Ideological Bias and Breach of the Fiduciary Duty of Care

Anthony M. Marino *

Marshall School of Business

University of Southern California

Los Angeles, CA 90089-1422

E-mail: amarino@usc.edu

August 16, 2025

Abstract

This paper analyzes a remedy for breach of the fiduciary duty of care in a principal-advisor relationship. The advisor's effort cost is ideologically too high for a certain investment. This bias can lead to a breach of care, if the advisor's investigative effort is unobservable by the principal. The principal audits and sues the agent, if there is a signal indicating a breach. We show that this process does not lead to efficiency. To correct this problem we offer a mechanism which can implement efficiency. The (active) court precommits to a function whose image value scales expectation damages and depends on the degree of bias of the agent and a minimum audit probability. The damage scale factor may inflate or deflate expectation damages. We consider cases where the principal is risk neutral and risk averse.

JEL Code: K13, K15

Key Words:

Breach, Fiduciary Duty

*I thank.my colleagues Joao Ramos and Odilon Camara their insightful comments. I also thank the Marshall School of Business for generous research support.

1. Introduction

Many individuals rely on the advice of experts. Examples include financial advisors, pension fund managers, lawyers, trustees, some property managers, some real estate agents, court appointed fiduciaries, executors and guardians. Those experts must exert effort to acquire information which in turn allows them to impart advice. If an advisor finds some investments, projects, or tasks distasteful in the sense that personal effort cost in acquiring necessary information is privately too high, then the advisor may not investigate such decisions properly. Such investments or tasks would be marginalized in the principal's portfolio and possible returns or welfare would be lost. The advisor's arbitrarily high effort cost for certain investments or tasks is a form of ideological bias which could lead to a breach of fiduciary duty to the principal. The remedy for such a breach is for the principal to take civil legal action against the advisor, conduct an audit of the advisor's behavior, and show that with a preponderance of evidence, there was a violation. This paper will analyze the efficiency properties of this audit process with a focus on individuals who are advised by a fiduciary with the above type of bias. We show that this mechanism cannot lead to efficiency, but we offer a process of collaboration by the court which relies on adjusted damages and leads to an efficient outcome.

By some estimates over 40 trillion dollars will be transferred between generations over the first half of the twenty-first century.[1] In recent times, most of this wealth is in the form of real estate and financial assets which require active management. Today's trustees and advisors in general have fairly broad powers and such powers along with hidden action and misalignment of incentives can lead to a breach of fiduciary duty. Breach can take the form of breach of loyalty (i.e., malfeasance) or breach of prudence or care (i.e., nonfeasance). As noted by Kelly (2022) courts in the past were reluctant to award other than compensatory damages, but more recently both federal and state

[1] See Havens and Schervish (2003) and Kelly (2022).

courts have ordered punitive damages for breaches of both loyalty and care in some cases.

A fiduciary must adopt the utility function of the principal and in making decisions carry out a duty of loyalty and a duty of care. We will concentrate on the latter duty as it applies to individuals being represented by a fiduciary advisor. Does the advisor exert the correct amount of effort in advising? If not, then there is a breach, but this type of breach is difficult to prove as it is hard for the principal to observe the effort of the advisor in investigating alternative decisions. It is a classic agency problem with hidden action and a possible divergence of incentives.

There is an early literature in the area of law and economics which discusses the remedies for breach of the duty of care in a fiduciary relationship. Cooter and Freedman (1991), Easterbrook and Fischel (1993), Sitkoff (2011) and Polinsky and Shavell (1998) discuss the basic economic efficiency argument in attempting to incentivize a fiduciary to exert due care. Kelly (2022) nicely summarizes this literature as it applies to trust law. The basic point is that if a duty of care is breached, then the principal must show with a preponderance of evidence that this breach occurred and that there were monetary damages. The compensatory damages are computed using the notion of expectation damages which states that the fiduciary must pay the principal an amount which is equal to the difference between the payoff that would be expected if due care had been exercised and the amount actually received. As noted by Kelly (2022), recently courts have also entertained punitive damages over and above these expectation damages. Punitive damages would be used to force the fiduciary to internalize the full amount of damages. This literature recommends that the court use a "total damage multiplier" applied to the expectation damages. See Polinski and Shavell (1998). The argument basically says that if D represents damages and ρ is the probability that the principal will uncover those damages, then efficiency will be obtained if the fiduciary is credibly assessed damages equal to $(1/\rho)D$. This causes the fiduciary to internalize the full damages $D = \rho(1/\rho)D$ as opposed to the lesser amount ρD. This argument is intuitive while being incomplete in that it treats the audit probability and the damages as exogenous. Each of these entities are functions of the agent's care level and what the principal expects the court's decision will be regarding damages. If the principal expects the court to multiply its expected damages by the reciprocal of its ρ, he could strategically lower ρ to inflate damages and at the same time pay less because ρ is costly. No equilibrium would result.

