Ethical standards and cultural assimilation in financial services*

Alan D. Morrison[†] and John Thanassoulis[‡]

April 2016

Abstract

We introduce ethical agents into an analysis of decision making in a profit-maximising firm. Agents can adopt a profitable new practice that may harm customers. Their decision reflects moral considerations, organisational culture, and compensation contracts. We analyse both utilitarian and deontological (duty-based) philosophical traditions. Cultural assimilation emerges as an equilibrium phenomenon. With sophisticated customers, the principal enables a culture that achieves the highest possible aggregate surplus and, under deontological ethics, both the principal and the customers would prefer to deal with less ethically committed agents. In contrast, the principal designs compensation to enable cultures that exploit naïve customers.

Keywords: Culture, ethics, behavioural norms, bonuses.

JEL Classification: D03, G02, G20.

1. Introduction

The financial crisis of 2008–09 cast a harsh light upon a number of practices in the financial services sector. For example, numerous commentators have pointed to questionable practices in the subprime mortgage lending market, where risky mortgage loans were sold to poor and financially unsophisticated people who did not understand the scale of the risks to which they were exposing themselves (see, e.g., Bar-Gill (2009) and Agarwal, Amromin, Ben-David, Chomsisengphete, and Evanoff (2014)). Financial institutions in the United Kingdom have been pilloried for using high-pressure tactics to sell Payment Protection Insurance (PPI) to

*We thank a seminar audience at Essex University Business School for helpful comments and discussions.
[†]Saïd Business School and Merton College, University of Oxford; CEPR; ECGI.
[‡]Warwick Business School, University of Warwick; Oxford-Man Institute, University of Oxford; Associate Member of Nuffield College, University of Oxford.

retail consumers who did not understand the product, and who often had no need for it.[1] In general, Célérie and Vallée (2013) find that structured products exhibit more financial complexity and have a higher "hidden markup" when distributed to less sophisticated investors. But malpractice is not restricted to the retail financial sector. Fixings of the standard money market LIBOR benchmark were systematically manipulated for years (Wheatley 2012), as were standard fixings in the foreign exchange markets.[2] At the time of writing, asset managers are accused in the EU of mis-selling by charging for active management when their funds simply track an index.[3]

In short, malpractice is endemic in the financial sector. Zingales (2015, p. 1348) states that "in the financial sector fraud has become a feature and not a bug." And, in an industry that relies upon trust, the systemic consequences of malpractice are potentially profound (for a discussion, see, e.g., Sapienza and Zingales 2012). Supervisors, rule-makers and commentators are therefore naturally concerned by the causes and implications of such practices. Several themes emerge from their public statements. First, supervisors increasingly appreciate that they cannot rely solely upon formal regulations to moderate market behaviour; it is frequently very hard for outsiders to assess actions, and, hence, internal controls and cultural standards are of critical importance. Those standards are created and transmitted by the senior members of the organisation. For example, Lord Adair Turner, chairman of the UK's financial regulator (the FSA) between 2008 and 2013, stated in a recent interview that

> "bank executives face the challenge of setting clearly from the top a culture which tells people that there are things they shouldn't do, even if they are legal, even if they are profitable and even if it is highly likely that the supervisor will never spot them."[4]

The importance of "tone from the top" to organisational culture is now widely recognised by supervisors (see also Adamson 2013), and many have concluded that top-down regulation is an important element of any response to behavioural problems in financial markets.

Second, notwithstanding our increasing appreciation of the social importance of culture in the financial sector, financial institutions appear frequently to foster dysfunctional cultures that fail to consider the potentially deleterious social consequences of practices that generate significant profits. For example, William C. Dudley (2014), President of the Federal Reserve

[1] See the "super complaint" brought to the UK Competition Commission by the UK's Citizen's Advice Bureau (Tutton and Hopwood Road 2005) and the Parliamentary Commission on Banking Standards's (2013a, p. 91) report.

[2] `http://www.nytimes.com/2015/05/21/business/dealbook/5-big-banks-to-pay-billions-and-plead-guilty-in-currency-and-interest-rate-cases.html`

[3] https://next.ft.com/content/d0c93bfa-c997-11e5-a8ef-ea66e967dd44

[4] http://www.bloomberg.com/news/articles/2012-07-24/turner-stakes-claim-for-boe-helm-as-bank-culture-reform-sought

Bank of New York, argues that regulators and bankers should address "cultural failure," and the G30 (2015) suggests that cultural problems were a contributing factor for the financial crisis; outside the immediate ambit of the regulators, Justin Welbey, the Archbishop of Canterbury, has spoken at length on this theme.[5]

Third, there is a growing consensus that executive pay has contributed to cultural problems in the financial sector. In the specific case of the UK's PPI mis-sellling scandal, Financial Services Authority's (2013, para. 22) written evidence to the Parliamentary Commission on Banking Standards identifies an excessive concern with targets and bonuses as a root cause of the mis-selling. Archbishop Welbey notes that individuals are generally more likely to violate norms of fairness when they are well-paid for doing so,[6] while President Dudley suggests that bonus deferral may be an effective way of changing cultural norms. In an attempt to redress the balance, regulators have imposed significant penalties upon individuals and firms that can be proved to have violated the law. At the time of writing, the SEC has launched enforcement actions against 198 entities and individuals, and levied penalties amounting to almost $2bn, for misconduct in securities markets that led to, or arose from, the financial crisis: most concerned misrepresentation of pertinent facts about investments.[7] Foreign exchange manipulation resulted in fines totalling $10 bn[8], and, as at May 2015, fines resulting from the LIBOR scandal topped $9 bn.[9]

Fourth, there is widespread agreement that behavioural problems in financial markets are associated at least in part with moral failings (see, e.g., Graafland and van de Ven (2011), Luyendijk (2016), Tett (2009), Wilson (2012)). Indeed, recent experimental work by Cohn, Fehr, and Maréchal (2014) strongly suggests that bankers are more likely to behave dishonestly in a professional than a personal context. It seems clear that we cannot consider bank regulations independently of moral standards.

In this paper, we present a simple economic model within which we can consider the relationship between ethics, culture and compensation policies. In particular, we exhibit situations in which a profit-maximising principal chooses to use compensation policy deliberately to subvert ethical decision making and to disrupt the transmission of valuable cultural norms. Our analysis thus requires us to model ethical decision making, to explain the formation and transmission of cultural values, and to relate both to compensation policies.

Notwithstanding the apparent ease with which some commentators identify immoral be-

[5]See, for example, the transcript of his 3rd October 2013 House of Lords speech at `http://www.archbishopofcanterbury.org/articles.php/5150/archbishop-calls-for-culture-change-at-financial-institutions`.

[6]See the speech cited in note 5, in which Archbishop Welbey draws an explicit connection between mis-selling and performance bonuses.

[7]`http://www.sec.gov/spotlight/enf-actions-fc.shtml`

[8]`https://next.ft.com/content/23fa681c-fe73-11e4-be9f-00144feabdc0`

[9]`http://www.cfr.org/united-kingdom/understanding-libor-scandal/p28729`

haviour, ethics are essentially contested. Philosophers have been arguing about what is right for more than two millennia and a variety of rich philosophical analyses promulgate inconsistent notions of ethical behaviour. Some philosophers attempt to define ethical behaviour with reference to what makes a virtuous person (e.g., Aristotle (2009), MacIntyre (2007)); for Aristotle, virtue is inconsistent with many commercial activities. Another perspective, expressed most famously by Kant (2012 [1785]), identifies right actions by considering the rights that people have by virtue of their humanity. In contrast to Kant, a utilitarian tradition due initially to Jeremy Bentham (2007 [1789]) identifies right actions entirely with reference to their aggregate consequences for every affected person.

Economic analysis can shed some light upon the effect that different ethical standards have upon resource allocation; it cannot adjudicate between different visions of what is good or right. Our analysis examines the economic consequences of two types of moral standard. Agents operating under the first subscribe to a version of Bentham's utilitarianism, under which an action is right if it maximises total well-being. Utilitarians since Bentham have considered questions about the distinction between the rightness of an *act* and of a *rule* (see, e.g., Rawls 1955), but our agents define the morality of an action with reference to its immediate consequences. An agent in our model has higher utilitarian ethical standards to the extent that he places a higher weight upon the Benthamite utilitarian welfare measure, and a lower weight upon his own concerns.

This approach lends itself well to an economic model in which actors maximise something. Indeed, many modern economists tacitly accept Bentham's ethical stance when they identify welfare with aggregate surplus. Furthermore, several studies indicate that real-world managers actually use utilitarian moral standards to resolve ethical dilemmas (see Fritzsche and Becker (1984), Premeaux and Mondy (1993), Premeaux (2004)[10]).

We also consider a crude version of Kant's (2012 [1785]) *deontological*, or duty, ethics. Kant argues that actions are properly constrained by duties that exist a priori: that is, Kantian duties, such as the obligation to tell the truth, reflect duties that exist before we understand their context, and that do not depend upon their consequences. More generally, at least some of the time, duty-based ethics forbid us to trade one agent's well-being against another's: for example, Nozick (1974, p. 33) argues that viewing one agent's suffering as sufficiently compensated by gain elsewhere "does not sufficiently respect and take account of the fact that he is a separate person, that his is the only life he has." We cannot capture all of the nuances of this reasoning in a simple economic model: we consider instead a simple approach under which an ethical agent refuses to impose any sort of harm upon another.

[10]All three papers attempt to infer the philosophical basis upon which managers resolve ethical dilemmas by examining their written responses to short business vignettes. For example, two of the vignettes asked the subjects whether they would pay a bribe to enter a new market, and whether they would suppress evidence that a profitable product was potentially unsafe. Practitioners relied almost entirely upon utilitarian standards to resolve these questions.

The moral agents in our model must decide whether or not to adopt a practice that may harm customers. Some, *senior*, agents have access to a signal of the harm caused by the practice. Other, *junior*, agents have no such signal, but are able to observe the adoption decision of the senior agent. When they are paid appropriately, we demonstrate that ethical junior agents copy the adoption decision of their better-informed seniors. They do this because they believe that their seniors are morally responsible, and because they know that their seniors are better-informed than they are.

We interpret this type of imitation as *cultural assimilation*. Culture is a complex and hard-to-define term. But, in an organisational context, it is widely understood to incorporate a set of "taken-for-granted-assumptions" (Giorgi, Lockwood, and Glynn 2015) and tacit norms that are absorbed, possibly unconsciously, through participation in the life of the organisation (for widely cited discussions of the transmission of cultural norms, see Schein (2004), Swidler (1986), and Patterson (2014)). The cultural assimilation in our model transmits this type of cultural knowledge.[11] The cultural failings to which we refer above arise when harmful practices are adopted in this way by morally responsible agents.

Whether or not cultural assimilation occurs depends upon the compensation policy of the principal, which, in turn, depends upon the sophistication of its customers and upon the ethical standard to which agents hold themselves. We start by analysing equilibria with sophisticated customers, who anticipate and demand compensation for equilibrium strategies that harm them. In general, agents must be compensated for violating their ethical standards. So, when they are utilitarian, they must be paid to induce actions that reduce aggregate output. Such actions also reduce a sophisticated customer's willingness to pay. The consequence in our model is that it is never worth inducing a utilitarian junior agent to deviate from the senior agent's action choice. That is, the only equilibrium that our model admits for utilitarian agents and sophisticated customers features cultural assimilation.

In contrast, there is a trade-off between the needs of sophisticated customers, who can reflect the costs and benefits of adoption in their willingness to pay, and a duty-based ethical standard that prohibits the imposition of any ex post harm. In this case, the profit earned by adopting a surplus-enhancing practice may outweigh the costs of inducing agents to invoke it. We therefore find that, even with sophisticated agents, the junior agent may not adopt a practice that the senior agent believes to be acceptable, so that cultural assimilation may fail.

We consider the contrasting case where customers are naïve: that is, when they fail to anticipate the potential harm imposed upon them by the agents with whom they deal. This is a reasonable approximation to the situation in many retail financial sectors and, in particular, to situations where financial products are mis-sold. In this case, because customers do not

[11]For an economic model that identifies cultural practice with shared understandings of the rules for selection amongst multiple equilibria, see Kreps (1990).

demand compensation for harms imposed upon them, the only brake upon harmful activity is provided by the agents' moral scruples. Once again, the precise characteristics of the equilibrium depend upon the agents' moral standard. But, precisely because customers do not charge for surplus-reducing actions, agents invoke harmful practices more often when their customers are naïve. Moreover, because invocation is profitable for principals, they seek the cheapest way to encourage it. Hence, they enable cultures that exploit particular agents' moral weaknesses. As a result, cultural assimilation need not obtain in equilibrium: junior agents may prefer always to adopt, or never to do so.

Our analysis is consistent with recent criticisms of compensation contracts in retail financial services. In our model, success bonuses in those industries are designed to encourage agents to set aside their moral concerns, and, at least when agents have utilitarian moral standards, economic surplus would be increased if they were disallowed. Moreover, because businesses with naïve consumers make more profits when their employees are morally lax, our analysis suggests that retail financial businesses have a strong incentive to seek such employees and to pay them what it takes to ignore their moral scruples. A cap on bonus payments in those industries would weaken that incentive and, hence, might result in higher average moral standards amongst retail financial advisors.

Our analysis complements work on compensation and culture in the financial sector. Several authors suggested in the wake of the financial crisis that inappropriate compensation contracts can induce short-termism and so undermine financial stability (e.g., Thanassoulis (2012, 2013), Fahlenbrach and Stulz (2011), and Efing, Hau, Kampfkotter, and Steinbrecher (2015)). Inderst and Ottaviani (2009) analyse the relationship between organisational structure, compensation, and misselling. But no prior work considers either the relationship between compensation and organisational culture, or the way that compensation and moral standards interact. And, while economists have modelled culture as a variety of devices that resolve communication or coordination problems in social life (Kreps (1990), Crémer (1993), Crémer, Garicano, and Prat (2004), Carillo and Gromb (1999, 2002), Van den Steen (2010a, 2010b)), none has shown how cultural learning can be subordinated to an appropriately designed compensation contract. Nor are we aware of any work in economics that combines the study of morals, compensation, and cultural learning.

Our analysis is presented as follows. Section 2 presents our model. Section 3 presents the model solution. Section 4 concludes. Proofs are collected in the Appendix.

