine, (426), *The City of God*, Rome.

Bacharach, M., Guerra, G., and D. Zizzo, (2007), 'The self-fulfilling property of trust: An experimental study', *Theory and Decision*, 63, 349-388.

Bagehot, W. (1866), 'What a Panic Is and How It Might Be Mitigated', *The Economist*, vol. 24 number 1185, 12 May, page 554.

Bauman, T. and E. Rose (2011), 'Selection or indoctrination: Why do economics students donate less than the rest?', *Journal of Economic Behavior and Organization*, 79(3), 318-327.

Becker Gary S.(1981), *A Treatise on the Family*, Harvard: Harvard University Press.

Belk, R. and Wallendorf, M. (1990), 'The sacred meanings of money', *Journal of Economic Psychology*, 11(1), March, 35-67.

Bentham, J., (1948/1789), *An Introduction to the Principles of Morals and Legislation*, ed. W. Harrison. Oxford: Blackwell.

Berger, A., Klapper, L.F., and Turk-Ariss, R., (2009), 'Bank competition and financial stability', *Journal of Financial Services Research*, 35, 99-118.

Bowie, N. (2017), *Business Ethics: A Kantian Perspective*, 2nd edition, CUP.

Bowles, S. (2011), 'Is liberal society a parasite on tradition', Philosophy and Public Affairs, 39(1), 46-81.

Bowles, S. (2016), *The Moral Economy: Why good incentives are no substitute for good citizens*, Yale University Press, New Haven.

Bowles, S. and S. Polania-Reyes (2012), 'Economic incentives and social preferences: substitutes or complements', Journal of Economic Literature, 50(2), June, 368-425.

Cipriani, P., Lubian, D. and A. Zago (2009), 'Natural born economists?', Journal of Economic Psychology, 30, 455-468.

Cohn, A., Fehr, M. and Marechal (2014), 'Business culture and dishonesty in the banking industry', *Nature*, 516, December, 86–89.

Collard, D. (1978), *Altruism and Economy: A study in non-selfish economics*, OUP, New York.

Danckert, S. (2018), ''Absolutely and utterly disgusting': Westpac's advice led to couple's ruin', *Sydney Morning Herald*, 19 April.

Downie, R. (1990), 'Professions and Professionalism', Journal of Philosophy of Education, 24(2), December, 147-159.

Durkheim, E. (1915), *Elementary Forms of Religious Life*, tr. J. W. Swain, Allen and Unwin.

Edmonds, T. (2013), *The Independent Commission on Banking: The Vickers Report,* House of Commons Library, SNBT 6171, Business and Transport Section, 30 December.

Elster, J. (1983), *Sour Grapes: Studies in the Subversion of Rationality,* CUP

Epstein, S. R. and M. Prak (2008), *Guilds, Innovation, and the European Economy, 1400-1800.*

Erat, S. and U. Gneezy (2010), 'White lies', *Management Science*, 58(4), 723-733.

Erhard, W. and M. Jensen, (1998), 'Putting Integrity into Finance: A Purely Positive Approach', NBER Working Paper 19986, http://www.nber.org/papers/w19986 .

Fligstein, N. and A. Roehrkasse, (2016), 'The causes of fraud in the financial crisis of 2007 to 2009', *American Sociological Review*, 81(4), 617-643.

Fourcade, M., Ollion, E. and Y. Algan, (2015), 'The Superiority of Economists', *Journal of Economic Perspectives*, 29(1), 89-114.

Frank, R., Gilovich, T. and D. Regan (1993), 'Does studying economics inhibit cooperation?', *Journal of Economic Perspectives*, 7(2), Spring, 159-171.

Frank, B. and G. Schultze (2000), 'Does economics make citizens corrupt?', *Journal of Economic Behavior and Organization*, September, 101-113.

Frey, B. (1997), *Not just for the money: an economic theory of personal motivation*, Elgar, Cheltenham.

Frey, B. and Meier S. (2003), 'Are political economists selfish and indoctrinated? Evidence from a natural experiment', *Economic Inquiry*, July, 448-462.

Gneezy, U. and Rustichini, A. (2000), 'A fine is a price', *The Journal of Legal Studies*, 29(1), 1-17.

Gold, S. and P. Miller (2016), *Philosophical Foundations of Fiduciary Law*, OUP.

Haidt, J. (2013), *The Righteous Mind*: *Why Good People are Divided by Politics and Religion*, Random House, New York.

Haldane, A. G., Brennan, S. and Madouros, V. (2011) 'What is the contribution of the financial sector: miracle or mirage?', *The Future of Finance: The LSE Report*. London School of Economics and Political Science.

Hausman, D. (2012), *Preference, value, choice and welfare*, CUP, New York.

Hayek, F. (1945), 'The Use of Knowledge in Society', *American Economic Review*, 35(4), 519-530.

Hill, C. A. and Painter, R. W. (2015), *Better Bankers, Better Banks*, University of Chicago Press.

IMF (2009), 'From Recession to Recovery', Chapter 3, *World Economic Outlook*, April.

Hogan, W. (2018), 'Prospects and Challenges for Australia's Financial System', Centre for Policy Market and Design Workshop on Banking, Health and Education, University of Technology Sydney, November 22, https://www.uts.edu.au/sites/default/files/2019-01/CPMD%20Presentation%20%20Banking%20Nov%2018%20update.pdf

HoLC, (2013), House of Lords Commission on Banking Standards in the UK, *Changing Banking for Good*, 12 June, http://www.parliament.uk/bankingstandards

Jaffer, S., Morris, N., Sawbridge, E. and D. Vines, (2014), 'How changes to the Financial Services Industry eroded Trust' in ed. Morris, N. and D. Vines, (2014), *Capital Failure: Rebuilding Trust in Financial Services.* Oxford: Oxford University Press.