We formulate a simple model of a fiduciary advisor and a principal. The advisor exerts effort which is unobservable to the principal in order to gather information on two decisions to be made for the principal. On one of these decisions, the advisor is negatively biased, and such bias can cause a breach of duty of care absent the threat of legal action by the principal. If the principal suspects that the duty of care has been breached by a fiduciary, the principal will sue and audit the fiduciary. It is costly for principal to raise the probability that the audit will uncover the truth regarding the agent's information acquisition effort. We show that the audit mechanism cannot lead to an equilibrium in which information acquisition efforts are efficient, if the agent has bias with regard to a decision. We offer a mechanism involving what we call an active court which, when added to the audit, can lead to efficiency. Our process is in the tradition of the Polinksi and Shavell total damage multiplier in that it induces efficiency by the fiduciary. It is more explicit and complete, because it eliminates the ambiguity and lack of endogeneity present in the simple total damage multiplier argument. The active court precommits to a function whose image value δ scales expectation damages. This function represents a rule which depends on the degree of bias discovered after an audit and a parametric minimum audit probability (a burden of proof) to be implemented by the principal in order to release damages. The image value δ can deflate or inflate expectation damages depending on the degree of bias of the advisor, so that damages may be discounted or punitive. By precommiting to this function, the court sets a uniform rule to be faced by all plaintiffs. That is, the court does not optimally choose δ assuming that one particular plaintiff will optimize their audit probability given that δ, as would be the case where the court is a Stackelberg leader and where its choice of δ affects each individual plaintiff's best response audit probability. We present our results for the cases where the principal is risk neutral and risk averse. The biased advisor is assumed to be risk neutral.

Section 2 presents the model. Section 3 considers a risk neutral principal. Section 4 characterizes the case of a risk averse principal. Section 5 concludes.

2. The Model and the First Best

Let us consider an advisor making decisions on behalf of a principal. Suppose that there are different activities, tasks or investments that require a decision to either go ahead or to reject. Each of these activities has a possible payoff to the principal. For the purpose of streamlining discussion in what follows, let us refer to these entities as investments. We consider a very simple model where there are two possible investments. Investment i has a net return in the set $\{L_i, H_i\}$, $H_i > L_i, i = 1, 2$. The ex ante probability of H_i is $\pi_i \in (0, 1)$. Let $M_i = \pi_i H_i + (1 - \pi_i)L_i$, denote the ex ante expectation of cash flow for investment i. In the basic model, both the principal and the advisor are risk neutral, but we will consider a risk averse principal in Section 4. We assume

A.1 $H_i > 0$, $L_i, M_i < 0$.

Under A.1, both the advisor and the principal do not want to invest in investment i if the low cash flow is the actual state, and, likewise, they do not want to go forward if there is no information, M_i. The advisor can generate a noisy signal of investment magnitude through information acquisition effort e_i directed towards investigating investment i. With probability e_i the agent will learn that the return is H_i or L_i, and with probability $(1 - e_i)$ no information will be produced by the advisor's effort and expected return is the prior, M_i. If the advisor learns that the return is H_i after exerting information effort, then the advisor will invest. If the advisor learns L_i or gains no information so that expected return is M_i, then no investment will be made. We assume

A.2 $e_i \in (0,1), i = 1, 2.$

For an unbiased advisor, the cost of information acquisition effort on investment i is given by the function $c(e_i)$. We assume

A.3 $c'(e_i), c''(e_i) > 0$, $c'''(e_i) \geqq 0$, for $e_i > 0, c'(0) = 0$, and $c'(1) = \infty$.

Assumption A.3 invokes the Inada conditions for effort and states that the marginal and total costs of effort are positive and rising. If an advisor has an effort cost which deviates from $c(e_i)$, then that advisor will have a bias.

