This is the peer reviewed version of the following article:

Restoring Trust in Finance: From Principal–Agent to Principled Agent

Published in the

The Economic Record

First published: December 2019

which has been published in final form at

https://onlinelibrary.wiley.com/doi/abs/10.1111/1475-4932.12494

This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving."

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

Gordon Menzies, University of Technology Sydney
Thomas Simpson, Blavatnik School of Government, University of Oxford
Donald Hay, 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
There is evidence that many contemporary financial firms are untrustworthy, relative to other professions and to the past. We seek a partial explanation in the overuse of incentive contracts, and an implicit moral training for professionals which misuses cost-benefit analysis for matters of integrity. The most straightforward representation of Economic Man chooses an 'optimal' amount of deceit (moral optimization), in spite of experimental evidence that some agents rule out untruthfulness *a priori* (moral prioritization). We suggest that the restoration of trustworthiness may be aided by less reliance on incentive contracts and less use of cost-benefit analyses for matters of integrity.

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 managers have power over shareholders and customers, and therefore a social responsibility to

[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. We also thank without implication Peter Anstey, Adam Bennett, 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 Carl Rhodes. All correspondence received by gordon.menzies@uts.edu.au at UTS, PO Box 123, Broadway, Sydney.
[1] In this article 'banking' covers the activities of all kinds of financial intermediaries and 'manager' stands for an individual or group in charge of a financial intermediary.
[2] In this article 'banking' covers the activities of all kinds of financial intermediaries and 'manager' stands for an individual or group in charge of a financial intermediary.

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 many contemporary financial firms are 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)

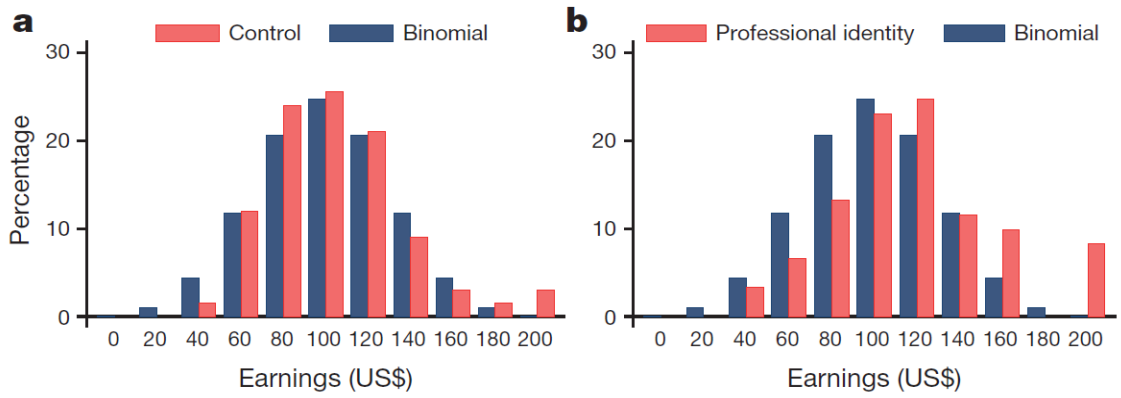
A decade later, the 2017-2019 Australian Royal Commission into Misconduct in the Banking, Superannuation and Financial Services Industry has also exposed instances of untrustworthiness. One testimony relates how a client had paid fees to construct a financial plan based around living in an investment property. She felt misled, because this is 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 behaviour is selected for media exposure and courtroom scrutiny. However, there is experimental evidence which raises concerns too. In Cohn et al. (2014), 128 bank employees from a large international bank were given a coin flipping task and rewarded for the 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 advance. Prior to the coin task, the control group of bankers was asked about the use of their leisure time and their hobbies, priming them to think in terms of their domestic identity. The treatment group of bankers was asked about their work life, priming them with their professional identity.

The experimental subjects flipped the coin out of sight, mirroring hidden action by the agent in the classic Principal-Agent setup (Jensen and Meckling 1976). No individual deceit could be detected, but group cheating was detectable when the results were compared with the binomial distribution of payoffs, which are bell-shaped for truthful disclosure (Figure 1). The individually-rational play clearly involves some degree of cheating.

Figure 1: The untrustworthiness of bankers



In figure 1, 'a' is the control group and 'b' is the treatment group. The dark binomial distribution bars represent the expected frequencies of payoffs if all tosses are reported truthfully, and the light bars are the findings. 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 focused on their domestic identity.[3]

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. We are not so naïve as to think that other professionals cannot ever lie, but we instead take this experiment as evidence that bankers have a trustworthiness problem relative to their peers in other professions.

Furthermore, bankers appear to have a relative trustworthiness problem when compared to the past, at least in the UK.[4] In some ways this is a hopeful observation. Different behaviour in the past suggests that a restoration to previous levels of trustworthiness is an evidence-based aspiration (for the UK see Haldane et al. 2011, Jaffer et al. 2014, Martin 2016, Mayer 2013, Morris and Vines 2014, Offer 2014, The Economist 2017, Turner 2010, Woolley 2010).

It is not uncommon to make a general claim that contemporary financial firms are

[3] The experiment implies a moral boundary between work and home, a theme explored within the Ethics of Care literature (Tronto 2013).

[4] We explore this in appendix 1, relying primarily on the well-documented and widely-accepted ethical demise of the City (of London), though we also refer to the US and Australia (Fligstein and Roehrkasse 2016 and Hogan 2018). To make the point analytically, we apply Bowles's (2011) stylized evolutionary-game theory model of institutional and cultural change as a conjecture explaining the ethical demise of the City. We call the negative feedback loop based on Bowles's model, whereby untrustworthiness requires widespread implementation of incentive contracts which themselves promote untrustworthiness, a 'non-virtuous circle'. We also describe in Appendix 2 how competition policy is difficult to implement successfully in finance, implying deregulation may have had some role to play in the historical unravelling of trustworthiness (Berger et al. 2009, IMF 2009, Noe and Peyton Young 2014 illustrated by Wolf 2010).

untrustworthy relative to other professions and to the past, and to hope for change. However, in this paper we wish to be more specific than this. By highlighting two particular reasons why untrustworthiness might be prevalent we are then able to recommend corresponding strategies to help restore trustworthiness to levels attained in other professions and in the past.

One potential reason is that finance is a profession that is 'all about money' and so the literature on money priming has potential relevance (Vohs 2015). For example, Belk and Wallendorf (1990) claim that money in contemporary society is (pg. 36) 'revered, feared, worshipped, and treated with the highest respect', and it is not difficult to imagine negative effects on motivation that might arise. For example, in Oliver Stone's 1987 film *Wall Street* the anti-hero, Gordon Gekko, is a ruthless corporate raider who delivers the phrase 'greed is good' in a much down-loaded excerpt.[5]

Whatever plausibility this has, however, we are unable to pursue it in this paper because we are hampered by a lack of evidence. Studies showing money priming effects, where money is introduced in an incidental way in an experimental environment rather than as an incentive, are embroiled in the replication crisis. It is too early to tell if any money priming effects will survive scrutiny (Klein et al. 2014, Roher et al. 2015 and especially Vadillo et al. 2016).

The related motivation crowding out literature is unscathed by the replication crisis, and it suggests that money has the effect of attenuating moral motivations when it appears as an incentive. A classic motivation crowding out example is the study of six day-care centres in Haifa (Gneezy 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 had the apparent effect of transforming late-arrival from one kind of moral entity to another – from a morally reprehensible violation of a principle to a decision problem to be solved with cost-benefit analysis.

Financial incentives are widely applied in banking, so this is relevant for the prevalence of trust and trustworthiness. Based on the evidence for motivation crowding out, we suggest that there may have been an overuse of financial incentives. It follows that using them less may help restore trust.

A third explanation focusses on the type of training participants receive, either prior to entry in a university degree or through professional study (such as an MBA). Successful banking

[5] https://www.youtube.com/watch?v=VVxYOQS6ggk

requires that both loans and insurance contracts be monitored in the light of evolving conditions that operate at the firm level and at the level of the general business environment – the domains of microeconomics and macroeconomics. Important banking positions are therefore occupied by those with an economics training, and there is a body of evidence suggesting that economists lack 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, and concludes that both adverse selection and training contribute to the measured preferences of economists.