2. Model

We consider a business run by a profit-maximising principal with a *senior agent*, s, and a *junior agent*, j, who provide a service to customers. For clarity we use female pronouns to refer to s and male pronouns for j. All of the agents in our model are risk-neutral: the principal and the customers aim to maximise their expected income, and we discuss the agents' preferences in Section 2.3 below.

The value v of the service to any customer is $v \equiv N(\bar{v}, 1/\tau_v)$; each time it provides the service, the business earns a profit $\pi \in \{\underline{\pi}, \bar{\pi}\}$, where $P[\pi = \bar{\pi}] = p$. We write $\Delta_\pi = \bar{\pi} - \underline{\pi}$.

We study the emergence of a new working practice $\mathscr{P}$ for delivering the service. We write I_s, $I_j \in \{0, 1\}$ for the respective decisions of the senior and junior agent to invoke $\mathscr{P}$. I_s and I_j are not verifiable, but the junior agent can observe the senior agent's adoption decision. If $\mathscr{P}$ is adopted, then it raises the probability that $\pi = \bar{\pi}$ to $p + \Delta_p < 1$. But the practice could be *harmful*, in which case it imposes a cost $c > 0$ on customers. The ex ante probability that $\mathscr{P}$ is harmful is $h > 0$, and we assume that

$$c > \Delta_p \Delta_\pi, \tag{1}$$

so that harmful practices are surplus-reductive. The ex ante expected surplus generated by $\mathscr{P}$ is $\Delta_p \Delta_\pi - hc$. We write

$$\hat{h} \triangleq \frac{\Delta_p \Delta_\pi}{c}; \tag{2}$$

$\mathscr{P}$ is surplus-reductive precisely when $h > \hat{h}$.

The senior agent s has a superior understanding of $\mathscr{P}$, because she receives a private signal $\sigma \in [0, 1]$ of its type before she decides whether to adopt it. σ is drawn from a distribution with function $F_H(\cdot)$ if the practice is harmful, and $F_L(\cdot)$ if it is not. We assume that the corresponding density functions $f_H(\cdot)$ and $f_L(\cdot)$ exist and are differentiable, and that $f_H(\sigma)/f_L(\sigma)$ is strictly monotonically increasing in σ. High signals σ are therefore suggestive of a harmful practice. We assume that $f_L(1) = 0 = f_H(0)$, so that signals of 0 or 1 reveal the type of $\mathscr{P}$ with certainty.

The junior agent does not receive a signal of the practice's type, but she observes the senior agent's invocation decision I_s before she makes her own choice I_j.

2.1 *Timings*

The timing of the game is as follows.

At time $t = 0$, the practice $\mathscr{P}$ emerges, and the principal offers a remuneration scheme to the agents. The principal makes take-it-or-leave it fee offers ϕ_s and ϕ_j to the customers of the senior and junior agent, respectively. Each agent acquires one customer, who commits to pay ϕ_s or ϕ_j for the service.

At time $t = 1$, the senior agent receives her private signal σ of the practice's type, and makes her private invocation decision I_s. The junior agent observes I_s.

At time $t = 2$, the junior agent makes his private invocation decision I_j.

At time $t = 3$, customer benefits and business profits are realised. The time 0 remuneration contracts are executed.

2.2 *Contracts*

The profits realised by both agents are verifiable, but customer benefits are not. Agent contracts are therefore conditional upon the business' profits. The junior agent moves after the senior agent and cannot affect her profits. His contract is therefore an ordered pair (w^j, w_j), comprising his income in the respective cases where the business realises a high or a low profit from its customer. The senior agent's contract is a four-tuple $(w^{sj}, w^s_j, w^j_s, w_{sj})$, describing her income as a function of the profit she and her junior earn: here subscripts denote low profit, and superscripts denote high profits so that, for example, w^s_j is the senior agent's profit if she earns a profit $\bar{\pi}$ and her junior earns profit $\underline{\pi}$.

2.3 *Objective functions and ethical standards*

As noted above, the principal aims to maximise the expected profits earned by the business, and each customer wishes to maximise his or her expected income. Agents are different, in that we allow them to exhibit concern for the ethical consequences of their actions.

As discussed in the Introduction, ethical values are hard to include in a traditional economic model. For example, it would be very difficult meaningfully to incorporate ethical concerns founded upon Aristotelian virtue in a model whose agents maximise an objective function. Moreover, economic reasoning cannot arbitrate between different visions of the good; the most we can hope to achieve is to understand the way that agents' ethical stance affects the social ordering. We therefore use an approach that allows us to derive results using standard optimisation methods, and that incorporates in a simple fashion the fact that some people are more likely than others to consider the ethical consequences of their actions. For an agent who earns lifetime income w and who provides service valued by customers at v, we use the following general objective function:

$$u = (1 - \varepsilon)w + \varepsilon e(w, v). \tag{3}$$

The parameter $\varepsilon \in [0, 1]$ in Equation (3) measures the strength of agent's *ethical commitment*; the function e represents the agent's *ethical standard*, which depends upon w, v and possibly upon some properties of their distribution. Where necessary, we use s and j subscripts to distinguish between the senior and junior agents' ethical commitment.[12]

[12]This formulation inevitably rules out some ethical standards. For example, we cannot use it to model situations in which an agent believes that her customers have misguided preferences. This situation might arise, for example, if the customers suffer from drug addiction; some writers have suggested that it could be a consequence of social indoctrination that leaves its subject unable to identify her best interests (Levy 2005). We do not attempt to address this difficult and important class of problems in our first pass at this topic

We consider two simple ethical standards, which correspond roughly to two ethical traditions. First, we consider a classical *act utilitarian*, who evaluates the ethical worth of an action by calculating the immediate effect of that action upon each of the affected parties. We formalise this standard in our model as follows:

Definition 1. *An* act utilitarian *agent uses the following ethical standard to evaluate the practice $\mathscr{P}$:*

$$\begin{aligned} e_{Act} &\triangleq E[\textit{surplus due to act}] \\ &= E[\textit{surplus accruing to bank}] - E[\textit{cost to customer}]. \end{aligned}$$

Note that, when assigning ethical worth to a practice, an act utilitarian is unconcerned with his or her own income *per se*: the bank surplus in the above definition is gross of any wage payments. We cite evidence in the Introduction that managers often use act-utilitarian reasoning when they face real-world ethical dilemmas.

We also present a simple formulation of the type of duty-based (deontological) ethics discussed in the Introduction. Such ethics prohibit actions that impose certain harms, irrespective of their consequences. We identify these ethics in our model with an other-regarding standard that prohibits actions that harm customers.

Definition 2. *An* other-regarding *agent uses the following duty-based ethical standard to evaluate the practice $\mathscr{P}$:*

$$e_{Duty} \triangleq -E[\textit{harm to the customer}] \tag{4}$$

Hence, in line with the discussion in the Introduction, the duty-based ethical standard embraced by an other-regarding agent does not permit inter-personal utility tradeoffs.

2.4 *Equilibrium definition*

A *strategy* in our model is a probability θ that $\mathscr{P}$ is invoked.

Definition 3. *An equilibrium of the model comprises:*

1. *Remuneration contracts (w^j, w_j) and $(w^{sj}, w_j^s, w_s^j, w_{sj})$ between the principal and the agents;*
2. *A strategy $\theta_s(\sigma)$ for the senior agent that depends upon her private signal σ;*
3. *Strategies θ_j^1 and θ_j^0 for the junior agent in the respective cases where the senior agent does and does not invoke $\mathscr{P}$;*

and, of course, we cannot determine whether or not "misguided preferences" exist.

such that each agent's strategy is optimal given the other's, and the remuneration contracts maximize the principal's expected income given the agents' strategies.

3. Game Solution

We solve the game by backward induction, starting in Section 3.1 by characterising the respective invocation decisions I_s and I_j of the senior and junior agents. We rely in our argument upon the senior agent's posterior assessment of the probability that $\mathscr{P}$ is harmful, which we denote by $\eta(\sigma)$.

Lemma 1 is a consequence of the monotonicity of $f_H(\sigma)/f_L(\sigma)$: a higher signal renders the senior agent more confident that$\mathscr{P}$ is harmful.

Lemma 1. *For any level ε_s of ethical commitment, the senior agent's posterior assessment $\eta(\sigma)$ of the probability that $\mathscr{P}$ is harmful is increasing in σ.*

A senior agent with signal σ expects $\mathscr{P}$ to increase aggregate surplus precisely when $\eta(\sigma) \leq \hat{h}$: that is, precisely when

$$\sigma \leq \hat{\sigma} \triangleq \eta^{-1}(\hat{h}). \tag{5}$$

The junior agent updates her priors after observing the senior agent's invocation decision and in light of the senior agent's equilibrium strategy. We write h_j^I for the junior agent's posterior evaluation of the probability that $\mathscr{P}$ is harmful after a senior agent invocation decision $I \in \{0, 1\}$; for notational simplicity, we sometimes suppress the dependence of h_j^I upon the senior agent's strategy $\theta_s(\sigma)$. We demonstrate below that, in equilibrium, the junior agent becomes more convinced of $\mathscr{P}$'s moral worth when $I_s = 1$.

3.1 *Agent invocation decisions*

Proposition 1, which we prove in the Appendix, demonstrates that both agents adopt an equilibrium cut-off strategy: the senior agent invokes $\mathscr{P}$ precisely when her signal σ is low enough, and the junior agent involves $\mathscr{P}$ when her posterior h_j^I of the probability that the practice is harmful is low enough.

Proposition 1. *Suppose that the senior and junior agents have respective levels of ethical commitment ε_s and ε_j, and let $(w^{sj}, w_j^s, w_s^j, w_{sj})$ and (w^j, w_j) be the respective remuneration contracts of the senior and junior agents. Let $R \in \{Act, Duty\}$ denote the agents' ethical standard.*

1. *There exists $\sigma^* \in [0, 1]$ such that the senior agent has the following optimal cutoff strategy:*

$$\theta_{s,R}(\sigma) = \begin{cases} 1 & \textit{if } \sigma < \sigma_R^*; \\ 0 & \textit{if } \sigma > \sigma_R^*. \end{cases} \tag{6}$$

If $0 < \sigma^*_R < 1$ *then the senior agent is indifferent between invocation and non-invocation when* $\sigma = \sigma^*_R$. $\sigma^*_{Duty} \le \sigma^*_{Act}$, *with strict inequality when* $\sigma^*_{Act} \in (0,1)$.

2. *For* $R \in \{Act, Duty\}$, *define* $\underline{h}_R$ *as follows:*

$$\underline{h}_R \triangleq \begin{cases} 0, & R = Duty, \\ \hat{h}, & R = Act. \end{cases} \tag{7}$$

Let

$$h^*_{j,R} \triangleq \underline{h}_R + \frac{1-\varepsilon_j}{\varepsilon_j}\frac{\Delta_p}{c}(w^j - w_j). \tag{8}$$

If agents follow ethical standard R, *then, given a senior agent invocation decision* $I \in \{0,1\}$, *the deputy agent has the following cutoff strategy* θ^I_j *:*

$$\theta^I_j = \begin{cases} 1, & \text{if } h^I_j < h^*_{j,R}; \\ 0, & \text{if } h^I_j \ge h^*_{j,R}. \end{cases} \tag{9}$$

When the ethical standard $R \in \{Act, Duty\}$ is either obvious from the context of the exposition or unimportant we reduce notational complexity by omitting the R subscript from θ_s, σ^*, h^l, and h^*_j.

The intuition for Proposition 1, and the interpretation of σ^*, is as follows.

Absent ethical concerns ($\varepsilon = 0$), the senior agent's decision to invoke $\mathscr{P}$ would be completely determined by her compensation contract. Hence, when $\varepsilon = 0$, either $\sigma^* = 1$ and invocation always occurs, or $\sigma^* = 0$ and it never does. Any variation in the senior agent's invocation choice must derive from her signal σ of the moral worth of $\mathscr{P}$, which can bite only if she has sufficient ethical commitment ε_s. Because $f_H(\sigma)/f_L(\sigma)$ is monotone in σ, the agent is relatively more certain that the practice reduces aggregate surplus for high σ, and that it is surplus-increasing for low σ. How she responds to this information depends upon her ethical standard. If she is an act-utilitiarian, then her ethical standard requires her to invoke $\mathscr{P}$ for low σ and not to do so for high σ; it follows that $\sigma^*_{Act} \in (0,1)$ for high enough ε_s. The duty-based standard of Equation (4) requires the agent never to invoke $\mathscr{P}$ if there is any chance that it is harmful to customers; when $\sigma = 0$ the practice cannot be harmful, in which case the other-regarding agent's invocation decision is determined by her compensation. Hence, either $\theta_{s,Duty}(0) = 0$ and $\sigma^*_{Duty} = 0$, or $\theta_{s,Duty}(0) = 1$, in which case $\sigma^*_{Duty} > 0$. $\mathscr{P}$ is always more morally acceptable for act utilitiarians than for other-regarding agents, so that $\sigma^*_{Duty} \le \sigma^*_{Act}$, with strict inequality when $\sigma^*_{Act} \in (0,1)$.

Similarly, the junior agent's invocation decision depends upon his moral judgement and his compensation. The former leads him never to invoke $\mathscr{P}$ when he has an other-regarding ethical standard, and, when he adheres to act-utilitarian standards, to do so precisely when

$\mathscr{P}$ is expected to increase aggregate surplus. In other words, the ethical standard $R \in \{Act, Duty\}$ requires the junior agent to invoke $\mathscr{P}$ precisely when $h < \underline{h}_R$; Equation (8) confirms that this rule applies when $w^j = w_j$. Differential pay for success and failure ($w^j \neq w_j$) alter the junior agent's cutoff $h^*_{j,R}$ by $\frac{1-\varepsilon_j}{\varepsilon_j}\frac{\Delta_p}{c}(w^j - w_j)$, and so induce the junior agent to deviate from the invocation strategy dictated by his ethical standard.

Proposition 1 states that, ceteris paribus, the senior agent is more willing to invoke $\mathscr{P}$ when she has a strong signal of the practice's moral worth. The junior agent's posterior assessment of the likelihood that $\mathscr{P}$ is harmful should therefore be lower when the senior agent invokes the practice than when she does not. Lemma 2 confirms this intuition.