Jensen, M., (2014), 'Integrity: Without it Nothing Works' (April 6, 2014). Rotman Magazine: The Magazine of the Rotman School of Management, April 6, Fall 2009, 16-20; *Harvard Business School NOM Unit Working Paper* 10-042; Barbados Group Working Paper No. 09-04; Simon School Working Paper No. FR 10-01. Available at SSRN: https://ssrn.com/abstract=1511274

Jensen, M., and Murphy, K. (1990), 'CEO Incentives: It's Not How Much You Pay, But How,' *Harvard Business Review,* 68(3), 138–153.

Jensen, M. and W. Meckling, (1976). 'Theory of the firm: Managerial behavior, agency costs and ownership structure', *Journal of Financial Economics* (October), 3(4), 305–360.

Kant, I., (1785, 2005) *Groundwork for the metaphysics of morals*, tr. Thomas Kingsmill Abbott (1829-1913), edited with revisions by Lara Denis (1969-). Peterborough, Ont.; Orchard Park, NY: Broadview Press.

Khurana, R. (2007), *From higher aims to hired hands: The social transformation of American business schools and the unfulfilled promise of management as a profession*, Princeton University Press.

Krause, E. A. (1996), *Death of the Guilds*, Yale University Press, New Haven.

Lydenberg, S. (2014), 'Reason, Rationality and Fiduciary Duty', in *Cambridge Handbook of Institutional Investment and Fiduciary Duty*, eds. James P. Hawley, Andreas G. F. Hoepner, Keith L. Johnson, Joakim Sandberg, and Edward. J. Waitzer, 287-99, CUP, Cambridge.

Martin, I. (2016) *Crash, Bang, Wallop: The Inside Story of London's Big Bang and a Financial Revolution that Changed the World*. London: Hodder and Stoughton.

Mayer, C. (2013), *Firm Commitment*: *why the corporation is failing us and how to restore trust in it*, OUP.

Macintosh, N. B., (2007),'Accounting - Truth, Lies, or 'Bullshit?': A Philosophical Investigation', August 13, Available at SSRN: https://ssrn.com/abstract=1006875 or http://dx.doi.org/10.2139/ssrn.1006875 .

Mill, J. S., (1974/1843), *The Collected Works of John Stuart Mill, Volume VIII - A System of Logic Ratiocinative and Inductive,* Part II, Chapter 9, ed. J. Robson. Routledge, London.

Morris, N. and D. Vines, (2014), *Capital Failure: Rebuilding Trust in Financial Services.* Oxford: Oxford University Press.

Noe, T. and H Peyton Young, (2014), 'The Limits to Compensation in the Financial Sector', in *Capital Failure: Rebuilding Trust in Financial Services.* Oxford: Oxford University Press.

Offer, A. (2014), 'Narrow Banking, real Estate and Financial Stability in the UK, c. 1870-2010' in British Financial Crises since 1825, eds. N. Dimsdale and A. Hotson, chapter 9, 158-173.

Olgvie, S. (2011), *Institutions and European Trade: Merchant Guilds, 1000–1800*, CUP.

Peters, G. M. (1987), 'The Use of Lies in Negotiation', Ohio State Law Journal, 48(1), un-paginated.

Parliamentary Library (Department of the) (1995), 'Foreign Bank Policy in Australia', Parliamentary Research Service Current Issues Brief No. 8 1995-96, Australian Government Printer.

Salz, A. (2013), 'Salz Review: An Independent Review of Barclays' Business Practices', April.

Sen, A. (1977), 'Rational Fools: A Critique of the Behavioral Foundations of Economic Theory', *Philosophy and Public Affairs*, 6(4), (Summer), 317-344.

Smith, A. (1759/2002), *The Theory of Moral Sentiments*, ed. Knud Haakonssen, Cambridge University Press, Cambridge.

Smith, A. (1776/1999), *The Wealth of Nations*. Penguin, London.

Sox, H. C. (2007), The Ethical Foundations of Professionalism: A Sociological History, *CHEST (medical ethics),* Vol 131, 1532–1540.

The Economist, (2017), 'Barclays and four former executives are charged with fraud', print edition, Finance and Economics section, 22nd of June.

Titmuss, R., (1970), *The Gift Relationship: From Human Blood to Social Policy*, New Press

Turner, A., (2010), 'What Do Banks Do? Why Do Credit Booms and Busts Occur? What Can Public Policy Do About It?' in *The Future of Finance*, 3-63, London School of Economics.

US Senate, (2010), 'Testimony: Lloyd Blankfein Chairman Goldman Sachs', Senate Permanent Subcommittee on Investigations, Chair: Senator Carl Levin, 28 April, https://www.youtube.com/watch?v=jS9r1Dk-Zg8

Wight, J. (2015), *Ethics in Economics*, Stanford University Press, California.

Williams, B. A. O. (1973), *A Critique of Utilitarianism* in J. J C. Smart and Williams, Utilitarianism, for and against, 75-150, Cambridge, CUP.

Vaughan versus Menlove (1837), 3 Bing NC 468, 132 ER 490 (Court of Common Pleas), sitting judges; Tindal, C. J., Park, J. and J. Vaughan.

Wolf, M. (2010),'Why and How Should We Regulate Pay in the Financial Sector' in *The Future of Finance*, 227-237, London School of Economics.