We do not pursue an adverse selection explanation in this paper. 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 it is not clear how much adverse selection will matter in the finance industry. Indeed, we think it likely that people who choose training in manufacturing, pharmaceuticals, telecommunications and information technology are more profit-driven than, say, nurses, but the adverse selection operating in these instances was not enough to show up as untrustworthiness for these professions in Cohen et al. (2014), or to show up historically in scandals of a banking-scale proportions with banking-scale frequency. Thus, while we suspect some adverse selection is in operation, we lack evidence about its importance relative to other professions.

Instead, we explore a training effect recently highlighted by the original popularizer of the Principal-Agent framework (Jensen and Meckling 1976).

> '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. 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)

The implication is that teaching business students to model ethical choices using cost benefit analysis amounts to a very specific moral training, with a learning objective that optimality is always desirable.[6] Jensen suggests that this a normative 'great failure', but the work of Erat

[6] Consistent application of cost benefit criteria guarantees optimality for well-behaved functions, so the two frameworks point to the same choice. For example, suppose we maximize *f* under a binding constraint.

and Gneezy (2010) goes further, suggesting that this modelling might even may fail positively as a generalization of 'the way things actually are'. They find that a significant proportion of agents tell the truth as a matter of principle, even when it hurts everyone, including themselves.

In this paper call the use of cost-benefit analysis in moral environments moral optimization, and we outline an alternative account, called moral prioritization, where a moral principle overrides utility maximization. Jensen's comments above align him to Sen (1977), who argues that sometimes commitments can and should override utility maximization. Thus, what we call moral prioritization is a Sen-style commitment to moral principle.

The paper is structured as follows. In section 2 we formalize Jensen's (2014) perspective by distinguishing moral optimization from moral prioritization. In section 3 we show the transformational significance of this for Jensen's own Principal-Agent framework. If moral prioritization is feasible, Jensen and Meckling (1976) becomes a means of measuring the cost of adopting a second-best incentive contract when the first-best full-information contract is feasible, rather than a prescription to always use incentive contracts when there is hidden action. In section 4 we rely on the motivation crowding out literature to suggest that financial incentives can be harmful. It follows from sections 3 and 4 that a reduced reliance on incentive contracts would yield a double dividend of promoting efficiency (if truth-telling could be relied upon) and restricting the undesirable effects of motivation crowding out. We also refer in section 4 to the literature which suggests that an economics training might crowd out pro-social motives, echoing Jensen's (2014) point. In section 5 we discuss how practical it is to restore trust in finance to levels attained in the past or in other professions. As well as a reduced reliance on incentive contracts, there is some scope for altering professional norms in the workplace, and reforming economics training in such a way that students are exposed to other representative agents alongside Economic Man. However, we are not utopian – our discussion is about the restoration of finance, not its transformation.

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

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

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

2 Moral Optimization and 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 important economic effects flow from lying: It exposes customers to fraud by bankers who understand financial products better than they do, and (as explained in the next section) it forces shareholders to relate to managers according to the Principal-Agent model, so that socially inefficient incentive contracts must be offered to bank managers who cannot be trusted to give a reliable account of their activities.

Central to the idea of trustworthiness is the notion that someone's commitment can be relied upon, even if it ceases to be in their interests, or the interests of those they care about. When that is applied to lying, a trustworthy person will tell the truth even if, as Jensen (2014) might put it, lying passes a cost benefit analysis.

Ordinary common-sense morality recognizes this phenomenon 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. Common-sense morality contrasts with utilitarianism, however, which countenances lying if the consequences of a lie are sufficiently beneficial that it outweighs the costs.

We have already mentioned our experimental warrant to take the phenomenon of common-sense morality seriously. 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.[7] The mainstream economic solution is straightforwardly determined in this situation because:

> 'The utilitarian approach [moral optimization] … argues that one should lie in such situations. … a person should weight benefits against harm and happiness against unhappiness. The act of lying in itself carries no bad consequences.' (2010, pg. 724)

[7] This terminology draws on the standard definition of a Pareto improvement.

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 evidence in mind we create a distinction 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 includes a regard for others to a greater or lesser extent (Collard 1978, Becker 1981, Hausman 2012). In contrast, moral prioritization rejects the framework of preference satisfaction when modelling some ethical acts, allowing profit- and utility-maximizing to be overridden (Williams 1973, Sen 1977). The essence of moral optimization is balancing one's own interests against another's using cost benefit analysis. The essence of moral prioritization is over-riding this balancing of interests.

So, to follow through with the example of truth-telling, a cost benefit analysis (moral optimization) recommends an optimal amount of deceit, if the benefits to me, or those I love, are high enough. But decisions about lying need not be made in this manner. Individuals might act according to the principle: 'You should not lie!' as one third of Erat and Gneezy's (2010) subject pool appeared to do. For such individuals the principle trumps evaluation of costs and benefits. The fact that some people do not act according to the moral principle does not count against the phenomenon that many do. Moral prioritization is a principled eschewing of cost benefit analysis, even when its components include shared interests and empathy.

On a theoretical level, Sen (1977) laid the groundwork for legitimizing moral prioritization within mainstream economic theory by suggesting that preference satisfaction tries to do too many things.

> 'A person is given *one* preference ordering, and as and when the need arises this is supposed to reflect his interests, represent his welfare, summarize his idea of what should be done, and describe his actual choices and behaviour. Can one preference ordering do all these things? A person thus described may be "rational" in the limited sense of revealing no inconsistencies in his choice behaviour, but if he has no use for these distinctions between quite different concepts, he must be a bit of a fool.' (Sen 1977, pp. 335-336, original italics)

Engelen (2017) is one example of someone who, having rejected the proposition that preference satisfaction models cover all behaviour, turns to what Sen (1977) calls commitment (and we call moral prioritization) for some behaviours.

However, Sen's suggestion has been resisted. Hausman (2012) defends preference orderings as a universal modelling device, although concedes that there are instances where most economists uses moral prioritization in their models. He gives the example of maximizing utility with a fixed amount of money, noting that the budget constraint cannot be expanded by stealing irrespective of how much it might satisfy preferences.

Consider figure 2A, which is a standard undergraduate diagram of utility maximization. A consumer has a budget set determined by their income, and they maximize utility subject to this set. As shown by the arrow, the consumer has income taken away from her, where the budget sets for income are solid, and must suffer a decline of utility as a result. Hausman points out that what is missing in a standard undergraduate diagram like 2A, but is included in 2B, is a parameterization of theft. The optimal amount of theft could be calculated and it provides a cushion for utility reduction. Clearly, theft would increase as utility falls.

Figure 2A: Undergraduate Diagram

(reduced utility from lower income)

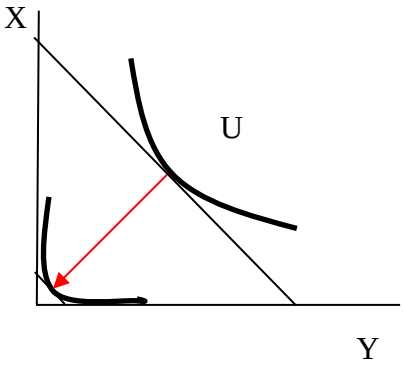
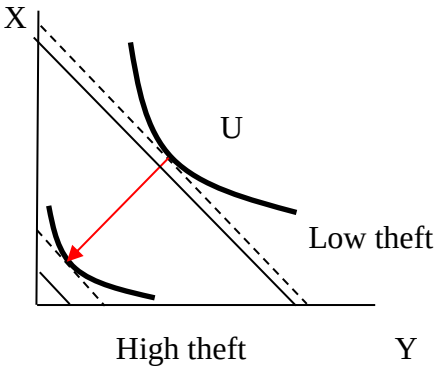


Figure 2B: Optimal theft

(dashed budget set includes receipts from theft)



The widespread use of 2A in textbooks (and journal articles) rather than 2B is a formal recognition of what has been noted at least since Adam Smith (1759), that certain ethical norms are necessary for a well-functioning market system. We might optimize over many choice variables in a problem but the bedrock norms – such as no stealing in figure 2A – are not subjected to the principle of optimality.