We want to describe an advisor implementing fiduciary duty in researching and then choosing investments. Consider the following problem:

$$\max_{\{e_1,e_2\}} \sum_{i=1}^{2} e_i \pi_i H_i - \sum_{i=1}^{2} c(e_i). \tag{FD}$$

Let $e \equiv (e_1, e_2)$ and $S(e) \equiv \sum_{i=1}^{2} e_i \pi_i H_i - \sum_{i=1}^{2} c(e_i)$. The objective function $S(e)$ consists of the expected value of assets managed less the unbiased advisor's effort cost of information acquisition. Because the fiduciary is the expert, we assume that the unbiased advisor's effort costs are less than those of the principal for each vector of effort provision. Therefore, the appropriate cost of effort function for the fiduciary is his unbiased effort cost. We will think of the principal's or advisee's objective as the maximization of value less the *unbiased fiduciary's* cost of effort. The problem FD also entails the advisor implementing only those investments that the advisor knows will generate a high return, given the mechanism of compensation.

Some fiduciaries receive fixed payments, some are paid on a project basis, and some are paid hourly, but most are paid through a percentage of the value of assets managed, and that percentage is market driven. Fewer are paid a fixed payment which is also market driven. If the compensation mechanism is a share, $s \in (0,1)$, of the total value created, the agent receives expected gross payment of

$$s \sum_{i=1}^{2} e_i \pi_i H_i.$$

If the agent is paid a fixed payment, then a gross payment of

$$F$$

is the agent's payoff, where $F < H_i$, $i = 1, 2$.

Note that problem FD generates net value maximization for the unbiased agent acting as the principal in that it invokes an investment only if the agent and principal expect it to generate a high return. Also note that FD does not specify a participation constraint for the advisor, because the advisor is a free lance consultant being paid on a market determined percentage or fixed basis and possibly working for multiple principals. The above description of the implementation of fiduciary duty is accurate for a majority of fiduciary relationships.

Problem FD embodies the notion that the advisor places themselves into the shoes of the principal and maximizes the principal's value less unbiased effort cost or surplus. Under this scheme, the unbiased advisor would generate social optimality or surplus maximization in his choice of efforts. For notational convenience, let $z_i \equiv \pi_i H_i$. The socially optimal allocation and the solution to FD equates the marginal benefits of e_i to their true or unbiased marginal costs:

$$z_i - c'(e_i) = 0, i = 1, 2. \tag{1}$$

3. A Biased Advisor and the Principal's Response

We wish to investigate the case where the advisor might be ideologically biased against investment 2. This might be an investment with a high carbon footprint, one with a low ESG rating, or one with a low DEI score. Generally, the advisor is suboptimally obsessed with avoiding certain investments which he believes are socially irresponsible. The investment manager prioritizes environmental, social, or governance factors over maximum financial returns, thereby breaching their fiduciary duty to the principal. Examples this type of bias might include unwarranted animus towards fossil fuel companies, gun companies, companies practicing legal labor relations, and companies obeying environmental regulations. Aronowitz (2023), Hupart (2025), Mulvany (2025), and Talgo (2023) discuss recent ESG lawsuits against fiduciaries of retirement plans and investment management companies. Crisuolo and Odom (2022) discuss ESG investing concerns for trustees, and McGowan gives a recent summary of the law concerning the pitfalls ESG investing.

We model this type of bias as emanating from the agent applying a greater than socially optimal effort cost to investment 2. We express this as $\alpha c(e_2)$,where $\alpha > 1$. If the advisor were to be biased with respect to investment 2, then the advisor would maximize $S^a(e) \equiv \sum_{i=1}^{2} e_i z_i - c(e_1) - \alpha c(e_2)$ and implement efforts as the solution to

$$z_1 - c'(e_1) = 0, \text{ and} \tag{2}$$

$$z_2 - \alpha c'(e_2) = 0. \tag{3}$$

We denote the socially optimal solution to (1) as e^* and the biased privately optimal solution to (2)-(3) as e^a. Note that this model of a biased advisor is equivalent to a model where the advisor undervalues the benefit of investment 2 by internalizing the lower value $\frac{e_2 z_2}{\alpha}$. The type of bias characterized here is one of subpar research on certain investments induced by excessive effort cost associated with that research.

The following result follows from Assumptions A.1-A.3. All proofs are provided in the Appendix.