Lemma 2. $h^1_j(\sigma^*) \leq \eta(\sigma^*) \leq h^0_j(\sigma^*)$: *that is, after senior agent invocation, the junior agent's posterior assessment of the likelihood that $\mathscr{P}$ is harmful decreases, while his assessment that $\mathscr{P}$ is harmful increases if the senior agent does not invoke the practice.*

For $R \in \{Act, Duty\}$, Lemma 2 allows us to characterise three relevant regions for $h^*_{j,R}$, each of which corresponds to a three different cultural norms:

Definition 4. *Suppose that the agents have ethical standard $R \in \{Act, Duty\}$. Then:*

1. *$h^*_{j,R}$ lies in the* rejection region *if $h^*_{j,R} < h^1_j$. In this case, the junior agent never invokes $\mathscr{P}$;*
2. *$h^*_{j,R}$ lies in the* cultural assimilation region *if $h^1_j \leq h^*_{j,R} < h^0_j$. In this case, the junior agent makes the same invocation decision as the senior agent;*
3. *h^*_j lies in the* acceptance region *if $h^0_j \leq h^*_{j,R}$. In that case, the junior agent always invokes $\mathscr{P}$.*

When $h^*_{j,R}$ lies in the rejection and acceptance regions we say that the junior agent accepts *and* rejects *$\mathscr{P}$, respectively; when $h^*_{j,R}$ lies in the cultural assimilation region we say that the junior agent's invocation decision is* culturally determined.

Figure 1 illustrates the strategic regions identified in Definition 4 as a function of the trigger levels σ^* and $h^*_{j,R}$. The Figure illustrates the case where $h < \hat{h}$ so that, ex ante, $\mathscr{P}$ is expected to increase surplus. We demonstrate in the Appendix (Lemma 10) that the boundaries h^0_j and h^1_j are increasing functions of σ^* with start and end points as illustrated. To understand the figure, note that, as the senior agent's trigger value σ^* increases, she is more willing to invoke $\mathscr{P}$. It follows that the senior agent's invocation of $\mathscr{P}$ sends a weaker signal of quality at higher trigger levels σ^* and, hence, corresponds to a higher junior agent posterior assessment h^1_j. Similarly, because the senior agent is more inclined to invoke $\mathscr{P}$ at higher σ^*, her rejection of the practice at higher σ^* sends a strong signal that the practice is harmful and, hence, results in a higher posterior junior agent assessment h^0_j.

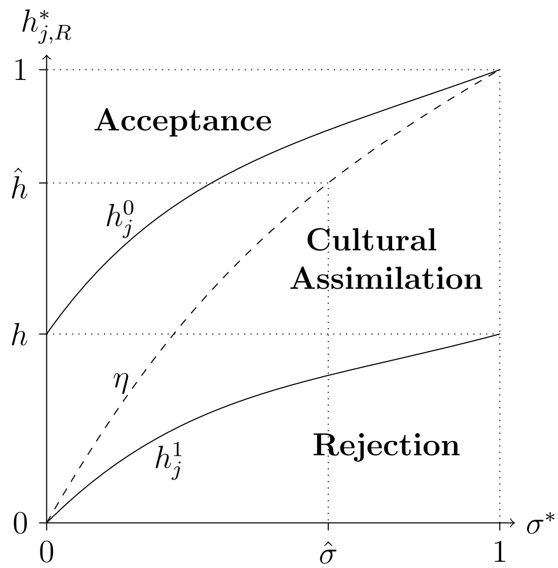


Figure 1. **Junior agent strategy regions**. For $R \in \{Act, Duty\}$, the junior agent's strategy depends upon the relationship between the trigger level $h^*_{j,R}$ and the values of h_j^0 and h_j^1. In the acceptance and rejection regions respectively, the junior agent always and never invokes $\mathscr{P}$. In the cultural assimilation region, the junior agent adopts the same social practices as the senior agent. The Figure illustrates the case where $h < \hat{h}$ so that the practice is ex ante expected to be surplus-enhancing.

In the region between h_j^1 and h_j^0, the junior agent follows the senior agent's lead, and accepts precisely those practices that she is prepared to endorse. Hence, for trigger values (σ^*, h_j^*) that fall in this region, the firm exhibits the type of cultural learning that we identify in the Introduction; even without any personal understanding of the moral status of the practice $\mathscr{P}$, the junior agent is prepared to adopt it when he sees the senior agent doing so.

The trigger values σ^* and h^* are determined endogenously as a function of the ethical commitments ε_s and ε_j of the agents, as well as of their compensation contracts $(w^{sj}, w_j^s, w_s^j, w_{sj})$ and (w^j, w_j). The principal therefore determines the equilibrium strategies of the agents and, hence, the cultural norms of the firm; we solve his problem in the following Section.

3.2 *The Principal's Problem*

The principal selects remuneration contracts (w^j, w_j) and $(w^{sj}, w_j^s, w_s^j, w_{sj})$ so as to implement cut-off values h_j^* and σ^* and, by extension, determines the relationship between h_j^* and the invocation boundaries $\left\{h_j^1(\sigma^*), h_j^0(\sigma^*)\right\}$. These choices therefore determine whether the junior agent culturally assimilates: that is, follows the senior's invocation decision. We assume that the agents are constrained to receive more for high than for low profitability:

$$w^j \geq w_j,\; w^{sj} \geq w_s^j,\; w_j^s \geq w_{sj}. \tag{10}$$

This is a natural requirement if the agent can hide or deliberately lower his/her profits.

Lemma 3 is an immediate consequence of Equations (8) and (10):

Lemma 3. *For $R \in \{Act, Duty\}$, the principal can implement any $h^*_{j,R} \in [\underline{h}_R, 1]$, where $\underline{h}_R$ is defined in Equation (7). The cheapest way to do so is to set $w_j = 0$ and*

$$w^j = \underline{w}^j_R(h^*_{j,R}) \triangleq \frac{\varepsilon_j}{1-\varepsilon_j}\frac{c}{\Delta_p}(h^*_j - \underline{h}_R). \tag{11}$$

The principal can use bonus payments to induce agents to set aside their moral objections to a social practice. But, because agents cannot be paid more for low profits than high (Equation (10)), bonuses can only induce agents to *expand* the range of beliefs for which they adopt the practice beyond the baseline dictated by their moral standard: that is, to induce junior agents to employ a cut-off above $\underline{h}_R$. Lemma 3 identifies the cost of doing so.

Corollary 1. *An act-utilitiarian junior agent's trigger level $h^*_{j,Act}$ can lie in the Rejection region if and only if ex ante $\mathscr{P}$ is expected to reduce surplus.*

Proof. Equation (8) and the fact that $w^j \geq w_j$ together imply that $h^*_{j,Act} \geq \hat{h}$. Then, because $h^1_j(\sigma^* = 1) = h$ (see Figure 1 and Lemma 10 in the Appendix), $h^*_{j,Act}$ can lie within the Rejection region if and only if $h > \hat{h}$: that is, if and only if $\mathscr{P}$ is surplus-reductive.

Like the junior agent, a senior agent with ethical standard $R \in \{Act, Duty\}$ is guided by her ethical standard to invoke $\mathscr{P}$ precisely when she assigns probability $\eta(\sigma) < \underline{h}_R$ that $\mathscr{P}$ is harmful: that is, when Condition (12) is satisfied:

$$\sigma < \underline{\sigma}_R \triangleq \begin{cases} 0, & \text{if } R = Duty; \\ \hat{\sigma}, & \text{if } R = Act. \end{cases} \tag{12}$$

Since success payments are constrained weakly to exceed failure payments, the senior agent can be induced by high enough bonus payments to employ a cut-off that exceeds $\underline{\sigma}_R$. However, as Lemma 4 demonstrates, this result is complicated when agents are act-utilitarians by the possible effect that the senior agent's invocation decision I_s has upon the junior agent's choice I_j.

Lemma 4. *Suppose that the agents have ethical standard $R \in \{Act, Duty\}$.*

1. *The principal can implement any cutoff $\sigma^* \geq \underline{\sigma}_R$. The cheapest way to do so depends upon the choice of h^*_j as follows:*
   (a) *If h^*_j lies in the Cultural Assimilation Region then $w^s_j = w^j_s = w_{sj} = 0$ and*

$$w^{sj} = \frac{\varepsilon_s}{1-\varepsilon_s}\frac{c}{\Delta_p}\left(\eta(\sigma^*) - \underline{h}_R\right)\frac{1}{2p+\Delta_p}; \tag{13}$$

(b) *If h_j^* lies in the Acceptance Region then $w_s^j = w_{sj} = 0$ and w^{sj}, w_j^s are selected so as to satisfy Equation* (14)*:*

$$(p+\Delta_p)w^{sj} + (1-p-\Delta_p)w_j^s = \frac{\varepsilon_s}{1-\varepsilon_s}\frac{c}{\Delta_p}\left(\eta(\sigma^*) - \underline{h}_R\right). \tag{14}$$

(c) *If h_j^* lies in the Rejection Region then $w_s^j = w_{sj} = 0$ and w^{sj}, w_j^s are selected so as to satisfy Equation (15):*

$$pw^{sj} + (1-p)\, w_j^s = \frac{\varepsilon_s}{1-\varepsilon_s}\frac{c}{\Delta_p}\left(\eta\left(\sigma^*\right) - \underline{h}_R\right) \tag{15}$$

2. *If $R = Act$, $1-2p-\Delta_p < 0$, and $\hat{h} < h_j^0$, then the principal can implement any cutoff $\sigma^* < \hat{\sigma}$ by implementing h_j^* in the cultural assimilation region, and setting $w^{sj} = w_s^j = w_{sj} = 0$ and*

$$w_j^s = \frac{\varepsilon_s}{1-\varepsilon_s}\frac{c}{\Delta_p}\left(\eta(\sigma^*) - \underline{h}_R\right)\frac{1}{1-2p-\Delta_p}. \tag{16}$$

The proof of Lemma 4 appears in the Appendix.

Part 1 of the Lemma is analogous to Lemma 3: it states that the principal can induce a senior agent with ethical standard $R \in \{Act, Duty\}$ to adopt any cut-off value $\sigma \geq \underline{\sigma}_R$: in other words, that appropriate success bonus payments can induce the senior agent to invoke $\mathscr{P}$ even when her posterior assessment is that the practice is morally wrong. The bonus payments are increasing in the width of the range $\sigma \in [\underline{\sigma}_R, \sigma^*]$ of bad signals that the principal wishes to induce the senior agent to ignore.

Part 1(a) of the Lemma corresponds to the case where the junior agent's actions mirror the senior agent's. Paying the senior agent for junior agent success therefore heightens the senior agent's incentive to invoke $\mathscr{P}$ and, hence, the cheapest wage contract compensates the senior agent only when both agents succeed, as in Equation (13). The cost of inducing the senior agent to set aside her moral scruples is higher when she is more ethically committed; hence, the bonus payment w^{sj} is increasing in ε_s.

Part 1(b) obtains when the junior agent always invokes $\mathscr{P}$. Since the senior agent cannot affect the junior agent's behaviour in this case, her bonus is paid conditional upon her own success, as in Equation (14), and it does not matter how the payments are spread between the cases where the junior agent does and does not generate a high profit. The probability weights in Equation (14) reflect the fact that the junior agent always invokes the practice and so generates a high profit with probability $p + \Delta_p$.

Part 1(c) corresponds to the case where the junior agent rejects $\mathscr{P}$. As in the Acceptance case (part 1(b)), the senior agent's actions in this case cannot influence the junior agent's choice, and analogous reasoning applies: payments are once again spread between the cases where the junior agent does and does not generate high profit. Note the junior agent in this case generates a high profit with probability p.

Part 2 of the Lemma applies only when agents are act-utilitarian, and has no analogue for the junior agent. When it obtains, $\sigma^* < \hat{\sigma}$, so that the senior agent is induced not to invoke a morally desirable practice whenever $\sigma^* \leq \sigma < \hat{\sigma}$.[13] This can occur only when cultural assimilation occurs because, if the junior agent were always to invoke $\mathscr{P}$, a senior agent wage constrained to rise with profits (Equation (10)) can only ever increase adoption rates. By Lemma 3, it is impossible to implement a junior agent cutoff h_j^* below $\hat{h}$; cultural assimilation can therefore be achieved only when $\hat{h} < h_j^0$, as in the statement of the Lemma. Within the Cultural Assimilation Region paying the senior agent a bonus w^{sj} when both agents succeed incentivises invocation and so cannot cause the senior agent to reject a surplus-enhancing practice. It follows that the only lever left to the principal is to reward the senior agent for low junior agent profit by setting $w_j^s > 0$. Doing so has two consequences. On the one hand, invoking $\mathscr{P}$ renders high senior agent profit more likely; on the other, it causes junior agent invocation, and so reduces the chances of low junior agent profitability. This tradeoff resolves itself against invocation when the probability $(p + \Delta_p)(1 - p - \Delta_p)$ of profitabilities $(\bar{\pi}, \underline{\pi})$ with invocation is less than the corresponding probability $p(1-p)$ without: that is, when $1 - 2p - \Delta_p < 0$, as in the statement of the Lemma.

We are now in a position to determine the principal's preferred remuneration contracts, (w^j, w_j) and $(w^{sj}, w_j^s, w_s^j, w_{sj})$. The principal's pricing policy can then be determined and, hence, the principal's profits. Our results depend upon customer sophistication, which we characterise in definition 5:

Definition 5. *Customers are* sophisticated *if they correctly anticipate the equilibrium of the game and are prepared to pay the expected value of their services in that equilibrium. Customers are* naïve *if they do not appreciate the consequences of $\mathscr{P}$, and always pay $\bar{v}$ for the service.*

The principal can use the senior agent's compensation contract to select a σ^* value, and must then decide in which of the regions of Figure 1 to locate the junior agent's trigger h_j^*. We consider the respective cases where customers are sophisticated and unsophisticated in Sections 3.3 and 3.4.

3.3 *Equilibrium with sophisticated customers*

We present our analysis for act-utilitarian and other-regarding agents in turn.

[13] $\mathscr{P}$ can never be socially desirable under our version of duty-based ethics: this is the reason that part 2 of the Lemma applies only to act-utilitarians.

3.3.1 Act-utilitarian agents

Proposition 2. *Suppose that the firm's customers are sophisticated and that agents are act-utilitarians. Then the principal sets wages equal to the outside option of zero, irrespective of the state of the world; furthermore, $\sigma^* = \hat{\sigma}$ and the junior agent's investment decision is culturally determined.*

The proof of Proposition 2 is in the Appendix. The intuition for the result is as follows.