Anderson (2001) joins in the debate between Sen and Hausman and makes the methodological point that whilst it may be *possible* to describe ethics in terms of preference satisfaction, it may not be the *best* procedure. In her view, the generation of other competing analytic tools is a way of gauging the adequacy of preference satisfaction. In view of the unresolved debate on

this matter, we follow her approach by accepting the validity of *both* moral optimization and moral prioritization as potential tools in order to model ethical choices.

Thus, we are not ruling out moral optimization (and hence using cost benefit analysis for ethical issues). Mainstream economic modelling has found moral optimization to be a very useful tool. Becker (1981) showed that moral optimization is easy to append to standard economic models, and this goes some way to explaining its popularity. Typically, the welfare of others is added into preferences (or a 'utility function') as just another 'good' and standard analysis can then proceed.

This approach seems especially reasonable when the nature of the ethical choice involves two things that are intrinsically good. In Becker (1981), an agent is deciding how much of a resource pool to access for themselves and how much to give to family members. In many plausible situations both actions could be described as ethical, so the idea of balancing considerations at the margin – i.e. compromising – seems both true to how people actually behave, and to how they should behave. In finance, by extension, one might imagine a professional deciding how much to charge a client. The difference between the maximum the client will pay and a scheduled fee defines a surplus. If the professional has the power to unilaterally decide what they charge the client they may well charge less than the maximum, depending on their regard for the client's feelings and financial circumstances. All this is amenable to preference-satisfaction modelling – what we have called moral optimization.

However, not all ethical considerations fall neatly into this framework. There are certain bedrock values for which 'optimal' violations seem ethically nonsensical, just like optimal stealing seemed out of place in figure 2B. We might note that, at least in OECD countries, the notion of an 'optimal amount of workplace violence' is not something that is easy to say, let alone implement. Indeed, even if a manager decided that the optimal amount of workplace violence were zero, the very act of optimizing this choice variable would be regarded as a morally reprehensible act.

> '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).

William's comment is also helpful for understanding how moral prioritization differs from a more sophisticated moral optimization that properly incorporates externalities. Even if the fully optimal amount (i.e. incorporating externalities) of, say, workplace violence is zero, Williams

would still claim that the calculation itself was unethical, regardless of the outcome. In other words, 'means' matter as much as 'ends'. Naturally, Hausman, who believes everything can be described in preferences, would take the opposite side in this ongoing debate.[8]

What might finance look like if the optimal amount of deceit aroused the same moral indignation as the optimal amount of theft, or the optimal amount of workplace violence? Erhard and Jensen (1998) claim that deceit among corporate management is sometimes hidden through redescription. They have their own terminology for trustworthiness, and in the following quote having 'Word-4 integrity' is being someone for whom 'what you say is so'.

> 'In everyday language violating Word-4 is "lying". When we use terms other than lying to describe a violation of Word-4 we inadvertently encourage the sacrifice of integrity. We have observed perfectly honest upstanding people in their roles as board members condone manipulation of financial reports because it does not occur to them as lying—it occurs to them as just part of what it means to ... "manage earnings".'
>
> (Erhard and Jensen 1998, page 17)

As in Jensen (2014), the authors are claiming that truth-telling should be a bedrock value in finance, not subject to optimality. In our terminology, truth-telling in finance should be an instance of moral prioritization rather than moral optimization.

3 Moral Prioritization and the Principal Agent Model

In this section we show the transformational significance of Jensen's (2014) quote for his own Principal-Agent framework (Jensen and Meckling 1976). Put simply, trustworthy communication implies that hidden action becomes discoverable. That is, the principal simply asks the agent, who then tells the truth. If the agent is trustworthy, the first-best solution becomes attainable, and the meaning of Jensen and Meckling (1976) is changed. It becomes a means of measuring the cost of adopting a second-best incentive contract when the first best is feasible, rather than a prescription to always use incentive contracts when there is hidden action. That is, if more truth-telling is really attainable through better training, as Jensen

[8] 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 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). 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. 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.

implies, then Jensen (2014) should override Jensen and Meckling (1976) in finance, on efficiency grounds alone.

We now illustrate this using a standard textbook exposition of the Principal Agent model (Milgrom and Roberts 1992, appendix to chapter 6 *Moral Hazard and Performance Incentives*). We make some cosmetic changes to their notation, but the setup is identical.

In Table 1 the principal hires an agent and relies on the effort *e* of the agent to generate revenues *R* for the principal. Effort is either high (*e=2*) or low (*e=1*) and high effort makes high revenues more likely as follows (Milgrom and Roberts, 1992, Table 6.5, pg. 201):

Table 1: Probability of Revenues for Different Levels of Effort		
	Revenue	
Effort level	Low R=10	High R=30
e=2 (high)	Prob=1/3	Prob=2/3
e=1 (low)	Prob=2/3	Prob=1/3

We assume the principal pays a lower case wage, *w*, to the agent when *R=10* and an upper case wage, *W*, when *R=30*. If *w=W* the agent receives a constant wage un-incentivized by outcomes. Milgrom and Roberts assume agent satisfaction is a concave transformation of the wage, net of effort cost, $\sqrt{wage}-(e-1)$. The concavity is a device to ensure the agent dislikes risk.

The agent can earn a wage for another company with satisfaction equal to unity (for simplicity). In order to guarantee the agent's participation the principal must offer satisfaction at least equal to this outside option. Principal satisfaction is expected profits, defined as expected revenues minus the expected wage: $E(\pi)=E(R\text{-}wage)$.

There are two optimal contracts. Contract O applies when effort is Observable. The principal offers the agent a steady wage of *W=w=4* if a high effort is forthcoming, and nothing otherwise. The agent sets *e=2* and $E(\pi)=70/3-12/3=58/3$. Agent satisfaction is $\sqrt{wage}-(e-1)=\sqrt{4}-(2-1)=1$, as required.

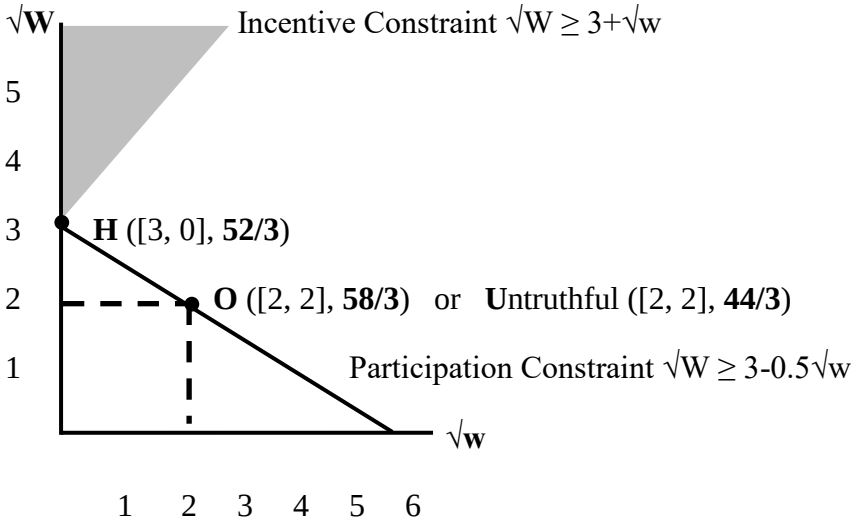
Contract H applies when effort is Hidden (and, implicitly, when agents cannot be relied upon to disclose it when asked). The solution involves meeting two constraints at minimum cost to the principal. The Participation constraint says that agent must be paid enough such that the expected satisfaction from participating exceeds what is available elsewhere, namely $1/3.\left(\sqrt{w}-(2-1)\right)+2/3.\left(\sqrt{W}-(2-1)\right)\geq 1$. The Incentive Compatibility constraint says that a high level of effort is preferred to a low one. That is, the expected satisfaction from e=2 must not

only exceed unity, as in the Participation constraint, but also that it must exceed what is available when e=1 namely $2/3.\left(\sqrt{w}-(1-1)\right)+1/3.\left(\sqrt{W}-(1-1)\right)$.