Lemma 1. Let A.1-A3 hold. There exists an unique interior maximizer e^* *of (1). Further, for each* $\alpha \geqq 1$, *there exists a unique interior maximizer of (2)-(3),* $e^a > 0$. *The maximizer* e_2^a *is strictly decreasing in* α, *while* e_1^a *is invariant in* α, *so that* $e_1^* = e_1^a$ *and* $e_2^* > e_2^a$, with $e_2^a \to e_2^*$, as $\alpha \to 1$.

We concentrate on the fiduciary's duty of care (as opposed to the duty of loyalty) for the principal which includes acting with prudence, with diligence, with competence, with the ability to manage risk, with the desire to document all decision making, and with the intent to act in the best interest of the beneficiary. Due care is the level of care taken, when the fiduciary implements their duty of care. When civil legal action is taken against a fiduciary, the court must determine the standard or level of care that the fiduciary should have taken and then the court must examine the level of care actually taken by the advisor. If the advisor did not take due care and this can be linked to a monetary loss, then the court assesses monetary damages to be paid by the advisor to the principal. The magnitude of those damages would depend on the extent to which the advisor's care level deviates from due care.[2]

In our model, the duty of care is to implement the socially optimal effort in information acquisition, and due care level would be the socially optimal effort level, e_i^*. We will model the possibility of successful suit and assessment of damages by assuming that the principal will trigger an audit of the advisor's actions if there is an ex post observation of H_i and the advisor did not invest in i. In the event of an audit, the principal will observe the advisor's objective function and actions and will discover the advisor's care (effort) level e_i. The probability of discovery of e_i is ρ and the cost of the audit process is $g(\rho)$, $g : [0, 1) \to \mathbb{R}$. We assume that the court observes ρ, and, with probability ρ, the defendant wins the case against the advisor. The function g satisfies

[2]Duty of care is distinct from duty of loyalty. the latter refers to misappropriation by the fiduciary. See Cooter and Freedman (1991).

A.3 $g(0) = 0, g', g'' > 0$, for $1 > \rho > 0$, $g'(0) = 0$, and $g'(1) = \infty$.

With probability $(1-\rho)$, e_i can not be discerned and the case is dropped. If the advisor's effort is discovered and at least due care is taken, the case is dropped. If less than due care is discovered for e_i, the court assesses damages to be paid to the principal in the amount of

$$\delta(1-s)(e_i^* - e_i)H_i. \tag{4}$$

Thus, the amount of damage is the principal's lost high outcome share of the cash times the difference between expected returns at due care versus actual care taken. These damages are scaled per dollar by a positive number $\delta > 0$ to be determined by the court. We will discuss determination of the adjustment δ at a later point. If the advisor receives a fixed payment, then (4) becomes the special case where $s = 0$. All of the results that follow hold under either compensation method, so for efficiency we will present results for the fixed percentage case. An advisor intending to exert due care does not face damages (4). An advisor who intends to exert less than due care on investment i would face the following ex ante expected penalty

$$\rho\delta(1-s)(e_i^* - e_i)z_i. \tag{5}$$

The penalty is then imposed only if $(e_i^* - e_i) > 0$, and it is imposed with probability ρ.

Given the audit mechanism (5), the advisor's effort $e_1 = c'^{-1}(z_2)$ will equal e_1^* and there will be no penalty if there is an audit. Effort on the second investment will be the solution to

$$\max_{\{e_2\}} z_2e_2 - \alpha c(e_2) - \rho\delta(1-s)z_2(e_2^* - e_2). \tag{6}$$

The advisor's selection of e_2 is described by

$$z_2[1 + \rho\delta(1 - s)] - \alpha c'(e_2) = 0, \tag{7}$$

at an interior solution $e_2 \in (0, 1)$. At such a solution, the second order condition, SOC, is met $(-\alpha c''(e_2) < 0)$, and we have that effort e_2 can be expressed as a function of ρ,

$$e_2(\rho) = c'^{-1}(\frac{z_2[1 + \rho\delta(1 - s)]}{\alpha}) \tag{8}$$

with

$$\frac{\partial e_2}{\partial \rho} = \frac{z_2\delta(1 - s)}{\alpha c''(e_2)} > 0. \tag{9}$$