Because sophisticated customers pay exactly their expected income from the firm's service, the principal earns the total expected surplus from service provision, less the expected value of any wages. The expected surplus is maximised if the senior agent invokes $\mathscr{P}$ whenever $\sigma \leq \hat{\sigma}$ and the junior agent's invocation decision is culturally determined. Because every agent has non-zero ethical commitment, the principal can induce this behaviour by paying all agents a zero wage. Because this contract maximises surplus and minimises expected wage payments, it is identified in Proposition 2 as optimal.

Note that, while the principal can achieve higher ex post profits by inducing junior agents always to accept $\mathscr{P}$, sophisticated customers anticipate the harm imposed upon them by this policy, and they force the principal to absorb that harm by paying lower prices for the service.

Lemma 5. *When the firm's clients are sophisticated and agents are act-utilitarians, the principal's expected profits are unaffected by the strength of the agents' ethical commitment.*

Lemma 5 is a consequence of the fact that, when customers are sophisticated, the firm induces the socially optimal investment policy. Since the firm has nothing to gain from paying its agents to set aside their ethical concerns, the cost of doing so, which depends upon their ethical commitment ε, is irrelevant.

The agents are optimally paid zero in Proposition 2 because we have normalised their outside option to zero. If the outside option were positive, the principal would still seek the cheapest way to ensure that all agents invoke $\mathscr{P}$ precisely when $\sigma \leq \sigma^*$. In this case, paying a flat wage equal to the outside option will induce a senior agent with positive ethical commitment to adopt the surplus-maximising invocation policy; in that case, paying the junior agent a flat wage induces cultural assimilation. Hence, a state-independent wage remains optimal when the agents are act-utilitarians. Furthermore, the expected cost of any other compensation scheme cannot be lower and, if compensation is not state-independent, it can serve to move the agents away from the surplus-maximising adoption policy and, hence, to lower expected profits.

3.3.2 Other-regarding agents

Surplus is maximised with sophisticated customers in Proposition 2 because act-utilitarian agents do not need to be paid to select the surplus-maximising invocation policy. In contrast, other-regarding agents who adopt the duty ethical standard of Definition 2 will only invoke $\mathscr{P}$ if they are paid to do so. Consequently, as we demonstrate in Proposition 3, surplus is not maximised with other-regarding agents.

Proposition 3. *Suppose that the firm's customers are sophisticated and that agents are other-regarding. Then* $\sigma^*_{Duty} < \hat{\sigma}$ *with* $\eta(\sigma^*_{Duty}) < \hat{h}/(1+\varepsilon_s/2(1+\varepsilon_s))$*, and the principal uses bonus contracts with wages above zero for agents whose ethical standards are low enough. Equilibria with junior agent cultural assimilation, acceptance, and rejection are all possible.*

To understand Proposition 3, note that, while surplus is maximised when $\sigma^* = \hat{\sigma}$, a $\sigma^* > 0$ can be achieved when agents are other-regarding only by paying success bonuses that increase in σ^*. The principal earns the expected surplus less the expected wage bill. A marginal reduction in σ^* from $\hat{\sigma}$ has no effect upon surplus, but lowers expected wages. Hence $\sigma^*_{Duty} < \hat{\sigma}$.

That cultural assimilation, acceptance, and rejection are all feasible in equilibrium with other-regarding agents is most easily demonstrated with examples.

Cultural assimilation.—Suppose that $\varepsilon_s = \varepsilon_j = 0$. Then σ^* can be achieved with arbitrarily small wage payments. In particular, the surplus-maximising equilibrium of Proposition 2, with $\sigma^* = \hat{\sigma}$ and cultural assimilation can be achieved; it is optimal because it maximises expected surplus, all of which the principal receives.

Acceptance.—Suppose that $\varepsilon_s = 1$ and $\varepsilon_j = 0$. Then the senior agent will never invoke $\mathscr{P}$: $\sigma^*_{Duty} = 0$ and the senior agent is paid zero. In contrast, an arbitrarily small payment can induce junior agent acceptance. Since the senior agent's action is uninformative, junior agent acceptance is surplus-maximising whenever $\mathscr{P}$ is ex ante expected to be surplus-enhancing (i.e., when $\hat{h} > h$).

Rejection.—Suppose that $\varepsilon_s = 0$ and $\varepsilon_j = 1$. Then junior rejection is inevitable, and an arbitrarily small bonus induces the senior agent to set $\sigma^* = \hat{\sigma}$.

Precisely which type of equilibrium obtains with other-regarding agents depends upon the properties of the density functions $f_L(\cdot)$ and $f_H(\cdot)$ and, hence, it is not possible in general to provide a complete equilibrium characterisation away from the neighbourhoods identified in the above examples. We present complete numerical examples in Section 3.5.1 below.

In contrast to the case with act-utilitarian agents, the principal makes higher profits when its agents are less ethically committed:

Lemma 6. *When the firm's clients are sophisticated and agents are other-regarding, the principal's profit is diminishing in the strength of the agents' ethical commitment.*

To understand Lemma 6, recall that other-regarding agents are required by their moral standard never to invoke $\mathscr{P}$ and, hence, do so only when their success bonus is large enough to overwhelm their moral qualms. The required scale of the bonus is lower, and the cutoff σ^*_{Duty} is higher, when the agents are less ethically committed. Hence, lowering agents' ethical commitment serves both to lower the expected wage bill for a given level of surplus and, by raising σ^*_{Duty}, to raise surplus. Both effects serve to increase the principal's expected profit.

The principal in our model makes a take-it-or-leave it offer to the customer, who is therefore indifferent over the agent's ethical standard. But we could extend our analysis to allow for surplus to be ex-ante shared between the principal and the customer. If we were to do so, then Lemma 6 demonstrates that, faced with other-regarding agents, sophisticated customers would have a positive preference for less ethical agents. This situation can be thought of as one in which ethical agents attempt to look out for the customer's best interests by refusing to inflict any harm upon him. A sophisticated customer may (correctly) believe that he is better placed to assess his expected well-being than his market counterparties and, hence, may prefer to deal with agents who do not even consider his well-being.

3.4 *Equilibrium with naïve investors*

We now consider a firm whose customers are naïve and, hence, pay $\bar{v}$ for the firm's service, irrespective of whether $\mathscr{P}$ is invoked. Once again, we consider act-utilitarian and other-regarding agents in turn.

3.4.1 Act-utilitarian agents

In order to present our formal result as succinctly as possible, we introduce some new notation, and we define a new formal term. First, we define the *ethical power* $\mathscr{E}_i \in \Re_{\geq 0} \cup \{\infty\}$ of agent $i \in \{s, j\}$ to be the following transformation of the agent's ethical commitment:

$$\mathscr{E}_i \triangleq \frac{\varepsilon_i}{1-\varepsilon_i}. \tag{17}$$

Second, we say that the agents' ethical powers are *limited by* $(\kappa_s, \kappa_j) \geq (0, 0)$ precisely when

$$\frac{\mathscr{E}_s}{\kappa_s} + \frac{\mathscr{E}_j}{\kappa_j} \leq 1. \tag{18}$$

Ethical limits place an upper bound upon the agents' ethical powers as illustrated in Figure 2.

We can now state our main result.

Proposition 4. *Suppose that agents are act-utilitarians and that the firm's customers are naïve. Then the principal selects $\sigma^*_{Act} \geq \hat{\sigma}$. There exist ethical constraints $K^{Ass} = (\kappa_s^{Ass}, 0)$,*

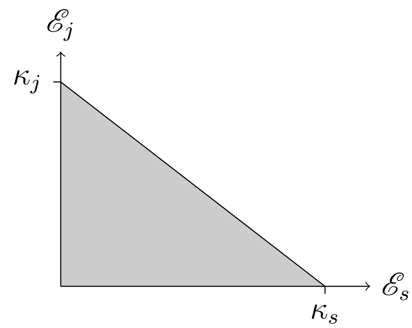


Figure 2. **Ethical limits**. Agents whose ethical powers are limited by (κ_s, κ_j) have ethical powers $(\mathscr{E}_s, \mathscr{E}_j)$ that lie within the shaded region.

$K^{Acc} = (\kappa_s^{Acc}, \kappa_j^{Acc})$ and $K^* = (\kappa_s^*, \kappa_j^*)$ such that:

1. *The optimal σ^*_{Act} in the Cultural Assimilation Region is greater than $\hat{\sigma}$ if $(\mathscr{E}_s, \mathscr{E}_j)$ are limited by K^{Ass};*
2. *The optimal σ^*_{Act} in the Acceptance Region is greater than $\hat{\sigma}$ if $(\mathscr{E}_s, \mathscr{E}_j)$ are limited by K^{Acc};*
3. *The optimal σ^*_{Act} in the Rejection Region is strictly greater than $\hat{\sigma}$;*
4. *The principal induces the junior agent to accept $\mathscr{P}$ if and only if $(\mathscr{E}_s, \mathscr{E}_j)$ are limited by K^*.*

The proof of the Proposition appears in the Appendix, along with expressions for K^{Ass}, K^{Acc} and K^*. Its intuition is as follows.

When customers are naïve, they do not charge the firm for the harm that they anticipate from $\mathscr{P}$. It is therefore in the principal's interest that $\mathscr{P}$ be implemented; only the moral concerns of the agents prevent this from happening. The principal can overcome those concerns by paying a sufficiently high success bonus to the agents; the scale of the necessary bonus reflects the moral concerns of the agents. When the junior agent's actions are culturally determined, the principal need only pay the senior agent: the cost of incentivising $\sigma^*_{Act} > \hat{\sigma}$ therefore reflects only the senior agent's moral power, so that it is worth setting $\sigma^*_{Act} > \hat{\sigma}$ only for ethical limits of the form K^{Ass} in Proposition 4. In contrast, in the Acceptance region, as well as inducing the senior agent to set aside her moral scruples, the principal must also compensate the junior agent for doing so. It is more expensive to incentivise the junior agent to accept $\mathscr{P}$ when σ^*_{Act} is high, because the senior agent's invocation decision is then less informative; the necessary limits on $(\mathscr{E}_s, \mathscr{E}_j)$ therefore have the form of K^{Acc}. If the practice is ex ante expected surplus-reductive then compensation contracts that implement the Rejection Region for the junior agent exist. These require the junior agent to reject a practice even when the senior agent has accepted it. This is possible only if the senior agent is willing to implement surplus-reductive practices: that is, if $\sigma^*_{Act} > \hat{\sigma}$.

Finally, the principal decides whether to force junior agent acceptance by trading off the extra profits thereby achieved against the costs of persuading the junior agent to ignore the senior agent's actions. The latter costs are low, and the principal forces acceptance, when the agents' ethical powers are sufficiently limited, as in part 4 of the Proposition.

We have established that with naïve consumers the principal will choose to set a trigger for the senior agent $\sigma^*_{Act} \geq \hat{\sigma}$ with strict inequality for sufficiently ethically limited agents. Because it is harder to induce ethically committed agents to accept a harmful practice, those agents receive higher bonuses and, hence, generate lower profits for the firm. Lemma 7, which is proved in the Appendix, confirms this intuition.

Lemma 7. *Suppose that the firm's clients are naïve and that agents are act-utilitarians. Then the principal's expected profits are declining in the strength of the ethical commitment of the agents.*

It follows that a firm with naïve consumers and act-utilitarian agents prefers to hire agents with low ethical power. The principal benefits by incentivising the senior agent to invoke practices whose expected surplus is overall negative. The lower the senior's ethical commitment, the cheaper this is.

Our analysis identifies three mechanisms by which act-utilitarian agents can be incentivised to invoke surplus-reductive practices. If the practice is ex ante expected to be surplus-reductive then the principal could implement the Rejection Region for the junior and incentivise senior agent adoption. This approach is optimal when the junior agent is sufficiently ethically committed. If the practice is ex ante surplus-enhancing, or if it is surplus-reductive and the junior agent is not too ethically committed, then the principal has two choices. First, she can induce cultural assimilation, and reward the senior agent only when both senior and junior agents are successful: that is, with a bonus based on the whole team's performance. Second, it could incentivise the junior agent always to adopt $\mathscr{P}$, by paying a bonus to the junior that could exceed the senior agent's bonus, which, if it is non-zero, is contingent only upon the senior agent's profitability. The second approach is preferred when both agents have sufficiently low ethical commitment; the first is adopted for intermediate levels of ethical commitment. In general, as stated in Lemma 7, all wages fall as the agents' ethical commitments fall.

3.4.2 Other-regarding agents

As stated in Proposition 3, cultural assimilation, acceptance, and rejection are all possible for other-regarding agents even when their customers are sophisticated. In that case, customers reduce their willingness to pay in response to surplus-reducing actions. Hence, because it is expensive to induce higher σ^*_{Duty}, it is always the case with sophisticated customers that

$\sigma^*_{Duty} \leq \hat{\sigma}$.

In contrast, surplus-reducing actions have no effect upon naïve customers' willingness to pay. Hence, because a higher σ^*_{Duty} raises firm profitability, $\sigma^*_{Duty} > \hat{\sigma}$ is possible when customers are naïve. We have therefore proved the following result:

Proposition 5. *Suppose that the firm's customers are naïve and that agents are other-regarding. Then equilibria with junior agent cultural assimilation, acceptance, and rejection are all possible. Any $\sigma^*_{Duty} \in [0, 1]$ is possible.*

Finally, we have the following analogue of Lemma 7:

Lemma 8. *When the firm's clients are naïve and agents are other-regarding, the principal's profit is diminishing in the strength of the agent's ethical commitment.*

3.5 *Numerical example*

Equilibria with other-regarding agents cannot be fully characterised analytically, because they depend upon the specific properties of the density functions f_L and f_H. This section presents our model's equilibria for a specific case with other-regarding agents.