The participation and incentive compatibility inequalities simplify to $\sqrt{W} \geq 3-0.5\sqrt{w}$ and $\sqrt{W} \geq 3+\sqrt{w}$ respectively in Figure 3. The vertical axis shows the square root of the wage when *R* is high, namely $\sqrt{W}$, and the horizontal axis shows the square root of the wage when *R* is low, namely $\sqrt{w}$. Using square roots means that the two constraints can be drawn as straight lines in figure 3. The Incentive Compatibility constraint is shown by the shaded area, and the participation constraint holds anywhere above the bold diagonal. Both constraints overlap and are therefore satisfied in the shaded area, but the point of minimum cost to the principal is *H* - the bottom tip of this area where $W=3^2=9$ and $w=0$.

The solutions O and H are shown by the triple *([√w, √W],* ***E(π))*** where the first two terms locate the point on Figure 2, and the (bold) expected profits determines the social desirability of the contract.

Figure 3: Wage Contract and Profits
for Hidden Action and Observable Action
(Figure 6.1, Milgrom and Roberts, 1992; Payoffs ([√W, √w], **E(π)**)



Milgrom and Roberts show the adjusted wage is equally satisfying in both H and O, and we refer the reader there for details. Since the agent is equally well off in both scenarios, the difference in expected profits, namely *58/3 – 52/3,* is the difference in social welfare.

We now show an additional point U, which is not in Milgrom and Roberts (1992). U stands for 'Untruthful' and it represents the outcome of a principal offering an agent a fixed wage contract (which is why it is the same point as O on the diagram), who accepts the contract, announces

e=2 but does e=1. The agent can lie because the revenue outcome in table 1 is stochastic and so they can blame low revenue on bad luck. We assume the principal has to pay the agent in this case so that a lying strategy is clearly optimal for a maximizing agent in a one-shot game if the agent is offered contract O. Choosing effort level *e=1* gives assured payoff $\sqrt{wage} - (e-1)$ *=2* since the agent is offered a wage of 4. The outcome for the principal is compromised by the agent's lie, giving *E(π)=E(R-wage)=44/3*. If the principal knows the agent type, it is clearly suboptimal to offer such a contract to an untruthful maximizing agent. It is socially sub-optimal too: the difference between first best profits versus profits at U are 58/3-44/3=14/3 and the extra gain for the agent (1 extra wage unit) leaves society with a welfare loss: 1-14/3 = -11/3

However, if truth-telling is feasible because the agent follows the moral sentiment of the Jensen (2014) rather than the Jensen and Meckling (1976), the economic analysis of the Principal-Agent problem serves a different purpose. Instead of advocating incentive contracts, as would be the case if unobservable action implied undiscoverable action, we now interpret the difference in social welfare *58/3–52/3* as the loss arising from failing to use feasible truth-telling in this economic environment. If truth-telling is not feasible *58/3–52/3* becomes the cost of untruthfulness.

Agent truthfulness in this framework is a violation of individual rationality, because any agent styled like Economic Man will accept the equal wage contract, *W=w=4* in exchange for high effort, put in low effort, but then lie by claiming the effort level was high. However, it is the nature of moral prioritization to over-ride maximizing behaviour, and we here remind the reader of the evidence of Erat and Gneezy's (2010) which is relevant for a trustworthy agent.

The formal model is helpful to see precisely how the contracts are transformed by trustworthy truth-telling, but we do not want the maths to obscure the point, which can be made very simply: We have shown in this section that a trustworthy agent subverts the presumption that what is unobservable is undiscoverable, allowing the first best to be attained.

Overriding individual rationality is an uncommon modelling choice within economics, though, as we noted above, optimal stealing is generally ruled out of consumer theory. However, in other professions like medicine, this overriding is assumed to be possible, so that we rely on truthfulness when we ask our doctor a question (Downie 1990). The relevant norm of truthfulness here is of moral prioritization, which over-rides individual rationality, rather than moral optimization. We would be unwise to consult a doctor who practiced optimal deceit, however other-regarding she might be.

In the same way, the principal in the Principal-Agent problem cannot rely on an agent telling them about hidden action in a workplace if the agent practices optimal deceit. However, if the principal knows the agent practices moral prioritization with respect to truth telling, there is an escape from Jensen and Meckling's (1976) inefficient incentive contracts.

4 Motivation Crowding Out in Finance

Frey (1997) defines motivation crowding out as the process whereby external (extrinsic) stimuli remove good internal (intrinsic) motivations. It is common to use motivation crowding out to explain how financial incentives affect work effort. Here, however, the focus is on how the external stimuli of financial incentives and economics training effect the intrinsic desire to tell the truth.

Motivation crowding out arising from financial incentives provides an explanation of why bankers at work show a permissive attitude to moral requirements like telling the truth (Cohn et al. 2014). Bonus-based pay culture could undermine bankers' moral motivations, leading them to act on the basis of material self-interest, constrained at best by the letter of the law. Bonuses could potentially frame banking as conducted only for money, by signaling that their job is only to maximize profit. Moral considerations on how to act, including around obligations to tell the truth to clients and shareholders, are implicitly downgraded.

Another possible mechanism corrosive to trust is that bonuses communicate that 'we do not trust managers'. The erosion of social preferences here would take the form of the experimental phenomenon of 'trust responsiveness' (Bacharach et al. 2007) whereby the communication of a lack of trust from party A to party B leads to party B actually becoming less trustworthy – a proverbial 'self-fulfilling prophecy'.

A third concern about bonuses comes from a former Chief Economist of one of Australia's 'Big Four' banks. His claim, which is new to the literature as far as we are aware, is that subjectively apportioned bonuses create an unhealthy dependency between the manager and the recipient, where the recipient is rewarded for mimicking the values of the manager, rapidly disseminating a new culture.

> '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. I believe this was a factor in changing the culture of bankers from 1995 to 2010, reinforcing the influence of global banking culture.' (Hogan 2018, slide 18)

Whatever the particular source, Hogan is critical about the impact of motivation crowding out in Australian finance, as it increasingly mimicked global banking culture over 1995 to 2010.

> '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. Based on my observation of the market, and dealing with counterparties, this lead to a cultural change in banking, where it [became] all about the money with a focus on short term profitability.' (Hogan, op. cit.)

There is now a significant body of international evidence to warrant concern about motivation crowding out. An early contribution was due to Titmuss (1970), who eloquently argued that financial incentives can evacuate intrinsic motivation, much like the aforementioned day-care example where a fine increased late coming (Gneezy and Rustichini 2000). 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.

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 even crowd in, good motivations (Frey 1997). On balance, however, crowding out is more often observed in experiments (see Bowles and Polania-Reyes 2012 for an extensive review and Bowles 2016 for a popular discussion).[9]

Thus, less reliance on incentive contracts could yield a double dividend. Such a strategy would not only promote efficiency if agents were trustworthy (as outlined in section 3 above), but also restrict the undesirable effects of motivation crowding out.

While bonuses are one extrinsic stimulus that can cause motivation crowding out, another is the kind of economics training received. We have already flagged the misuse of cost-benefit analysis is section 2 as a potential problem, but the impact of the economics training is wider than this.

The latter decades of the twentieth century where noteworthy for the pro-market cultural tides

[9] An example of crowding in is given by Bowles (2016). In variants of the public goods game with punishments, all agents have to reveal their contributions. In some cultures, low contributions invoke punishment by fellow players, and this acts to hold up the contributions of everyone (in other cultures, however, punishments meet with retaliation and contributions fall across subjects).

that reached their high-water mark during the Reagan/Thatcher era.[10] Their conception of economics placed a good deal of reliance on Adam Smith's 'invisible hand' metaphor, reiterated and developed by Arrow and Debreu (1954) and Hayek (1945) and popularized by Milton Friedman. Reliance on the metaphor, mediated through an increasingly insular economics training,[11] provided some with a rationalization – a strategy for reducing cognitive dissonance by adapting belief to desire (Elster (1983: 123, 156)) – for a particularly narrow vision for the economy. Important qualifications to the invisible hand, such as problems associated with public goods and externalities, where not highlighted.