Increases in the discovery probability increase the advisor's information acquisition effort on investment 2. An interior solution can be rewritten as

$$z_2 = \frac{\alpha c'(e_2)}{(1 + \rho\delta(1 - s))}, \tag{10}$$

so that efficiency results if

$$\rho = \frac{\alpha - 1}{\delta(1 - s)}. \tag{11}$$

The ratio $(\alpha - 1)/\delta(1 - s)$ would have to be fractional, that is, $\alpha < \delta(1 - s) + 1$, for it to be possible for the principal to issue a feasible discovery probability which generates efficiency.[3] This will be thought of as the case where bias is not too severe. As an example suppose that the court is neutral and $\delta = 1$. For typical fiduciaries paid on a percentage basis, the fraction $(1 - s)$ is close to unity ranging from 0.98 to 0.995. The number $(\alpha - 1)$ would have to be less than 1.98 for it to be feasible to implement efficiency through a fractional discovery probability. For example, if

[3]If the advisor receives a fixed payment then $s = 0$ and a fractional ρ simply requires $\alpha < \delta + 1$.

the advisor possesses an effort cost $\alpha = 1.5$ times higher than optimal and $s = 0.98$, then the ρ implementing efficiency would equal 0.51. If $\alpha \geqq \delta(1-s)+1$, then no fractional discovery probability issued by the principal would invoke efficiency. This would be the case where bias is fairly severe. If $\alpha < \delta(1-s)+1$, then efficiency or under provision of effort is possible, depending on whether ρ is equal or less than $\frac{\alpha-1}{\delta(1-s)}$. For this feasible case, it remains to be seen whether the efficient solution is optimal for the principal.

The following result establishes the existence of an interior solution to the advisor's problem.

Lemma 2. Let A.1-A.3 hold. There exists a unique maximal e_2 solving $z_2[1 + \rho\delta(1-s)] - \alpha c'(e_2) = 0$.

Next, let us consider the principal's choice of a discovery or verification probability. The principal uses the advisor's incentive compatibility constraint $e_2 = e_2(\rho)$ to affect the agent's information acquisition effort on the second investment and solves

$$\max_{\{\rho\}} z_2 e_2(\rho) - c(e_2(\rho)) + \rho\delta(1-s)z_2(e_2^* - e_2(\rho)) - g(\rho). \tag{12}$$

The objective function (12) consists of total surplus plus expected damage recovery less the cost of ρ. An interior solution to (12), $\rho \in (0,1)$, would satisfy

$$[z_2 - c'(e_2)]\frac{\partial e_2}{\partial \rho} - z_2\rho\delta(1-s)\frac{\partial e_2}{\partial \rho} + \delta(1-s)z_2(e_2^* - e_2(\rho)) - g'(\rho) = 0. \tag{13}$$

In the region where there are damages and an audit, the marginal benefits of raising ρ consist of the surplus increment, $[z_2 - c'(e_2)]\frac{\partial e_2}{\partial \rho} \geqq 0$ and the marginal damage recovery, $\delta(1-s)z_2(e_2^* - e_2) \geqq 0$. The marginal costs are the lost damages as a result of raising e_2, $z_2 e_2 \rho\delta(1-s)\frac{\partial e_2}{\partial \rho}$, and the direct marginal cost of raising ρ, g'. The Appendix discusses the SOC to the principal's problem. We assume that the SOC is met. Let us denote this interior solution as ρ^r and $e^r = (e_1^r, e_2^r(\rho^r))$. In the following lemma, we show that a solution to (13) exists.

Lemma 3. Let A.1-A.3 hold and suppose that the SOC to the principal's problem is met. There exists a unique maximal $\rho^r \in (0,1)$ *solving (13).*

We saw above that if bias is severe ($\alpha \geqq \delta(1-s)+1$) then there is no feasible audit probability leading to efficiency. In general, can the audit mechanism lead to efficiency? The following proposition answers this question in the negative.

Proposition 1. Suppose that both the principal and the agent are risk neutral. Let assumptions A.1-A.3 hold and suppose that the SOC to the principal's problem is met. At the principal's equilibrium, the audit mechanism results in $e_1^r = e_1^*$ *being efficient, and effort* $e_2^r < e_2^*$ *being under supplied.*

Proposition 1 says that the principal will not voluntarily choose an audit probability that would generate efficiency in the choice of information acquisition effort for the project against which the fiduciary has bias. In fact, it is optimal to generate under supply of this effort under the audit mechanism. The reason can be seen in the net marginal benefit terms of ρ in the FOC (13) to the principal's problem. At the $\rho = \frac{\alpha-1}{\delta(1-s)}$ which would generate efficiency this marginal benefit is negative, implying that the optimal ρ must be a lesser value, $\rho^r < \frac{\alpha-1}{\delta(1-s)}$.