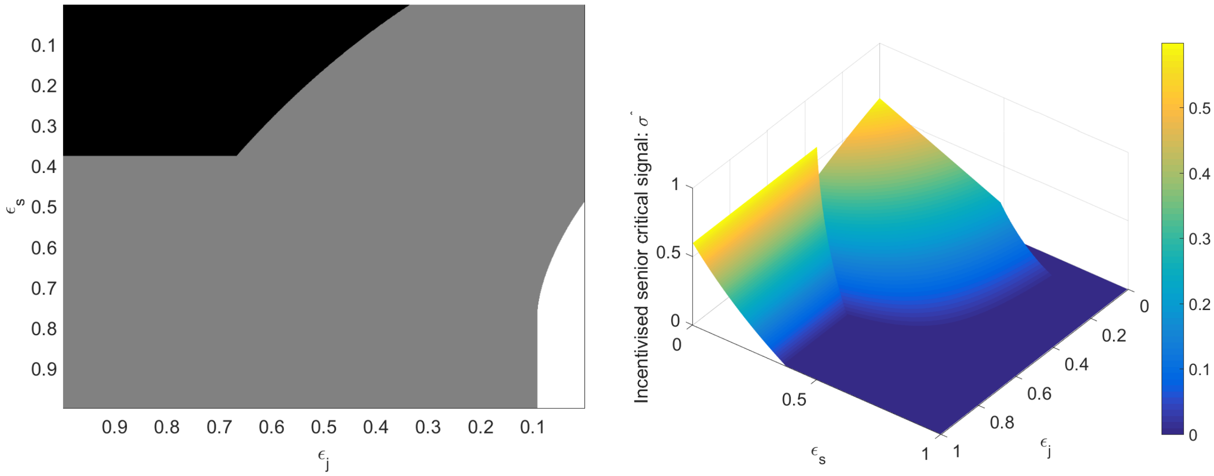
We consider density functions $f_H(\sigma) = 2\sigma$ and $f_L(\sigma) = 2(1 - \sigma)$ and set the prior probability h that $\mathscr{P}$ is harmful to $\frac{1}{2}$. This yields $\eta(\sigma^*) = \sigma^*$, $h^1_j(\sigma^*) = \sigma^*/2$, and $h^0_j(\sigma^*) = (\sigma + 1)/2$. We also set $\hat{h} = 0.6 > h = 1/2$, so that $\mathscr{P}$ is ex-ante surplus-enhancing.

3.5.1 Sophisticated customers and other-regarding agents

Figure 3 illustrates equilibria with sophisticated customers and other-regarding agents. The discussion after Proposition 3 demonstrates that, in this case, cultural assimilation occurs when $\varepsilon_j = \varepsilon_s = 0$, that junior agent acceptance occurs when $\varepsilon_j = 0$ and $\varepsilon_s = 1$, and that junior agent rejection occurs when $\varepsilon_j = 1$ and $\varepsilon_s = 0$. Figure 3a confirms that this is the case with our example distribution functions: rejection occurs in the black region, cultural assimilation in the grey region, and acceptance in the white region.

Figure 3b plots the senior agent's cut-off signal σ^*_{Duty} as a function of ε_j and ε_s; recall that the senior agent invokes $\mathscr{P}$ when $\sigma < \sigma^*_{Duty}$. Invocation never occurs for high enough ϵ_s.

The comparative statics of σ^*_{Duty} with respect to ethical commitment are clear from Figure 3b when σ^*_{Duty} is positive. First, consider the case where junior agents reject $\mathscr{P}$; this occurs for low ε_s and high ε_j. Since the senior agent's action does not affect the junior agent's invocation decision within this region, σ^*_{Duty} is unaffected by ε_j; since a less ethical agent is more willing to invoke $\mathscr{P}$, the cutoff signal σ^*_{Duty} is decreasing in ε_s. Now consider the cultural assimilation region, which occurs for low ε_s and ε_j. Once again, σ^*_{Duty} is decreasing



(a) Junior agent equilibrium behaviour

(b) Cut-off signal σ^*_{Duty}

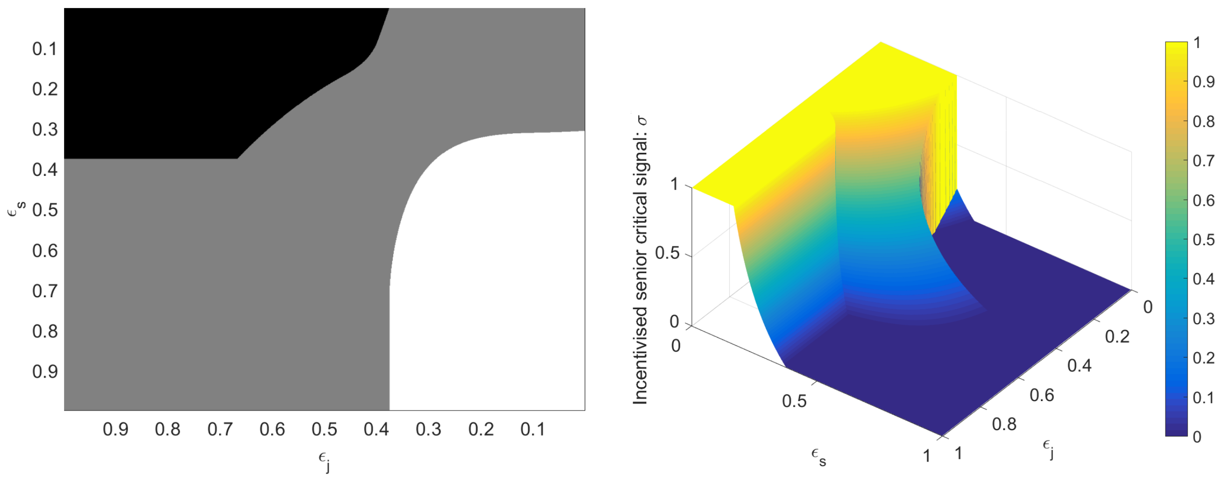
Figure 3. **Sophisticated customers and other-regarding agents**. Figure 3a illustrates the junior agent's equilibrium response to the senior agent's invocation decision. For the parameter values coloured black at the top left of the Figure, the junior agent always rejects $\mathscr{P}$, and for parameters coloured white at the bottom right the junior agent always accepts $\mathscr{P}$. Within the grey parameter region, cultural assimilation occurs, and the junior agent follows the senior agent's invocation decision. Figure 3b plots the senior agent's cutoff signal σ^*_{Duty} as a function of the ethical commitment ε_s and ε_j of the senior and junior agents.

in ε_s within this region. But, as illustrated in Figure 3b, σ^*_{Duty} is also decreasing in ε_j in the cultural assimilation region. This follows because the senior agent's invocation decision is mirrored within the cultural assimilation region by the junior agent. But an other-regarding junior agent is required by his moral standard never to invoke $\mathscr{P}$; he therefore does so only if he receives a sufficiently large bonus. The required size of that bonus is increasing in ε_j and, hence, so too is the total cost to the principal of inducing invocation. As a result, the principal implements a lower σ^*_{Duty} when ε_j is higher.

Note that σ^*_{Duty} is discontinuous across the border between the cultural assimilation and acceptance regions. The reason is that, as the firm moves from the cultural assimilation region to the rejection region, it no longer spends anything inducing invocation by the junior agent. It therefore becomes cheaper to induce the senior agent to invoke $\mathscr{P}$ for a given ethical commitment ε_s; as a result, the principal is more willing to induce invocation so that, as illustrated in Figure 3b, σ^*_{Duty} increases at the border between the two regions.

3.5.2 Naïve customers and other-regarding agents

Figure 4 is the analogue of Figure 3 in the case where customers are naïve: Figure 4a indicates for each $(\varepsilon_j, \varepsilon_s)$ pair whether an acceptance, rejection, or cultural assimilation equilibrium obtains, and Figure 4b illustrates the senior agent's cutoff σ^*_{Duty} for $\mathscr{P}$ invocation as a function of ε_s and ε_j.



(a) Junior agent equilibrium behaviour

(b) Cut-off signal σ^*_{Duty}

Figure 4. **Naïve customers and other-regarding agents**. Figure 4a illustrates the junior agent's equilibrium response to the senior agent's invocation decision. For the parameter values coloured black at the top left of the Figure, the junior agent always rejects $\mathscr{P}$, and for parameters coloured white at the bottom right the junior agent always accepts $\mathscr{P}$. Within the grey parameter region, cultural assimilation occurs, and the junior agent follows the senior agent's invocation decision. Figure 4b plots the senior agent's cutoff signal σ^*_{Duty} as a function of the ethical commitment ε_s and ε_j of the senior and junior agents.

The intuition for the shapes of Figures 4a and 4b is similar to that of Figures 3a and 3b. However, because naïve customers do not charge for expected harm, it is cheaper to induce invocation. As a result, the acceptance region in Figure 4a is much larger than the corresponding region in Figure 3a. And for low ε_s and ε_j, cultural assimilation amounts to automatic invocation by the senior agent, followed by assimilation by the junior agent. Hence, in this region of the cultural assimilation region, $\sigma^*_{Duty} = 1$, as illustrated in Figure 4b.

4. Conclusion

We present a model in which morally aware agents decide whether to invoke a social practice that may harm their customers. They are motivated partly by their moral standards, and partly by their compensation contract, which is designed by an amoral profit-maximising principal. Our analysis relates compensation contracts, ethical standards, and organisational profitability.

Cultural assimilation occurs in our model when junior agents opt to follow the invocation decision of their senior colleagues. Hence, the importance of "tone from the top" emerges naturally in our set-up. In our model, the tone that emerges at the top of the organisation is determined partly by the ethical standards of the organisation's senior employees, and partly by their wage contract. The wage contract is affected in turn by sophistication of the organisation's customer base. When customers are sufficiently sophisticated to figure out the effects of compensation policies, they penalise the firm for surplus-reductive activities. As a result, firms with sophisticated customers and utilitarian agents leave their employees free to exercise their moral judgement, and, in expectation, they serve client interests effectively. In contrast, left to their own devices, other-regarding agents would not maximise surplus and, hence, they receive equilibrium bonuses, which are designed to induce invocation as close as possible to the surplus-maximising policy as is cost-effective.

In contrast, naïve customers are unable to penalise bad behaviour, so that a profit-maximising principal attempts to induce its agents to set aside their moral scruples; in our set-up, they do this in pursuit of success bonuses. One of the equilibria that arises with naïve consumers retains cultural assimilation, but does so with perverse senior agent incentives, so that the tone from the top encourages harmful social practices whenever they are sufficiently profitable.

Our analysis sheds further light upon the United Kingdom's PPI mis-selling scandal, which we outline in the Introduction. The evidence concerning PPI strongly suggests that it was costly and pointless for most customers, but valuable for a few. The aggressive sales techniques documented by the CAB and the Banking Standards Commission were a profitable but potentially socially very damaging practice. Precisely because customers were naïve, it was particularly important that this practice be tempered by the moral sensibilities their financial advisors. And yet, in line with our model, it appears that compensation schemes and sales cultures were designed actively to undermine those sensibilities.

PPI selling bonuses appear to have been in sharp contradistinction to those deployed in the optimal contracting literature, which are designed to provide incentives that overcome a moral hazard problem that arises when managerial effort is not subject to contract. Giving managers "skin in the game" in this situation can move outcomes closer to the first best; in our model, doing so serves only to undermine managers' natural inclination to protect their customers. When values are important and customers are naïve, skin in the game could be harmful.[14]

We do not claim that performance-related payment is never justifiable in retail businesses whose customers are naïve. But our analysis does at least suggest that the decision to use such payments should be weighed against the possible effects that they may have upon

[14]This analysis may go some way towards resolving confusion expressed by the Archbishop of Canterbury, Justin Welbey, during oral submissions by the Group Chief Executive and Group Chairman of HSBC Holdings to the UK's Parliamentary Commission on Banking Standards. On that occasion, Archbishop Welbey said "We do not give skin in the game to civil servants, surgeons or teachers; there is a whole range of people who do not have that. It seems to me that you are running what you quite rightly describe as, and are putting huge effort into, a values-based organisation, with a strong values-based culture. Yet, at the end of the day, particularly for your most senior staff who are most important as regards setting values and culture, you seem to be saying that the only way you can motivate them to any significant extent is with cash, deferred or otherwise." (Parliamentary Commission on Banking Standards 2013b)

ethical decisions and the cultural context within which they are made. When retail financial services firms pay profit bonuses to sales people, they may undermine ethical standards in two ways: first, by encouraging employees to ignore their ethical concerns; and second, through their long term effect upon an employee pool in which it is cheapest to induce the desired tone-at-the-top amongst ethically uncommitted workers.

This reasoning suggests that optimal financial regulation is predicated upon the sophistication of the customer base. Customers who understand the impact that high-powered incentive schemes have upon the quality of service that they receive respond through the price system; principals are thus induced to nurture and to harness the ethical sensibilities of their employees. Bonus contracts here are designed to overcome the moral qualms of agents with other-regarding ethical standards, and will result in insufficient invocation from the perspective of an aggregate surplus maximising regulator. Against that backdrop, regulation can restrict itself to basic disclosure and anti-fraud provisions. But when customers lack the sophistication to understand the incentives that their counterparties face, or do not have the market power adequately to respond to those incentives, a more interventionist approach may be required and, in particular, direct regulation of compensation structures may be appropriate. This prescription is complicated by the facts that, first, it is hard in many cases for an outsider the gauge a customer's sophistication; and, second, that harmful practices of omission are far harder to detect in practice than those of commission. We leave a complete analysis of this problem for future work.

References

Adamson, C.. 2013. The importance of culture in driving behaviors of firms and how the FCA will assess this. Speech at the CFA Society UK Professionalism Conference, Financial Conduct Authority, London.

Agarwal, S., G. Amromin, I. Ben-David, S. Chomsisengphete, and D. D. Evanoff. 2014. Predatory lending and the subprime crisis. *Journal of Financial Economics*, 113:29–52.

Aristotle. 2009. *The Nicomachean Ethics*, Oxford University Press.

Bar-Gill, O.. 2009. The law, economics and psychology of subprime mortgage contracts. *Cornell Law Review*, 94:1073–1151.

Bentham, J.. 2007 [1789]. *An introduction to the principles of morals and legislation*, Dover.

Carillo, J. D., and D. Gromb. 1999. On the Strength of Corporate Cultures. *European Economic Review*, 43:1021 – 1037.

Carillo, J. D., and D. Gromb. 2002. Cultural Inertia and Uniformity in Organizations. Working paper, London Business School, London, U.K.

Célérie, C., and B. Vallée. 2013. What drives financial complexity? A look into the retail market for structured products. Research Paper FIN-2013-1013, HEC Paris.

Cohn, A., E. Fehr, and M. A. Maréchal. 2014. Business culture and dishonesty in the banking industry. *Nature*, 516:86–89.

Crémer, J.. 1993. Corporate Culture and Shared Knowledge. *Industrial and Corporate Change*, 3:351 – 386.

Crémer, J., L. Garicano, and A. Prat. 2004. Codes in Organizations. Discussion Paper 4205, Centre for Economic Policy Research, London.

Dudley, W. C.. 2014. Enhancing financial stability by improving culture in the financial services industry. Speech at the workshop on "reforming culture and behavior in the financial services industry", Federal Reserve Bank of New York, New York City.