The assumptions of Economic Man arguably developed into a moral norm with a weak sense of social responsibility, sometimes even lacking a conception of society itself.[12] It involved agents maximizing their financial wealth (Mill 1974/1843) or happiness (Bentham 1948/1789) and the paramount importance of 'preference satisfaction'. The conflation of orderings, self-interest and welfare (Sen 1977) creating serious terminological confusion, hampering the discussion of ethics within economics.[13] On a practical level, the upshot of this kind of economics training favored a pragmatic approach to ethical challenges, seeing them as problems to be solved with cost-benefit analysis, implicitly leading to an 'optimal amount' of wrong-doing.

The impact of this on business schools, management and global finance culture can be seen by considering the cultural transformation of the leading US business schools in the closing decades of the twentieth century.[14] Partly under the influence of Professor Michael Jensen, whom we have come across a number of times in this paper, a generation of students were

[10] The most common term for this is neo-liberalism, but this is sometimes used pejoratively and we want merely to describe the orientation to markets, rather than judge them.

[11] There is some evidence that economists are less likely to cite outside their field compared with other scholars. Fully 81% of economists' citations are drawn from within their discipline, as against 52% for sociology, 53% for anthropology, and 59% for political science (Fourcade et al. 2015).

[12] Lydenberg (2014) contrasts the 'rational person' of economics (economic man), whom he says pursues his own personal ends, with the 'reasonable person' of tort law who is defined 'in terms of the interests of oneself in relationship to society's interests and the interests of others in that society' (op cit., pg. 288). Mrs Thatcher is famously quoted as saying '… who is society? There is no such thing! There are individual men and women…' (Woman's Own 1987), although in fairness to her she claims that life is 'reciprocal' soon after in the same interview.

[13]The same lack of clarity affects the term 'utilitarianism' which was the historical precursor to preference satisfaction. As discussed in Collard (1978) Mill – the creator of economic man –recognized that utilitarianism was both used as an explanation for the behaviour of essentially selfish individuals, or as a moral vision which enjoins impartiality. Collard suggests that Mill believed, in a vague way, that education and social progress would close the gap between the two usages (Collard 1978, pg. 58). One hundred and fifty years hence, the gap remains, with the same word 'utilitarianism' uncomfortably stretched across it.

[14] The insights about US business schools are from Professor K. Ramanna, Blavatnik School of Government, University of Oxford, formerly of Harvard Business School, and I am grateful for him sharing a draft chapter contribution of his in a forthcoming book.

steered away from a stakeholder view of firm management towards one based on single-minded profit maximization assisted with Jensen's Principal-Agent contracts. (Khurana 2007).

Although Jensen asserted his models of firms were descriptive, his more popular writings (such as Jensen and Murphy 1990) promoted his contracts as good managerial practice in a normative sense, giving imprimatur to both the relentless profit-maximizing self-interest of the principal, and the agents who are unable to truthfully communicate.

Thus Adam's Smith's invisible hand ceased being a statement about how markets can work with a range of motivation exogenously given, and became instead a prescription that firms and managers should abandon a stakeholder view so as to align themselves to a self-interested competitive benchmark.

It is hardly surprising that economics training would be affected by these developments. On the 'supply side' the models produced in journals had been transformed by Jensen and Meckling (1976), and on the 'demand side' student were being sent out into a world of Jensen's imagining. Naturally, curricula had to adjust.

Evidence for motivation crowding out from economics training is found Frank et al. (1993). They survey a series of experiments with economics and non-economics 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 and Schulze (2000) and Frey and Meier (2003). 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, Bauman and Rose 2011).

It would be interesting to know how Jensen might have incorporated his post-1990s views into the 1970s Principal-agent model. The culture of professionalism (Downie 1990), and now Jensen himself, must believe that truth-telling is possible, for otherwise there would be no point in enjoining professional people to display personal integrity. Yet in the economics curricula of the late twentieth century truth-telling as a matter of principle had long been assumed out of existence.

5 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 sounding utopian. There may be

ways to ‘transform’ finance so that it is more ethical that other professions, or more ethical than it ever was, but we are setting ourselves a more modest goal.

With regards to our specific contribution, if our motivation-crowding-out critique of financial incentives and economics training is correct, then it follows that the restoration of trust would be aided by less reliance on incentive contracts in the workplace, and, by discouraging the use of cost-benefit analyses for matters of integrity, as Jensen (2014) implies. This implies training participants in the finance industry to think in terms of moral prioritization rather than moral optimization.

The change could be assisted by altering some professional norms in the workplace. It could also be aided by altering economics training in such a way that students are exposed to representative agents that are different to Economic Man, who optimizes over everything. A good start would be some exposure to the Reasonable Person of tort law, who is defined ‘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). Depending on how pluralistic a classroom is, one could also consider feminist or religious perspectives (Tronto 2013, or Menzies and Hay 2012). Another step would be to teach agency theory differently, bringing Jensen (2014) into conversation with Jensen and Meckling (1976), much along the lines of section 3 above.

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 firms that behave unethically will be driven out of a competitive market, and this may be true for sellers trying to sell rotten apples in a fruit market. However, the proverbial rotten apples of finance were not discovered during deregulation episodes in many economies during the late twentieth century, which gives us pause. Appendix 2 discusses why deregulation is unlikely to be effective at punishing unethical behaviour in finance. With respect to lying, the informational asymmetries are at least one problem – financial products and disclosures (to customers and shareholders) are complex, making it is hard to detect deceit.

Nor are we proposing a general clamp down on finance, greatly reducing its size and profitability, by a general re-regulation. Since we have claimed financial incentives are overused and that economics training could be altered at the margin, we think a more judicious path is to attempt these reforms before embarking upon such a course of action. Of course, this does not take away the social choice (which exists in many industries) between having a dynamic and yet destructive environment versus a stable and staid one, but that is not the choice we are analysing in this paper.

We close with some general comments which, while not original, could be relevant for policymakers who wish to restore trust in finance.

We earlier noted the work on lying aversion by Erat and Gneezy (2010), which saw fully one third of subjects decline to tell a Pareto White Lie. Evidence like this might go some way to explaining why many profession associations take the risk of enjoining their members to tell the truth, rather than relying on Jensen and Meckling's contracts. It also lends plausibility to the hope that finance might be expected to rise to the standards of these other professions.

Economists who are otherwise uncomfortable with moral philosophy might like to draw on Adam Smith (1759) to justify moral prioritization. He believed that moral obligations arise out of a fellow feeling for the community in which one lives; his 'impartial spectator' was devised as an attempt to show how an individual comes to understand what these obligations might be, and might change his or her actions as a result (Wight 2015). Smith's interest in the possible conflict between moral obligations and economic motivation was typical of his time, as was his wide-ranging consideration of the relevant issues (Oslington 2012). As discussed by Collard (1978, pg. 51 ff.) foundational modern thinkers such as Butler, Hume, Mill and Edgeworth all recognized a tendency not to count others' welfare as much as one ought to for the flourishing of society, except in enlightened moments of 'conscience', 'calm judgement' or 'calm moments'.

Another source of moral obligations which might underpin moral prioritization is deontological ethical theory. Deontology seeks good rules of action, such as the Kantian Categorical Imperative to 'act only in accordance with that maxim through which you can at the same time will that it become a universal law', or its corollary that people should never be treated merely as a means to an end (Kant 1785). Bowie (2017) pursues the argument that a consistent and generalized application of Kantian principles in Business could constitute a form of trustworthiness.

Alongside Kant's search for universal principles, there is a need for more 'local' rules that do not have to meet Kant's requirement of being applicable everywhere. In a work context, such local rules often constitute professional codes of conduct, 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).

Professionalization has arisen in occupations 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, what is of critical practical value is one's freedom; at school, education; in medicine, health. As these examples show, there are numerous other workplaces that maintain standards of professionalism, and where practitioners are not expected to exploit informational or monopoly power at the cost of those whom they serve.

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. Professionalization in finance would involve an examination of pay structures, to return to the motivation crowding out point made earlier.