Whether the bias is severe or not, it is possible for the court to generate an optimal efficient choice of ρ through certain precommitments. Because the court must be fair and uniform in its commitments to all plaintiffs, its rules must not base δ on a particular' plaintiff's optimal choice of ρ, given δ, as would a Stackelberg leader. One possibility for a uniform rule is that the the court could precommit to an adjustment δ per dollar of damages given by

$$\delta = \frac{(\alpha-1)}{\underline{\rho}(1-s)}, \text{ for } \rho \geqq \underline{\rho}, \tag{14}$$

where $\underline{\rho} \in (0,1)$ is a minimum probability of detection for which the court would issue damages. Each principal-plaintiff would face (14) and a "burden of proof" defined by $\underline{\rho}$. For example, $\underline{\rho} = 0.51$ would represent a verification probability demanding a preponderance of evidence. The function δ would represent a damage multiplier.[4] This function is increasing in the degree of bias, $(\alpha - 1)$, and decreasing in the burden of proof, $\underline{\rho}$, as well as the share of the payoff retained by the principal, $(1-s)$. The greater is the image value of $\delta(\cdot)$ for any $(\underline{\rho}, (\alpha-1), (1-s))$, the harsher is the court. In our example, if $\underline{\rho} = 0.51$, $\alpha = 1.2$, and $s = 0.01$, then $\delta = 0.396$. Here, bias is not too severe and δ has the effect of dampening the damages below the expected damages. Alternatively, in the same example, with more severe bias, $\alpha = 1.8$, we have $\delta = 1.584$, so that damages are punitive (greater than expected damages). A court precommiting to (14) will be termed "active".

With an active court, the principal's problem is modified to

$$\max_{\{\rho\}} z_2 e_2(\rho) - c(e_2(\rho)) + \frac{\rho}{\underline{\rho}}(\alpha - 1) z_2 (e_2^* - e_2(\rho)) - g(\rho), \text{ subject to } \rho - \underline{\rho} \geqq 0. \tag{15}$$

The following result characterizes the equilibrium with an active court.

Proposition 2. Suppose that both the principal and the agent are risk neutral. Let assumptions A.1-A.3 hold and suppose that the SOC to the principal's problem is met. If the court is active, the principal chooses $\rho = \underline{\rho}$ and the efficient solution $e^r = e^*$ is attained.

Our active court mechanism represents a more explicit and detailed explanation of the total damage multiplier outlined in the law and economics literature, e.g. Polinski and Shavell (1998). The intuition was that full damages are internalized by the biased party if expected damages are inflated to full damages by multiplying by the reciprocal of the probability of detection. With full internalization, efficiency results. The trouble with this argument is that the probability of

[4] The ratio $\delta = \frac{(\alpha-1)}{\underline{\rho}(1-s)}$ is an example of the damage multiplier discussed in Polinski and Shavell (1998), p.889.

detection is not exogenous nor are the expectation damages. The audit probability is a costly choice variable of the firm and damages are affected by the agent's choice of efforts. If the principal thought that the court would multiply its expected damages by the reciprocal of its ρ, it could strategically lower ρ to inflate damages and at the same time pay less cost of ρ. No equilibrium would result. Our process is different in that the court precommits to (14) for any plaintiff coming before the court and for any endogenous choice of ρ subject to its uniform rules. In our process the the court does not optimize over a choice of δ assuming that one particular plaintiff will optimize over their audit probability defining a best response function to that δ, as would be the case where the court is a Stackelberg leader in choosing δ.

4. A Biased Advisor and the Principal's Response with a Risk Averse Principal

In this section we discuss the basic problem with a risk averse principal and a risk neutral agent. The principal's utility function is denoted $u(\cdot)$, where u satisfies

A.4 (i) $u(0) = 0, u' > 0$, and $u'' < 0$, for all positive monetary values.

(ii) For any loss $L_i < 0$, the utility of the loss is $-u(-L_i)$.

Assumption A.4 (ii) asserts that the risk averse principal is not subject to loss aversion. We must modify Assumption A.1 to encompass the case of risk aversion. We assume

A.1' (i) $H_i > 0$, $L_i < 0$.

(ii) $\pi u(H_i) - (1 - \pi)u(-L_i) < 0$.