Efing, M., H. Hau, P. Kampfkotter, and J. Steinbrecher. 2015. Incentive pay and bank risk-taking: Evidence from Austrian, German, and Swiss banks. *Journal of International Economics*, 96:S123–S140.

Fahlenbrach, R., and R. Stulz. 2011. Bank CEO incentives and the credit crisis. *Journal of Financial Economics*, 99:11–26.

Financial Services Authority. 2013. Parliamentary Commission on Banking Standards: Written evidence from the Financial Services Authority. Report to parliamentary committee, FSA.

Fritzsche, D. J., and H. Becker. 1984. Linking management behavior to ethical philosophy—An empirical investigation. *Academy of Management Journal*, 27:166–175.

G30. 2015. *Banking conduct and culture: A call for sustains and comprehensive reform*, Group of Thirty.

Giorgi, S., C. Lockwood, and M. A. Glynn. 2015. The Many Faces of Culture: Making Sense of 30 Years of Research on Culture in Organization Studies. *Academy of Management Annals*, 9:1–54.

Graafland, J. J., and B. W. van de Ven. 2011. The credit crisis and the moral responsibility of professionals in finance. *Journal of Business Ethics*, 103:6–5–619.

Inderst, R., and M. Ottaviani. 2009. Misselling through agents. *American Economic Review*, 99:883–908.

Kant, I.. 2012 [1785]. *Groundwork of the Metaphysics of Morals*, Cambridge University Press.

Kreps, D. M.. 1990. Corporate Culture and Economic Theory. in James E. Alt, and Kenneth A. Shepsle (ed.), *Perspectives on Positive Political Economy* . pp. 90 – 143, Cambridge University Press, Cambridge, MA.

Levy, A.. 2005. *Femal chauvinist pigs*, Free Press.

Luyendijk, J.. 2016. Big banks still have a problem with ethics and morality. *The Guardian*, 18th January, 2016.

MacIntyre, A.. 2007. *After Virtue: A Study in Moral Theory*, Gerald Duckworth & Co. Ltd., 3rd edn.

Nozick, R.. 1974. *Anarchy, State, and Utopia*, Blackwell, Oxford, UK (2002).

Parliamentary Commission on Banking Standards. 2013a. Changing banking for good (Volume II). Report HL Paper 27-II and HC 175-II, House of Lords and House of Commons, London, UK.

Parliamentary Commission on Banking Standards. 2013b. Minutes of Evidence HL Paper 27-III/HC 175-III. Oral Evidence: Stuart Gulliver, Group Chief Executive, HSBC Holdings and Douglas Flint, Group Chairman, HSBC Holdings, House of Lords and House of Commons, London, UK.

Patterson, O.. 2014. Making Sense of Culture. *Annual Review of Sociology*, 40:1–30.

Premeaux, S. R.. 2004. The current link between management behavior and ethical philosophy. *Journal of Business Ethics*, 51:269–278.

Premeaux, S. R., and R. W. Mondy. 1993. Linking management behavior to ethical philosophy. *Journal of Business Ethics*, 12:349–357.

Rawls, J.. 1955. Two concepts of rules. *Philosophical Review*, 64:3–32.

Sapienza, P., and L. Zingales. 2012. A trust crisis. *International Review of Finance*, 12:123–131.

Schein, E. H.. 2004. *Organizational culture and leadership*, Jossey-Bass.

Swidler, A.. 1986. Culture in Action: Symbols and Strategies. *American Sociological Review*, 51:273–286.

Tett, G.. 2009. *Fool's gold: How unrestrained global greed corrupted a dream, shattered global markets and unleashed a catastrophe*, Little Brown.

Thanassoulis, J.. 2012. The case for intervening in bankers' pay. *Journal of Finance*, 67:849–895.

Thanassoulis, J.. 2013. Industry structure, executive pay and short-termism. *Management Science*, 59:402–419.

Tutton, P., and F. Hopwood Road. 2005. Protection racket: CAB evidence on the cost and effectiveness of payment protection insurance. Report, Citizens Advice Bureau.

Van den Steen, E.. 2010a. Culture Clash: The Costs and Benefits of Homogeneity. *Management Science*, 56:1718–1738.

Van den Steen, E.. 2010b. On the origin of shared beliefs (and corporate culture). *RAND Journal of Economics*, 41:617–648.

Wheatley, M.. 2012. *The Wheatley Review of LIBOR.*

Wilson, H.. 2012. Banks were 'unethical,' says FSA's Martin Wheatley. *Daily Telegraph*, 30th June, 2012.

Zingales, L.. 2015. Does Finance Benefit Society?. *Journal of Finance*, 70:1327–1363.

APPENDIX

We use the following technical result to prove several of our main results.

Lemma 9. *For every* $0 \leq \sigma \leq 1$,

$$\frac{F_L(\sigma)}{F_H(\sigma)} \geq \frac{f_L(\sigma)}{f_H(\sigma)} \geq \frac{1-F_L(\sigma)}{1-F_H(\sigma)}. \tag{19}$$

Proof. By assumption, $f_L(s)/f_H(s)$ is declining in s. It follows that, for every $a > 0$,

$$f_L(s+a) \leq \frac{f_L(s)}{f_H(s)} f_H(s+a). \tag{20}$$

Integrating Equation (20) over $a \in [0, 1-s]$ yields the following expression:

$$\int_s^1 f_L(t)dt \leq \frac{f_L(s)}{f_H(s)} \int_s^1 f_H(t)dt,$$

or

$$\frac{1-F_L(s)}{1-F_H(s)} \leq \frac{f_L(s)}{f_H(s)}.$$

The other inequality is derived by integrating equation (20) over $a \in [-s, 0]$.

Proof of Lemma 1.

Given a signal σ, the senior agent assesses probability $\eta(\sigma)$ that $\mathscr{P}$ is harmful, where

$$\begin{aligned}\eta(\sigma) \triangleq P[\mathscr{P} \text{ harmful}|\sigma] &= \frac{P[\sigma|\mathscr{P} \text{ harmful}]\, P[\mathscr{P} \text{ harmful}]}{P[\sigma]} \\ &= \frac{hf_H(\sigma)}{hf_H(\sigma)+(1-h)f_L(\sigma)} \\ &= \frac{1}{1+\frac{1-h}{h}\frac{f_L(\sigma)}{f_H(\sigma)}}.\end{aligned} \tag{21}$$

Then, because $f_L(\sigma)/f_H(\sigma)$ is monotonically decreasing in σ, $\eta(\sigma)$ is increasing in σ.

Proof of Proposition 1.

We start by proving part 2 of the Proposition. It is convenient to define $\mathbb{1}_A$ to be 1 if agents are act-utilitiarian, and to be 0 otherwise.

After the senior agent's invocation decision $I \in \{0, 1\}$, let h_j^I be the junior agent's posterior assessment of the probability that $\mathscr{P}$ is harmful. Then his expected utility from not invoking $\mathscr{P}$ is

$$u_j^1 \triangleq (1 - \varepsilon_j)(pw^j + (1-p)w_j) + \varepsilon_j \mathbb{1}_A(\underline{\pi} + p\Delta_\pi + \bar{v}),$$

and from doing so is

$$u_j^1 \triangleq (1 - \varepsilon_j) \left((p + \Delta_p)w^j + (1 - p - \Delta_p)w_j\right) + \varepsilon_j \left(\mathbb{1}_A(\underline{\pi} + (p + \Delta_p)\Delta_\pi + \bar{v}) - h_j^I c\right).$$

Hence, the junior agent invokes $\mathscr{P}$ precisely when $u_j^1 \geq u_j^0$; this requirement reduces to Conditions (8) and (9).

We now prove the first part of the Proposition. The senior agent derives the following utility from not invoking $\mathscr{P}$:

$$\begin{aligned}
u_s^0 \triangleq (1 - \varepsilon_s) \Big[&\left(1 - I_j^0\right) \left(w^{sj}p^2 + (w_j^s + w_s^j)p(1-p) + w_{sj}(1-p)^2\right) \\
&+ I_j^0 \left(w^{sj}p(p + \Delta_p) + w_j^s p(1 - p - \Delta_p)\right. \\
&\qquad \left. + w_s^j(1-p)(p + \Delta_p) + w_{sj}(1-p)(1 - p - \Delta_p)\right)\Big] \\
&+ \varepsilon_s \mathbb{1}_A \left(\underline{\pi} + p\Delta_\pi + \bar{v}\right) \\
= (1 - \varepsilon_s) \Big[&w^{sj}p^2 + (w_j^s + w_s^j)p(1-p) + w_{sj}(1-p)^2 \\
&+ I_j^0 \Delta_p \left((w^{sj} - w_j^s)p + (w_s^j - w_{sj})(1-p)\right)\Big] \\
&+ \varepsilon_s \mathbb{1}_A \left(\underline{\pi} + p\Delta_\pi + \bar{v}\right),
\end{aligned}$$

where I_j^1 is 1 if $h_j^1 \leq h_j^*$ and 0 otherwise, and I_j^0 is defined similarly. The senior agent derives the following utility from invoking $\mathscr{P}$:

$$\begin{aligned}
u_s^1 \triangleq (1-\varepsilon_s)\Big[&\big(1-I_j^1\big)\big(w^{sj}(p+\Delta_p)p + w_j^s(p+\Delta_p)(1-p) \\
&\quad + w_s^j(1-p-\Delta_p)p + w_{sj}(1-p-\Delta_p)(1-p)\big) \\
&\quad + I_j^1\big(w^{sj}(p+\Delta_p)^2 + (w_j^s+w_s^j)(p+\Delta_p)(1-p-\Delta_p) \\
&\qquad\qquad + w_{sj}(1-p-\Delta_p)^2\big)\Big] \\
&+ \varepsilon_s\left(\mathbb{1}_A(\underline{\pi} + (p+\Delta_p)\Delta_\pi + \bar{v}) - \eta(\sigma)c\right) \\
= (1-\varepsilon_s)\Big[&w^{sj}p(p+\Delta_p) + w_j^s(p+\Delta_p)(1-p) \\
&\qquad + w_s^j(1-p-\Delta_p)p + w_{sj}(1-p-\Delta_p)(1-p) \\
&+ I_j^1\Delta_p\big((w^{sj}-w_j^s)(p+\Delta_p) + (w_s^j - w_{sj})(1-p-\Delta_p)\big)\Big] \\
&+ \varepsilon_s\left(\mathbb{1}_A(\underline{\pi} + (p+\Delta_p)\Delta_\pi + \bar{v}) - \eta(\sigma)c\right)
\end{aligned}$$

The senior agent will choose to invoke $\mathscr{P}$ precisely when $u_s^1 - u_s^0 \geq 0$. This requirement reduces to Condition (22):

$$\begin{aligned}
&\Delta_p\Big[w^{sj}\left(p + I_j^1\Delta_p + p(I_j^1 - I_j^0)\right) + w_j^s\left(1-p-I_j^1\Delta_p - p(I_j^1-I_j^0)\right) \\
&\qquad + w_s^j\left(-p - I_j^1\Delta_p + (1-p)(I_j^1-I_j^0)\right) + w_{sj}\left(-1+p+I_j^1\Delta_p - (1-p)(I_j^1-I_j^0)\right)\Big] \\
&\geq \frac{\varepsilon_s}{1-\varepsilon_s}\left(\eta(\sigma)c - \mathbb{1}_A\Delta_p\Delta_\pi\right)
\end{aligned} \tag{22}$$

It follows immediately that $\frac{\partial}{\partial\sigma}\left(u_s^1 - u_s^0\right) = -\varepsilon_s c\eta'(\sigma)$. Hence, because $\eta(\sigma)$ is increasing in σ, the benefit to the senior agent of invoking $\mathscr{P}$ is decreasing in σ. The result follows. Note that σ^*_{Act} is guaranteed to lie between 0 and 1 for high enough ε_s. This follows immediately from the observation that $[\Delta_p\Delta_\pi - \eta(\sigma)c]$ is positive when $\sigma = 0$ and, by Assumption (1), negative when $\sigma = 1$.

Proof of Lemma 2.

The following expressions are an immediate consequence of Bayes' Law:

$$h_j^1 = \frac{1}{1 + \frac{1-h}{h}\frac{F_L(\sigma^*)}{F_H(\sigma^*)}}; \qquad h_j^0 = \frac{1}{1 + \frac{1-h}{h}\frac{1-F_L(\sigma^*)}{1-F_H(\sigma^*)}}. \tag{23}$$

Then $\eta(\sigma^*) \geq h_j^1$ if and only if

$$\frac{hF_H(\sigma^*)}{hF_H(\sigma^*) + (1-h)F_L(\sigma^*)} \leq \frac{hf_H(\sigma^*)}{hf_H(\sigma^*) + (1-h)f_L(\sigma^*)}.$$

This expression reduces to $F_H(\sigma^*)/F_L(\sigma^*) \leq f_H(\sigma^*)/f_L(\sigma^*)$, which is true by Lemma 9. That $\eta(\sigma^*) \leq h_j^0$ follows similarly.

We can now prove Lemma 10, which establishes the properties of the junior agent's strategic regions that are illustrated in Figure 1:

Lemma 10.

1. *$h_j^0(\sigma^*)$ is increasing in σ^* with $h_j^0(0) = h$ and $h_j^0(1) = 1$;*
2. *$h_j^1(\sigma^*)$ is increasing in σ^* with $h_j^1(0) = 0$ and $h_j^1(1) = h$;*
3. *$\eta(\sigma^*)$ increases from $\eta(0) = 0$ to $\eta(1) = 1$.*

Proof. All of the results of Lemma follow from Equations (21), (23), and the assumption that $f_L(1) = 0 = f_H(0)$.

Proof of Lemma 4.