The foregoing suggestions are not radical, but neither are they straightforward to implement. Restoring trust and trustworthiness cannot be achieved by institutional reforms unless they are accompanied by a change in outlook. In particular, the keeping of rules – either general Kantian rules or professional codes of conduct – will safeguard the system only to the extent that bankers desire to be ethical.

There are at least two barriers that stand in the way of this change in outlook.

First, there is an apparent 'moral boundary' between work and home in Cohn et al. (2014). The bankers who lied did so when primed to think about work, but not about home. This discrepancy is a pivotal interest for the Ethics of Care research program (Tronto 2013), which wants to 'deconstruct' the moral boundary between home and work, allowing care (defined in some detail by Tronto (2013)) to cross over from the former to the latter. Walzer (1983) explores the meanings of care that should apply to home and to the workplace.

The second barrier standing in the way of a changed outlook is that the motivational power of money may not be fully understood. Money priming has suffered recently at the hands of the replication crisis, but we would caution against completely abandoning this kind of research. Indeed, the 'sacred' meaning attached to money in sociology echoes a pre-modern tradition which cautions against the motivational dangers of money (Heilbroner 2000). The influential example of Augustine (426) is analysed by Cameron (2011). Augustine's model of ethics asserts the interplay of emotions and intellect. He conjectured that our understanding of the world is determined by our 'loves', which can include things as well as people. In the absence of a correct 'ordering of loves' a person can be obsessed by something so that 'it fills the horizon and the desire for it displaces other desires worth having. Instead of abundance, all we

see is scarcity' (Cameron 2011, pg. 53). Augustine sees money as having a personality, which can be served and loved. [15] According to this view, restoring trust in finance is as much about 'ordering of loves' as it is about good rules. Writing in this tradition, Welby (2013) says rules in finance have limited usefulness if people do not desire the social goods that the rules are designed to foster.

To conclude, we have argued that finance may have over-used incentive contracts, and we agree with Jensen (2014) that many finance professionals have received a very particular ethical training that uses crude cost-benefit analysis for ethics. It is difficult to imagine restoring trust in bankers while ever their workplace culture applies cost benefit analysis to all moral decisions, for *moral optimization* will always prescribe an optimal amount of a 'bad' like deceit. We argued that sometimes it is better if a worthwhile principle overrides utility- or profit-maximization. We called this *moral prioritization* and discussed how to reinstate it in finance.

While it is easy to make claims with hindsight, economists (and society generally) may have been too optimistic about Adam Smith's invisible hand operating in the financial system, and too reluctant to ask financial market participants to tell the truth with the same frequency as other professions. Perfection is unattainable, but nevertheless society can and should expect more principled behaviour from more principled agents.

Appendix 1: Finance Was More Trustworthy in the Past

Cohn et al. (2014) does not explain the lack of truth-telling by bankers – it only establishes its existence and correlates it with their workplace. But the natural question which arises is whether this has always been the case and, if not, what might have caused the change. Our reading of the evidence is that the untrustworthiness observed in Cohn et al. (2014) is contingent and recent. In coming to this view, we rely on the well-documented and well-accepted narrative of the transformation of The City (of London) financial centre from a service-orientated profession to a workplace for egoistic profit maximizers.

We first tell the story historically, based on secondary sources,[16] and at the end of this appendix we tell the story analytically, using a modified version of Bowles's (2011) stylized

[15] Augustine draws on the New Testament, which is responsible for the famous saying 'For the love of money is a root of all kinds of evil' (Paul, in Timothy 6:10). This is often misquoted as the less defensible 'money is the root of all evil' most famously by Ayn Rand in her (1957) defence of markets in the novel *Atlas Shrugged*.
[16] Our historical analysis draws extensively from Jaffer et al. (2014). See also Martin (2016) and Offer (2014).

evolutionary-game theory diagram of institutional and cultural change. We describe a 'non-virtuous circle' whereby increased offering of incentive contracts erodes virtue, requiring that more incentive contracts be offered, and so on.

For most of the 20th Century British banking 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. This is the origin of the term 'gentlemen bankers', collectively referred to as the 'Club'.

The banking community at the time operated largely by self-regulatory agreement, 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, given the partnership arrangements, they were taking risks mostly with their own funds. Investment bankers depended very much on their reputation, which had developed through long-term relationships with clients and other counterparties within The City. Most banks had centralized, and demanding, inspection regimes which ensured that rules and procedures were strictly followed and clients were served well.

Growth and consolidation in British banking occurred over the latter half of the twentieth century spurred by general financial deregulation. At the so-called 'Big Bang' in 1986, fixed commission charges were abolished and the Stock Exchange changed from open outcry to electronic trading. Previously separate financial organisations began to merge, and capital markets became dominated by global investment banks with large capital bases. Bankers struck profit-sharing bargains with their new shareholders, and a bonus-pay culture took hold.

The arrival of overseas banks transformed and globalized the culture of The City. At least by 2008, this global culture was not known for its truthfulness. US issuers and underwriters of mortgage-backed securities (MBSs), in a violation of fiduciary duties to both shareholders and customers, misled shareholders about their own MBS holdings and bet against these assets even as they sold them to trusted clients. Most of the largest mortgage originators and mortgage-backed securities issuers and underwriters have been involved in regulatory settlements, and have paid multibillion-dollar penalties (Fligstein and Roehrkasse 2016). UK

court cases since 2008 are testimony to the global contagion that was underway. Barclays and four former executives have recently been charged with fraud in 2008 (The Economist 2017) and Payment Protection Insurance (PPI) mis-selling was a growing problem from the mid-1990s.[17] A very large number of those policies were sold to clients who did not ask for them, did not understand them, or did not know that they would be unable to claim against them. By the time the fraud was uncovered, and the High Court ruled against the practice in 2011, it had become 'systemic' in the financial system (HoLC 2013).

The arrival of global banking culture to Australia came a bit later, but according to Hogan (2018, quoted in body of the paper) the new culture became obsessed entirely with short term financial gain, and was partly responsible for the abuses uncovered by the 2017-2019 Australian Royal Commission into Misconduct in the Banking, Superannuation and Financial Services Industry.

Returning to the UK narrative, the globalization of banking culture in The City was often accompanied by the formation of a 'markets division', managed by people who began 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 corresponding surge in the pay of other bank board members, which could be justified only by raising shareholder expectations of returns. Higher returns were achieved by increasing the levels of leverage and risk (Jaffer et al. 2014). Even those who did not receive bonus-based pay packages began to inhabit a banking culture generated by those who did. The bonus culture saw an evacuation of ethical motivations, which in turn required that people be motivated increasingly by financial incentives. The possible ways in which bonuses accomplish this are discussed in the main text.

Difficulties in making competition work in finance (discussed in appendix 2) meant that any misbehaving bankers were not driven out of the market. Rather, they began to drive its values, treating customers and shareholders in ways that would have been unacceptable during the era of the Club.

Financial products were increasingly sold on a *caveat emptor* (let the buyer beware) basis, and bankers maximized rents arising from market power and informational asymmetries (Woolley 2010). The simplest way to exploit the latter is by not disclosing to clients low probability but

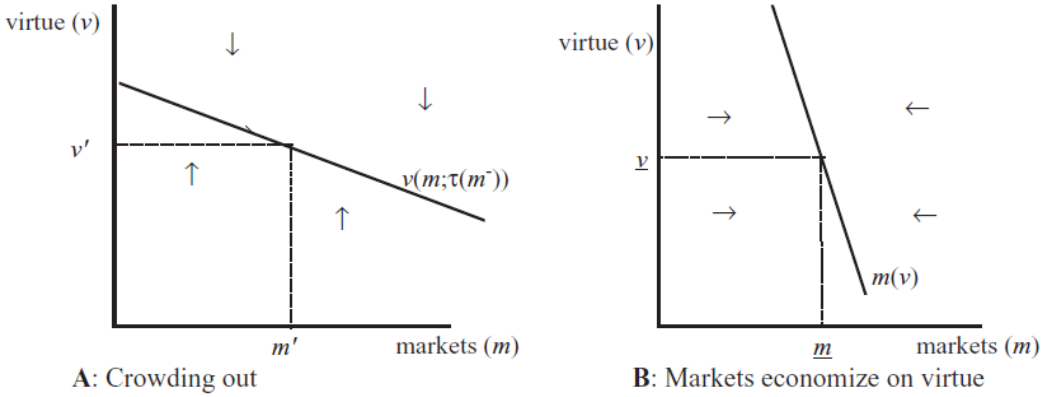
[17] PPI is an insurance facility that Banks and Building Societies sell to borrowers to protect them against the possibility that they might be unable to cover future repayments.

damaging ('tail') risks. With regards to shareholders, informational asymmetries allowed bankers to mislead shareholders (and clients) about the worth of their management services.