Assume first that $\eta(\sigma^*) - \underline{h}_R > 0$, as in part 1 of the Lemma. Any σ^* satisfies Condition (22) with equality. It is convenient to establish the following general expression for the principal's expected senior agent wage bill:

$$\begin{aligned}
\mathbb{E}[w] = F_h(\sigma^*) &\left[w^{sj}(p+\Delta_p)(p+I_j^1\Delta_p) + w_j^s(p+\Delta_p)(1-p-I_j^1\Delta_p)\right.\\
&\left.\quad + w_s^j(1-p-\Delta_p)(p+I_j^1\Delta_p) + w_{sj}(1-p-\Delta_p)(1-p-I_j^1\Delta_p)\right]\\
&+ (1-F_h(\sigma^*))\left[w^{sj}p(p+I_j^0\Delta_p) + w_j^s p(1-p-I_j^0\Delta_p)\right.\\
&\left.\qquad + w_s^j(1-p)(p+I_j^0\Delta_p) + w_{sj}(1-p)(1-p-I_j^0\Delta_p)\right]\\
= w^{sj}&p(p+I_j^0\Delta_p) + w_j^s p(1-p-I_j^0\Delta_p) + w_s^j(1-p)(p+I_j^0\Delta_p) + w_{sj}(1-p)(1-p-I_j^0\Delta_p)\\
&+ F_h(\sigma^*)\Delta_p\left[w^{sj}\left(p+I_j^1\Delta_p + p(I_j^1-I_j^0)\right) + w_j^s\left(1-p-I_j^1\Delta_p - p(I_j^1-I_j^0)\right)\right.\\
&\left.\qquad + w_s^j\left(-p-I_j^1\Delta_p + (1-p)(I_j^1-I_j^0)\right) + w_{sj}\left(-1+p+I_j^1\Delta_p - (1-p)(I_j^1-I_j^0)\right)\right],
\end{aligned} \tag{24}$$

The term that is multiplied by $F_h(\sigma^*)$ in Equation (24) corresponds to the case where $\mathscr{P}$ is invoked. In this case, Condition (22) holds with equality, so that we can re-write the expected wage as

$$\begin{aligned}
\mathbb{E}[w] = w^{sj}p(p+I_j^0\Delta_p) + w_j^s p(1-p-I_j^0\Delta_p) + w_s^j(1-p)(p+I_j^0\Delta_p)\\
+ w_{sj}(1-p)(1-p-I_j^0\Delta_p) + \frac{\varepsilon_s}{1-\varepsilon_s}c\left(\eta(\sigma^*) - \underline{h}_R\right)F_h(\sigma^*)
\end{aligned} \tag{25}$$

For part 1(a), note that when $h_j^1 \le h_j^* < h_j^0$ we have $I_j^1 = 1$ and $I_j^0 = 0$, so that the principal selects $(w^{sj}, w_j^s, w_s^j, w_{sj})$ so as to minimize the following expected wage bill, subject to the constraints $w^{sj} \ge w_s^j$, $w_j^s \ge w_{sj}$, $w_s^j \ge 0$, and $w_{sj} \ge 0$:

$$\mathbb{E}[w] = \left(w^{sj}p^2 + (w_j^s + w_s^j)p(1-p) + w_{sj}(1-p)^2\right) + \frac{\varepsilon_s}{1-\varepsilon_s}c\left(\eta(\sigma^*) - \underline{h}_R\right)F_h(\sigma^*). \tag{26}$$

This problem yields the following Lagrangian:

$$\begin{aligned}\mathscr{L} = &-\left(w^{sj}p^2 + (w_j^s + w_s^j)p(1-p) + w_{sj}(1-p)^2\right) - \frac{\varepsilon_s}{1-\varepsilon_s}c\left(\eta(\sigma^*) - \underline{h}_R\right)F_h(\sigma^*) \\ &+ \lambda\left[(1-\varepsilon_s)\Delta_p\left(w^{sj}(2p+\Delta_p) + (w_j^s + w_s^j)(1-2p-\Delta_p) - w_{sj}(2-2p-\Delta_p)\right)\right. \\ &\left. + \varepsilon_s(\underline{h}_R - \eta(\sigma^*))\right] + \mu^{sj}(w^{sj} - w_s^j) + \mu_j^s(w_j^s - w_{sj}) + \mu_s^j w_s^j + \mu_{sj}w_{sj}. \end{aligned} \tag{27}$$

The first order conditions for the principal's problem are as follows:

$$\frac{\partial \mathscr{L}}{\partial w^{sj}} = -p^2 + \lambda(1-\varepsilon_s)\Delta_p(2p+\Delta_p) + \mu^{sj} = 0; \tag{28}$$
$$\frac{\partial \mathscr{L}}{\partial w_j^s} = -p(1-p) + \lambda(1-\varepsilon_s)\Delta_p(1-2p-\Delta_p) + \mu_j^s = 0; \tag{29}$$
$$\frac{\partial \mathscr{L}}{\partial w_s^j} = -p(1-p) + \lambda(1-\varepsilon_s)\Delta_p(1-2p-\Delta_p) - \mu^{sj} + \mu_s^j = 0; \tag{30}$$
$$\frac{\partial \mathscr{L}}{\partial w_{sj}} = -(1-p)^2 - \lambda(1-\varepsilon_s)\Delta_p(2-2p-\Delta_p) - \mu_j^s + \mu_{sj} = 0. \tag{31}$$

Suppose that $\mu_{sj} = 0$. Then Equation (31) implies that $\mu_j^s < 0$, which is impossible. Hence $\mu_{sj} > 0$ and, by complementary slackness, $w_{sj} = 0$.

Now subtract Equation (30) from Equation (29) to get $\mu_j^s + \mu^{sj} = \mu_s^j$. If $\mu^{sj} = 0$ then setting $\mu_j^s = \mu_s^j = 0$ generates a contradiction from Equations (28) and (29) (except in a knife-edge case). So if $w^{sj} > 0$ we must have $w_j^s = w_s^j = 0$.

Similarly, if $\mu_j^s = 0$ we must have $\mu^{sj} = \mu_s^j > 0$, so that, if $w_j^s > 0$, we must have $w^{sj} = w_s^j = 0$.

If $w^{sj} > 0$ then it follows from Equation (22) (which holds with equality) that w^{sj} is given by Equation (13), and from Equation (26) that the expected wage payment to the senior agent is given by expression (32):

$$\frac{\varepsilon_s}{1-\varepsilon_s}\frac{c}{\Delta_p}\left(\eta(\sigma^*) - \underline{h}_R\right)\left(\frac{p^2}{2p+\Delta_p} + \Delta_p F_h(\sigma^*)\right). \tag{32}$$

Similarly, if $w_j^s > 0$ then w_j^s is given by Equation (16) and the expected wage payment to the senior agent is as follows:

$$\frac{\varepsilon_s}{1-\varepsilon_s}\frac{c}{\Delta_p}\left(\eta(\sigma^*) - \underline{h}_R\right)\left(\frac{p(1-p)}{1-2p-\Delta_p} + \Delta_p F_h(\sigma^*)\right). \tag{33}$$

The expected wage payment is lower when $w^{sj} > 0$ if and only if expression (32) is less than expression (33). For $1-2p-\Delta_p > 0$ this requirement reduces to $p > 0$. But if $1-2p-\Delta_p < 0$, $w_j^s < 0$ (since, by assumption, $\eta(\sigma^*) - \underline{h}_R) > 0$), which is impossible: it follows that, even in this case, the principal sets $w^{sj} > 0$, as in the statement of the Lemma.

For part 1(b) of the Lemma, note that when $h_j^1 < h_j^0 < h_j^*$ we have $I_j^1 = I_j^0 = 1$, so that

Equation (22) yields

$$(p+\Delta_p)(w^{sj}-w_s^j)+(1-p-\Delta_p)(w_j^s-w_{sj})=\frac{\varepsilon_s}{1-\varepsilon_s}\frac{c}{\Delta_p}\left(\eta(\sigma^*)-\underline{h}_R\right).$$

This expression and Equation (10) together imply that $w_s^j=w_{sj}=0$ and that w^{sj}, w_j^s satisfy Equation (14).

For part 1(c) of the Lemma, note that in the rejection region we have $h_j^*<h_j^1<h_j^0$ and so $I_j^1=I_j^0=0$. Equation (22) therefore yields the requirement that

$$p\left(w^{sj}-w_s^j\right)+(1-p)\left(w_j^s-w_{sj}\right)=\frac{\varepsilon_s}{1-\varepsilon_s}\frac{c}{\Delta_p}\left(\eta\left(\sigma^*\right)-\underline{h}_R\right)$$

As this requirement can remain satisfied by lowering w_j^s and w_{sj} by equal amounts, Equation (10) implies $w_{sj}=0$; similarly, $w_s^j=0$. Hence we have (15).

Part 2 of the Lemma corresponds to the case where $h_j^1\le h_j^*<h_j^0$ and $1-2p-\Delta_p<0$ so that, as above, the expected wage bill is cheaper with $w_j^s>0$ than with $w^{sj}>0$. When $\eta(\sigma^*)<\hat{h}$, Equation (22) yields a positive value for w_j^s, so that this payment is feasible.

Proof of Propositions 2 and 3

We write $R\in\{Act, Duty\}$ for the agents' ethical standard. The principal sets the customer's fees via a take-it-or-leave-it offer. The principal's profits are therefore $\Pi_{S,R}\triangleq\mathscr{S}-\mathscr{W}_R$, where $\mathscr{S}$ and $\mathscr{W}_R$ are the equilibrium expected surplus and the expected wage bill, respectively and the S subscript indicates that customers are sophisticated. We proceed by computing the optimal contract conditional upon cultural assimilation, acceptance, and rejection; we then compare the principal's expected profit from each contract. Our analysis generates a number of interim Lemmas that we use to prove Propositions 2 and 3.

Cultural Assimilation Region.—In the cultural assimilation region, we have

$$\mathscr{S}^{\text{Ass}}=2(\underline{\pi}+\bar{v}+(p+\Delta_pF_h(\sigma^*))\Delta_\pi-hcF_H(\sigma^*)). \tag{34}$$

The consumer anticipates that the practice is harmful with probability h, in which case each agent will invoke the practice with probability $F_H\left(\sigma^*\right)$. The consumer is therefore willing to pay at most $\bar{v}-hcF_H\left(\sigma^*\right)$ for the service, as in Equation (34). Note that using $\eta(\sigma)f_h(\sigma)=hf_H(\sigma)$ we have $\frac{\partial\mathscr{S}^{\text{Ass}}}{\partial\sigma^*}=2f_h(\sigma^*)c\left(\hat{h}-\eta(\sigma^*)\right)$, which is positive for $\sigma^*<\hat{\sigma}$, zero when $\sigma^*=\hat{\sigma}$, and negative for $\sigma^*>\hat{\sigma}$.

Wages depend upon the ethical standard, R. By Lemma 3, act-utilitarians must have $h_{j,Act}^*\ge\underline{h}_{Act}=\hat{h}$; the principal can therefore locate the junior agent in the Cultural Assimilation Region when $R=Act$ by setting $h_{j,Act}^*=\hat{h}$ and paying a state-independent wage of zero. In contrast, when agents are other-regarding, any $h_{J,Duty}^*\ge 0$ can be implemented, and, by Equation (11), the associated wage $\underline{w}^j_{Duty}(h^*_{j,Duty})$ is increasing in $h^*_{j,Duty}$. It follows from inspection of Figure 1 that the cheapest way to locate an other-regarding junior agent in the Cultural Assimilation Region is to set $h^*_{j,Duty} = h^1_j(\sigma^*)$. This reasoning yields the following expression for expected junior agent wage:

$$\mathscr{W}^{\text{Ass}}_{j,R} = \begin{cases} 0, & R = Act; \\ \frac{\varepsilon_j}{1-\varepsilon_j}\frac{c}{\Delta_p} h^1_j(\sigma^*)(p + \Delta_p F_h(\sigma^*)), & R = Duty. \end{cases} \tag{35}$$

Equations (32) and (33) then yield the following expression for the senior agent's expected wage:

$$\mathscr{W}^{\text{Ass}}_{s,R} = \begin{cases} \frac{\varepsilon_s}{1-\varepsilon_s}\frac{c(\eta(\sigma^*)-\underline{h}_R)}{\Delta_p}\left(\frac{p^2}{2p+\Delta_p} + \Delta_p F_h(\sigma^*)\right), & R = Duty \text{ or } R = Act \text{ and } \sigma^* \geq \hat{\sigma}; \\ \frac{\varepsilon_s}{1-\varepsilon_s}\frac{c\left(\eta(\sigma^*)-\hat{h}\right)}{\Delta_p}\left(\frac{p^2(1-p)}{1-2p-\Delta_p} + \Delta_p F_h(\sigma^*)\right), & R = Act,\ \sigma^* < \hat{\sigma} \text{ and } 1-2p-\Delta_p < 0. \end{cases} \tag{36}$$

Lemma 11. *Suppose that agents are act-utilitarian and that cultural assimilation is to be implemented. Then the principal optimally sets $\sigma^*_{Act} = \hat{\sigma}$ and pays both agents a state-independent wage of* 0.

Proof. $\mathscr{W}^{\text{Ass}}_{s,Act}$ is increasing in σ^* for $\sigma^* > \hat{\sigma}$ and decreasing for $\sigma^* < \hat{\sigma}$. Since $\mathscr{S}^{\text{Ass}}$ is maximised at $\sigma = \sigma^*$, it follows that, conditional upon opting for the cultural assimilation region, the principal optimally sets $\sigma^*_{Act} = \hat{\sigma}$; Equations (35) and (36) then imply that both agents receive a state-independent wage of zero.

Lemma 12. *Suppose that agents are other-regarding and that cultural assimilation is to be implemented. Then the principal optimally sets $\sigma^*_{Duty} < \hat{\sigma}$ and*

$$\eta(\sigma^*_{Duty}) < \frac{\hat{h}}{1 + \frac{1}{2}\frac{\varepsilon_s}{1-\varepsilon_s}}. \tag{37}$$

Proof. Since $\mathscr{W}^{\text{Ass}}_{Duty}$ is increasing in σ^* and $\mathscr{S}^{\text{Ass}}$ is maximised at $\hat{\sigma}$, we must have $\sigma^*_{Duty} < \hat{\sigma}$. If ε_s and ε_j are not too high then it is cost-effective for the principal to set $\sigma^*_{Duty} > 0$, so that the senior agent suppresses his moral concerns for some signals σ. When this is the case, the first order condition for the principal's problem is

$$\begin{aligned} &2f_h(\sigma^*)\left(\hat{h} - \eta(\sigma^*)\left(1 + \frac{1}{2}\frac{\varepsilon_s}{1-\varepsilon_s} + \frac{1}{2}\frac{\varepsilon_j}{1-\varepsilon_j}\frac{h^1_j(\sigma^*)}{\eta(\sigma^*)}\right)\right) \\ &\quad = \frac{\varepsilon_s}{1-\varepsilon_s}\frac{1}{\Delta_p}\left(\frac{p^2}{2p+\Delta_p} + \Delta_p F_h(\sigma^*)\right)\frac{\partial\eta}{\partial\sigma}(\sigma^*) + \frac{\varepsilon_j}{1-\varepsilon_j}\frac{1}{\Delta_p}(p + \Delta_p F_h(\sigma^*))\frac{\partial h^1_j}{\partial\sigma}(\sigma^*). \end{aligned} \tag{38}$$

The right hand side of Equation (38) is positive and $h_j^1(\sigma^*) < \eta(\sigma^*)$, so that

$$\hat{h} > \eta(\sigma^*)\left(1 + \frac{1}{2}\left(\frac{\varepsilon_s}{1-\varepsilon_s} + \frac{\varepsilon_j}{1-\varepsilon_j}\frac{h_j^1(\sigma^*)}{\eta(\sigma^*)}\right)\right) > \eta(\sigma^*)\left(1 + \frac{1}{2}\frac{\varepsilon_s}{1-\varepsilon_s}\right),$$

from which Equation (37) follows.