Over this period, the general view is that the conservative virtues of probity and truthfulness that had characterised banking culture were replaced by the pursuit of personal gain (Jaffer et al. 2014, Martin 2016 and Offer 2014). The Club had been based on delivering a service, but participants in the global banking culture that replaced the Club trained themselves to mislead customers and shareholders. Turner (2010) and Kolb (2010) catalogues the well-known unravelling of the system in 2008.

We are aware in the foregoing account there is a two-way causality between motivation crowding out and incentive contracts. On the one hand, banker bonuses crowded out good motives, but on the other hand if agents have poor motives it is natural to motivate them with high-powered incentive contracts (as this is likely to be the only motivation that will work).

An empirical modelling of this process is beyond the scope of this paper, and the prospects for a non-experimental investigation are probably poor, since attenuating virtue among agents is something they are likely to hide. (Cohn et al. 2014 succeed in uncovering it, but most experiments are a somewhat contrived environment). As an alternative, we describe this negative feedback loop using an adaptation of Bowles's (2011) stylized evolutionary-game theory diagram of institutional and cultural change. His original diagrams (labelled A to D) are reproduced below, with the addition of phase arrows.



A: Crowding out

B: Markets economize on virtue

Suppose a bank (principal) assigns a group of agents to be managers, whose type (trustworthy or untrustworthy) is known to the bank.

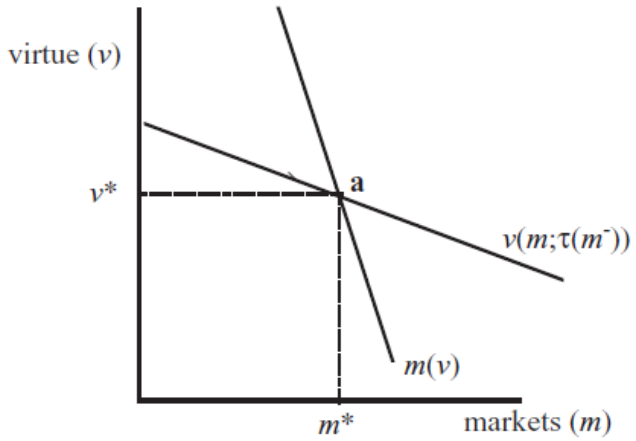
We let Bowles's virtue (v) be the number of truth-telling managers and we let the extent of markets (m) refer to the number of performance-based incentive contracts. Just as markets are

designed to function in an economy inhabited by Economic Man, Principal-Agent incentive contracts are designed to function in a firm inhabited by untrustworthy managers.

Panel A, the Crowding Out diagram, is downward sloping because offering incentive contracts communicates that the bank doesn't trust managers. The phase arrows (which are implied but not drawn in Bowles's original diagram) indicate how truth-telling by managers drifts to this 'message' of distrust if it is off the line. The second term of $v(m; \tau(m^-))$ shifts the whole curve down for a decline in Bowles's 'tradition' variable $\tau(m^-)$, which depends on all previous values of m. We interpret tradition as agents' use of moral prioritization rather than moral optimization. In Bowles's framework a decline occurs with an unspecified time lag, as the number of incentive contracts rises (though we make the effect instant when we use mathematics below). A maximizing untruthful agent faced with a fixed wage contract declares untruthfully that they have put in contracted effort and then blames any resultant low revenue on bad luck, making the offer of such a contract sub-optimal for the bank (who we assume still pays the agent). We might imagine cultural decline occurring when the number of incentive contracts based on Jensen and Meckling (1976) become so high that employees decide to read the original article, and then revert to their business school training Jensen (2014) that moral optimization is 'correct'.

Coming to Panel B, the Markets Economize on Virtue diagram, this is the number of incentive contracts a profit maximizing bank should offer given that it knows the type of the agent. When truth-telling is common only a few managers need incentive contracts but as more agents abandon truth-telling, more contracts have to be offered. The phase arrows indicate that if more than the minimum number are offered, the bank can be more efficient by replacing some contracts with fixed wage contracts, since they can rely on moral prioritization for the trustworthy agents, and that if too few are offered they have to increase the number since the fixed wage contract will be exploited by every untruthful agent.

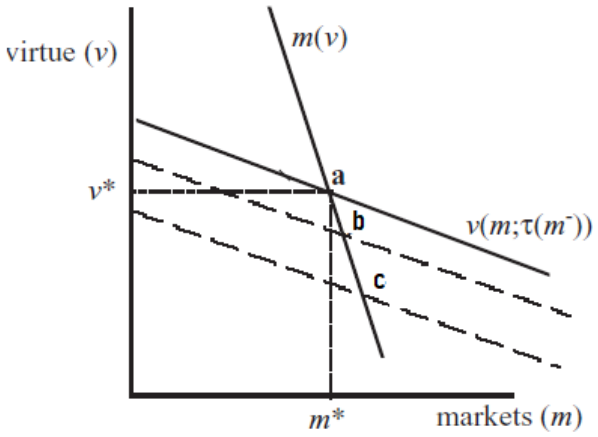
The intersection of the Crowding Out and Markets Economize on Virtue lines forms what Bowles (2011) calls a temporary cultural-institutional equilibrium.



C: A temporary cultural-institutional equilibrium

A negative feedback loop – what we call a non-virtuous circle – could be set off by any exogenous decline in the Crowding Out schedule, such as the influence of global banking culture described in Hogan (2018).

In terms of our stylized model, such a change in culture shifts down the Crowding Out schedule. That is, for a given number of incentive contracts fewer managers are truthful. On the path to point b in Panel D below, there is first of all a decline in the number of truthful managers. This requires the bank to offer more incentive contracts, leading to some more managers becoming untrustworthy via the 'banks don't trust managers' message, until point b becomes the new temporary cultural-institutional equilibrium with more incentive contracts and fewer truth-tellers than point a.



D: The long-run erosion of tradition

However, over time, the increased salience of incentive contracts, as they become more common, leads to an increased uptake of cost-benefit utilitarian ethics, i.e. moral optimization,

endogenously eroding whatever workplace culture of truth-telling remains. The Crowding Out curve drops to point c, and so on – creating the non-virtuous circle.

We can summarize Bowles's point by relabelling m(v) and v(m, $\tau(\bar{m})$) as MEV(m) and CO(m, $\{\tau(m)\text{-}e\}$) where MEV and CO stand for 'markets economize on virtue' and motivation 'crowding out', and we allow for the worst-case-scenario of an instant impact of m on tradition, as well as a shifter e which hurts tradition, such as the influence of global banking culture described in Hogan (2018). Taking the total derivative at the equilibrium MEV=CO:

$$dMEV = dCO$$
$$MEV_m dm = CO_m dm + CO_\tau \tau_m dm \quad - de$$
$$\frac{dm}{de} = \frac{1}{\{CO_m - MEV_m\} + CO_\tau \tau_m} > 0.$$

The denominator is positive for sufficiently large $CO_m - MEV_m$ which corresponds to a stability condition for the model. Thus an exogenous decline in Bowles's motivation crowding out function (via the shifter e above) will lead to a further uptake of incentive contracts, both because of necessity (moving down along the markets economize on virtue line m(v)) but also because the extra salience of contracts will over time encourage people to alter their moral frame (like the parents in Gneezy and Rustichini (2000)), leading to a shift down in the whole motivation crowding out schedule v(.) via an erosion of tradition (τ), as moral prioritization is abandoned in favour of moral optimization.

This is our so-called non-virtuous circle. A lack of trust becomes self-fulfilling as in the aforementioned phenomenon of 'trust responsiveness' (Bacharach et al. 2007) whereby the communication of a lack of trust from party A to party B leads to party B actually becoming less trustworthy. But in the non-virtuous circle, the fact that party B becomes less trustworthy additionally requires party A to motivate her with an incentive contract which, while it may function well as an incentive, introduces another round of negative (mis)trust responsiveness, and so on.

Appendix 2: Competition is Unlikely to Drive Out Bad Ethical Behaviour in Finance

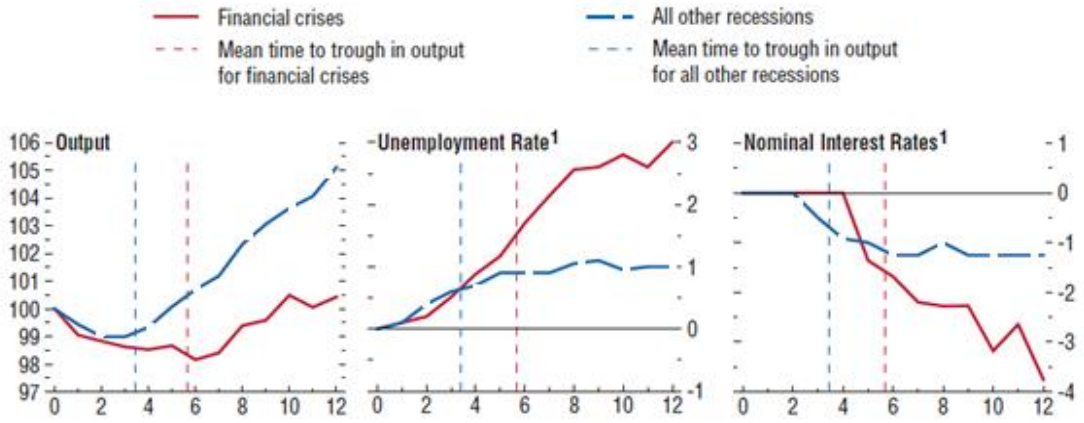
Our first concern about competition is that deregulation in finance will not necessarily destroy poorly performing firms. As was amply demonstrated in 2008, risks associated with bank failure are so large that the authorities cannot afford to let some poorly performing firms go bankrupt (Turner 2010). In the 2008 crisis not only did some bailed-out institutions access

public money, but some managers exited with substantial bonuses, leaving them free to mismanage again.

There has been an international regulatory response, not least in the form of Basel III. It has sought to strengthen incentives for good behaviour, protect depositors by increasing the quantity and quality of capital, enhance liquidity provision and introduce macroprudential policy. In the UK, 'ring-fencing' and 'bailing in' are discussed in the Vickers Report on the UK banking system (Edmonds 2013), and the intention of both is to hold high risk takers to account.

Nevertheless, recessions compounded by financial sector crises are deeper and longer than other recessions. Figure 4 (IMF 2009) shows these features averaged over worldwide financial sector recessions, and we note particularly that easier monetary policy is generally pursued. This means that even in scenarios where some errant banks fail without endangering the public purse, the likely easing of monetary policy during a financial crisis will protect some other errant banks. If errant firms exhibit unethical behaviour more than other firms, then competition policy is a poor tool for encouraging an ethical financial system.

Figure 4: Financial Sector Recessions are More Severe
(IMF calculations, quarters since peak in real output)



(IMF 2009, Figure 3.8, pg. 118)

Our second concern about competition is that it is very difficult for customers and shareholders to monitor banks, and accountability is a very important component of market discipline. The fundamental problem here is conceptual. Whatever accounting conventions are used, there is significant uncertainty about the measurement of risk, and this makes it difficult to adjust accounting profits for risk. Haldane et al. (2011) propose that this should be a priority in any reform of the accounting 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 in a position 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. The dangers are illustrated by Noe and Peyton Young (2014). They show how a manager can use derivatives to increase returns by generating extreme tail risk for the client. In normal times the client makes a good return and the manager gets a good bonus. However, during a tail event the investor loses everything, but the fund manager merely fails to get his or her bonus.

Any conflation of risk and return in measurement allow bankers to lie to ill-informed shareholders and customers about the risks of particular strategies. Furthermore, these measurement problems may paint a picture of an institution able to withstand the inevitable ‘bad draws’ of risk-taking, when in fact the institution is not sufficiently fortified.

The upshot of all these information problems is that there is a significant degree of difficult-to-quantify risk built into the financial system, which allows bank managers to lie to shareholders and customers should they wish to do so. It remains to be seen if advances in measurement, such as those sought by Haldane et al. (2011), will solve these problems, but such a development would be welcome.

Our third and final concern about competition policy is that attempts to reduce market power by encouraging new entrants may encourage risk taking and deception. Berger et al. (2009) suggest that new entrants are more likely to result in ‘competition fragility’ than to reap the benefits of the invisible hand. Competition fragility is the scenario where compressed bank margins tempt banks to pursue zero or negative expected excess return strategies, which none the less look very profitable in the short run. That is, strategies like Noe and Peyton Young (2014) become *even more likely* in competitive environments. As we just noted, these strategies are poorly understood, affording many opportunities for deception.

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.

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.

Cameron, A. (2011), *Joined-up life: A Christian account of how ethics works*, Wipf and Stock, Oregon.

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

Engelen, B. (2017), 'A New Definition of and Role For Preferences in Positive Economics', *Journal of Economic Methodology*, 24(3) 254-273.

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.

Heilbroner, R. (2000), *The Worldly Philosophers*, Penguin, London.

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.

Kiechel, W. (1987), "New Debate about Harvard Business School," *Fortune*, November 9, http://archive.fortune.com/magazines/fortune/fortune_archive/1987/11/09/69824/index.htm.

Klein, R. A., and 50 other authors at https://econtent.hogrefe.com/doi/abs/10.1027/1864-9335/a000178 (2014), 'Investigating variation in replicability: A "many labs" replication project', *Social psychology*, 45(3), 142.

Kolb, R. (2010), *The Financial Crisis of Our Times*, OUP.

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.

Menzies, G. and D. Hay (2012), 'Self and Neighbours', *The Economic Record*, 88, Special Issue, June, 137–148.

Milgrom, P. and J. Roberts, (1992), *Economics, Organization and Management*, Prentice Hall, New Jersey.

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.

Oslington, P. (2012), 'God and the Market: Adam Smith's Invisible Hand', *Journal of Business Ethics* 108, 429–438

Philippon, T. and A. Resheff, (2009), 'Wages and Human Capital in the US Financial Industry: 1909-2000, Quarterly Journal of Economics, 127(4), 1551-1609.

Rand, A. (1957,1992), *Atlas Shrugged*, Dutton, New York.

Rohrer, D., Pashler, H., and Harris, C. R. (2015), 'Do subtle reminders of money change people's political views?', *Journal of Experimental Psychology: General*, 144(4), e73.

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.

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

Tronto, J. (2013), *Caring Democracy: Markets, Equality, and Justice*, London and New York, New York University 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

Vadillo, M. A., Hardwicke, T. E., & Shanks, D. R. (2016), 'Selection bias, vote counting, and money-priming effects: A comment on Rohrer, Pashler, and Harris (2015) and Vohs (2015)', *Journal of Experimental Psychology*: General, 145 (5), 655-663.

Vohs, K. (2015), 'Money priming can change people's thoughts, feelings, motivations and behaviours: An update on 10 years of experiments', *Journal of Experimental Psychology*: General, 144 (4), e86-e93.

Welby, J. (2013), 'How do we fix this mess', Westminster, 22 April,

http://www.archbishopofcanterbury.org/articles.php/5050/how-do-we-fix-this-mess-archbishop-justin-on-restoring-trust-and-confidence-after-the-crash .

Walzer, M. (1983), *Spheres of justice: a defense of pluralism and justice*, New York: Basic.

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.

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.

Woman's Own, (1987), 'Margaret Thatcher interview', Journalist: Douglas Keay, 23 September, No. 10 Downing Street.

Woolley, P. (2010). 'Why are financial markets so inefficient and exploitative – and a suggested remedy', *The Future of Finance.* London: School of Economics and Political Science, 121–44.