Acceptance Region.—Using Figure 1 and Lemma 3, the cheapest way to locate the junior agent in the Acceptance Region is to set $h_j^* = h_j^0$, so that the expected junior agent's wage bill is

$$\mathscr{W}_j^{\text{Acc}} = \frac{\varepsilon_j}{1-\varepsilon_j}\frac{c}{\Delta_p}(h_j^0(\sigma^*) - \underline{h}_R)\,(p+\Delta_p)\,; \tag{39}$$

it is clear from Figure 1 that $\mathscr{W}_j^{\text{Acc}}$ is increasing in σ^* for $\sigma^* \geq \hat{\sigma}$.

Using Equation (14), the senior agent's optimal compensation contract can be implemented by setting

$$w^{sj} = \frac{\varepsilon_s}{1-\varepsilon_s}\frac{c}{\Delta_p}\frac{\eta(\sigma^*) - \underline{h}_R}{p+\Delta_p}; \qquad w_j^s = 0$$

so that, substituting these values and $w_{sj} = 0$ into Equation (24) with $I_j^0 = I_j^1 = 1$, the expected wage payment to the senior agent is

$$\mathscr{W}_s^{\text{Acc}} = \frac{\varepsilon_s}{1-\varepsilon_s}\frac{c}{\Delta_p}\left(\eta(\sigma^*) - \underline{h}_R\right)(p + \Delta_p F_h(\sigma^*))\,.$$

Once again, $\mathscr{W}_s^{\text{Acc}}$ is trivially increasing in σ^* for $\sigma^* \geq \hat{\sigma}$.

The expected surplus generated when the junior agent's trigger value is in the Acceptance Region is

$$\mathscr{S}^{\text{Acc}} = 2(\underline{\pi} + \bar{v} + p\Delta_\pi) + \hat{h}c(1 + F_h(\sigma^*)) - hc(1 + F_H(\sigma^*)),$$

from which it follows that $\frac{\partial \mathscr{S}^{\text{Acc}}}{\partial \sigma^*} = cf_h(\sigma^*)(\hat{h} - \eta(\sigma^*))$, which is negative for $\sigma^* > \hat{\sigma}$, positive for $\sigma^* < \hat{\sigma}$, and zero when $\sigma^* = \hat{\sigma}$.

Lemma 13. *Suppose that agents are act-utilitarians and that the acceptance region is to be implemented. Then the principal implements $\sigma^* = \hat{\sigma}$, sets the senior agent's wage to zero and pays the junior agent a success bonus $w^j = \underline{w}^j_{Act}(h_j^0)$.*

Proof. The principal's profit is equal to the expected surplus net of the expected wages: $\Pi_S^{\text{Acc}} \triangleq \mathscr{S}^{\text{Acc}} - \mathscr{W}_s^{\text{Acc}} - \mathscr{W}_j^{\text{Acc}}$. We have shown that this is decreasing in σ^* for $\sigma^* \geq \hat{\sigma}$, so that the principal implements $\sigma^* = \hat{\sigma}$. Equation (14) then implies that the senior agent's wage is zero; the junior agent bonus is derived before Equation (39).

Lemma 14. *Suppose that agents are other-regarding and that the acceptance region is to be* *implemented. Then the principal optimally sets* $\sigma^*_{Duty} < \hat{\sigma}$ *and*

$$\eta(\sigma^*_{Duty}) < \frac{\hat{h}}{1 + \frac{\varepsilon_s}{1-\varepsilon_s}}. \tag{40}$$

Proof. Since $\mathscr{S}^{\text{Acc}}$ attains a maximum at $\sigma^* = \hat{\sigma}$ and $\mathscr{W}^{\text{Acc}}_{j,Duty} + \mathscr{W}^{\text{Acc}}_{s,Duty}$ is increasing in σ^* we must have $\sigma^*_{Duty} < \hat{\sigma}$. The principal's first order condition for σ^* is

$$f_h(\sigma^*)\left(\hat{h} - \eta(\sigma^*)\left(1 + \frac{\varepsilon_s}{1-\varepsilon_s}\right)\right) = \frac{\varepsilon_s}{1-\varepsilon_s}\frac{1}{\Delta_p}(p + \Delta_p F_h(\sigma^*))\frac{\partial \eta}{\partial \sigma}(\sigma^*) + \frac{\varepsilon_j}{1-\varepsilon_j}\frac{1}{\Delta_p}(p+\Delta_p)\frac{\partial h^0_j}{\partial \sigma}(\sigma^*);$$

because the right hand side of this expression is positive, Equation (40) follows.

Rejection Region.—It is clear from Figure 1 that the cheapest way to place the junior in the Rejection Region is to set $h^*_j = \underline{h}_R < h^1_j$ and to pay a state-independent wage of 0. This is possible for act-utilitarian agents only if $h > \hat{h}$ and $\sigma^* > \left(h^1_j\right)^{-1}\left(\hat{h}\right) \geq \hat{\sigma}$.

The senior agent's optimal compensation contract can be implemented by setting

$$w^{sj} = \frac{\varepsilon_s}{1-\varepsilon_s}\frac{c}{\Delta_p}\frac{\eta(\sigma^*) - \underline{h}_R}{p},$$

so that, adapting Equation (26) to junior rejection, the expected wage payment to the senior agent is

$$\mathscr{W}^{\text{Rej}}_{s,R} = \frac{\varepsilon_s}{1-\varepsilon_s}\frac{c}{\Delta_p}\left(\eta(\sigma^*) - \underline{h}_R\right)\left(p + \Delta_p F_h(\sigma^*)\right).$$

$\mathscr{W}^{\text{Rej}}_{s,Act}$ is increasing in σ^* for $\sigma^* \geq \hat{\sigma}$, and $\mathscr{W}^{\text{Rej}}_{Duty}$ is increasing in σ^*.

The expected surplus when the junior agent's trigger value is in the Rejection Region is

$$\mathscr{S}^{\text{Rej}} = 2(\underline{\pi} + \bar{v} + p\Delta_\pi) + \hat{h}cF_h(\sigma^*) - hcF_H(\sigma^*),$$

from which it follows that

$$\frac{\partial \mathscr{S}^{\text{Rej}}}{\partial \sigma^*} = cf_h(\sigma^*)(\hat{h} - \eta(\sigma^*)). \tag{41}$$

Lemma 15. *Rejection can never be optimal when agents are act-utilitarian.*

Proof. For act-utilitarian agents, $\sigma^* \geq \hat{\sigma}$. Hence $\mathscr{S}^{\text{Rej}} - \mathscr{W}^{\text{Rej}}_{s,Act}$ is declining in σ^*. The principal therefore optimally lowers σ^* out of the Rejection Region.

Lemma 16. *Suppose that agents are other-regarding and that the rejection region is to be implemented. Then the principal optimally sets* $\sigma^*_{Duty} < \hat{\sigma}$ *and Condition* (40) *is satisfied.*

Proof. It follows from Equation (41) that $\mathscr{S}^{\text{Rej}}$ attains a local maximum at $\hat{\sigma}$. Since $\mathscr{W}_{s,i}^{\text{Rej}}$ s increasing in σ^*, we must have $\sigma^*_{Duty} < \hat{\sigma}$. When $\sigma^*_{Duty} > 0$, the principal's first order condition is

$$f_h(\sigma^*)\left(\hat{h} - \eta(\sigma^*)\left(1 + \frac{\varepsilon_s}{1-\varepsilon_s}\right)\right) = \frac{\varepsilon_s}{1-\varepsilon_s}\left(\frac{p}{\Delta_p} + F_h(\sigma^*)\right)\frac{\partial \eta}{\partial \sigma}(\sigma^*).$$

The right hand side of this expression is positive, and Condition (40) follows.

Proof of Proposition 2.—Lemma 15 implies that rejection can never be optimal with act-utilitarian agents. Lemma 11 implies that the principal extracts all of the expected surplus from $\mathscr{P}$ in the Cultural Assimilation region when $R = Act$ and, hence, has expected income $\mathscr{S}^{\text{Ass}}(\sigma^* = \hat{\sigma})$. By Lemma 13, the principal has expected income $\mathscr{S}^{\text{Acc}}(\sigma^* = \hat{\sigma}) - \mathscr{W}^{\text{Ass}}_{j,Act}(\sigma^* = \hat{\sigma})$. But

$$\begin{aligned} \mathscr{S}^{\text{Ass}} - \mathscr{S}^{\text{Acc}} &= c\left[h(1 - F_H(\hat{\sigma})) - \hat{h}(1 - F_h(\hat{\sigma}))\right] \\ &= ch(1-h)\frac{f_H(\hat{\sigma})(1 - F_H(\hat{\sigma}))}{f_h(\hat{\sigma})}\left[\frac{f_L(\hat{\sigma})}{f_H(\hat{\sigma})} - \frac{1 - F_L(\hat{\sigma})}{1 - F_H(\hat{\sigma})}\right], \end{aligned} \tag{42}$$

which is positive by Lemma 9. Proposition 2 follows immediately.

Proof of Proposition 3.—Lemmas 12, 14 and 16 together prove that, irrespective of whether junior agents accept, reject, or culturally assimilate, $\sigma^*_{Duty} < \hat{\sigma}$. The same Lemmas give the limit on $\eta(\sigma^*_{Duty})$ identified in the Proposition. That acceptance, rejection, and cultural assimilation are all possible is demonstrated with examples in the text following the statement of the Proposition.

Proof of Lemmas 5 and 6.

The Lemmas are a consequence of the expressions for $\Pi^{\text{Ass}}_{S,Act}$, $\Pi^{\text{Ass}}_{S,Duty}$, $\Pi^{\text{Acc}}_{S,Duty}$, and $\Pi^{\text{Rej}}_{S,Duty}$ and the Envelope Theorem.

Proof of Proposition 4

Because naïve customers do not charge for the harm caused by $\mathscr{P}$, the principal's profit in the Cultural Assimilation and Acceptance regions, respectively, is $\Pi^{\text{Ass}}_{N,Act} \triangleq \Pi^{\text{Ass}}_{S,Act} + 2hcF_H(\sigma^*)$ and $\Pi^{\text{Acc}}_{N,Act} \triangleq \Pi^{\text{Acc}}_{S,Act} + hc(1 + F_H(\sigma^*))$. First we confirm that incentivising the senior agent not to invoke surplus-enhancing practices cannot be optimal. By proposition 2, $\partial \Pi^{\text{Ass}}_{S,Act}/\partial \sigma^*\big|_{\sigma^* < \hat{\sigma}} > 0$, and so $\partial \Pi^{\text{Ass}}_{N,Act}/\partial \sigma^*\big|_{\sigma^* < \hat{\sigma}} > 0$. The case of $\partial \Pi^{\text{Acc}}_{N,Act}/\partial \sigma^*$ is analogous. Hence $\sigma^*_{Act} \geq \hat{\sigma}$.

It is then easy to prove that

$$\left.\frac{\partial \Pi^{\text{Ass}}_{N,Act}}{\partial \sigma^*}\right|_{\sigma^*=\hat{\sigma}} = 2c\hat{h}f_h(\sigma^*) - \mathscr{E}_s\frac{c}{\Delta_p}\eta'(\hat{\sigma})\left(\frac{p^2}{2p+\Delta_p} + \Delta_p F_h(\sigma^*)\right), \tag{43}$$

which is greater than 0 precisely when $\mathscr{E}_s$ is limited by the ratio of the first term in Equation (43) to its coefficient in that Equation. This proves part 1 of the Proposition.

Similarly, we have

$$\left.\frac{\partial \Pi^{\text{Acc}}_{N,Act}}{\partial \sigma^*}\right|_{\sigma^*=\hat{\sigma}} = \hat{h}cf_h(\hat{\sigma}) - \mathscr{E}_s\frac{c}{\Delta_p}\eta'(\hat{\sigma})\left(p + \Delta_p F_h(\sigma^*)\right) - \mathscr{E}_j\frac{c}{\Delta_p}\frac{\partial h^0_j}{\partial \sigma^*}(\hat{\sigma})\left(p + \Delta_p\right), \tag{44}$$

whence part 2 of the Proposition follows with $(1/\kappa_s^{Acc}, 1/\kappa_j^{Acc})$ equal to the respective coefficients of $\mathscr{E}_s$ and $\mathscr{E}_j$ in Equation (44), divided by $\hat{h}cf_h(\hat{\sigma})$.

Note that $\eta(\hat{\sigma}) > h^1_j(\hat{\sigma})$ so that we have to have $\sigma^*_{Act} > \hat{\sigma}$ in the Rejection Region.

The final part of the Proposition follows by using continuity at $\mathscr{E}_s = \mathscr{E}_j = 0$ and noting that

$$\Pi^{\text{Acc}}_{N,Act} - \Pi^{\text{Ass}}_{N,Act}\Big|_{\mathscr{E}_s=\mathscr{E}_j=0} = \hat{h}c(1 - F_h(\sigma^*)) > 0$$
$$\text{and } \left.\Pi^{\text{Acc}}_{N,Act} - \Pi^{\text{Rej}}_{N,Act}\right|_{\mathscr{E}_s=\mathscr{E}_j=0} = \hat{h}c > 0$$

Hence incentivising junior acceptance of $\mathscr{P}$ dominates rejection of $\mathscr{P}$ if the agents' ethical commitment is weak enough.

Proof of Lemma 7 and 8

The result is an immediate consequence of the expressions for Π^{Ass}_N, Π^{Acc}_N, Π^{Rej}_N, and the Envelope Theorem.