Under A.1' both the agent and the principal do not want to invest in project i if the low outcome or no information obtains. Only the high outcome warrants investment.

First, consider the altered FD problem. The objective function is

$$S(e) = e_1\pi_1 e_2\pi_2 u(H_1+H_2) + e_1\pi_1(1-e_2\pi_2)u(H_1) + e_2\pi_2(1-e_1\pi_1)u(H_2) - c(e_1) - c(e_2). \tag{16}$$

We can rewrite (16) by defining $\Delta \equiv [u(H_1+H_2) - u(H_1) - u(H_2)]$. Under A.4, u is strictly subadditive and $\Delta < 0$. The social optimum solves

$$\max_{\{e\}} e_1\pi_1 e_2\pi_2\Delta + e_1\pi_1 u(H_1) + e_2\pi_2 u(H_2) - c(e_1) - c(e_2). \tag{RFD}$$

The FOC for the revised social optimum are

$$S_1 = \pi_1 u(H_1) + e_2\pi_1\pi_2\Delta - c'(e_1) = 0, \tag{17}$$

$$S_2 = \pi_2 u(H_2) + e_1\pi_1\pi_2\Delta - c'(e_2) = 0. \tag{18}$$

Under our assumptions, the SOC to this problem are met if the Hessian of the objective function has a positive determinant, $|H| \equiv c_1''(e_1)c_2''(e_2) - (\pi_1\pi_2\Delta) > 0$.[5] We assume that this condition is satisfied.

Note that the introduction of risk aversion makes the two e_i substitutes in the social welfare function, whereas they were independents in the risk neutrality case. We have that $S_{12} = \pi_1\pi_2\Delta < 0$. The marginal social benefit of effort one is increased by less of effort two and conversely. As above, denote the social optimum as e^*.

[5] If this condiiton is met, the (SOC) are met since $c_i'' > 0$, from A.3.

A biased advisor solves

$$\max_{\{e\}} S^a(e) = e_1\pi_1 e_2\pi_2\Delta + e_1\pi_1 u(H_1) + e_2\pi_2 u(H_2) - c(e_1) - \alpha c(e_2).$$

would choose efforts according to

$$S_1^a = \pi_1 u(H_1) + \pi_1 e_2\pi_2\Delta - c'(e_1) = 0, \tag{19}$$

$$S_2^a = \pi_2 u(H_2) + e_1\pi_1\pi_2\Delta - \alpha c'(e_2) = 0. \tag{20}$$

The relevant Hessian is $|H^a| = \alpha c_1''(e_1)c_2''(e_2) - (\pi_1\pi_2\Delta)^2$ which is positive, if our previous assumption $|H| > 0$ is met.[6] Using standard comparative statics, we have that

$$\frac{\partial e_1}{\partial \alpha} = \frac{-c'(e_2)\pi_1\pi_2\Delta}{|H^a|} > 0, \text{ and} \tag{21}$$

$$\frac{\partial e_2}{\partial \alpha} = \frac{-c'(e_2)c''(e_1)}{|H^a|} < 0. \tag{22}$$

Thus, the advisor's adverse treatment of investment 2 with under exertion of discovery effort causes an over exertion of effort on investment 1 for which there is no bias. That is, too much focus is placed on the project for which there is no animus, while too little focus is placed on the distasteful investment. Denote the agent's optimum as e^a.

Under the Inada conditions, both e^* solving (17)-(18) and e^a solving (19)-(20) exist. The proof of Lemma 1 can be modified to demonstrate this result. If the respective SOC to these problems are met, then each is a maximum.

Now we suppose that the principal uses the audit mechanism to attempt to control bias. The

[6] If this condiiton is met, the (SOC) are met since $c_1'', \alpha c_2'' > 0$, from A.3.

agent will face expected damages given by

$$-\rho\delta\{[e_1^*\pi_1 e_2^*\pi_2(1-s)H_2 + e_1^*\pi_1(1-e_2^*\pi_2)(1-s)H_2]$$
$$-[e_1\pi_1 e_2\pi_2(1-s)H_2 + e_1\pi_1(1-e_2\pi_2)(1-s)H_2]\}.$$

Factor the term $(1-s)H_2$, and we can rewrite damages as the much simpler expression

$$\rho\delta(1-s)H_2(e_2^*\pi_2 - e_2\pi_2).$$

The agent's problem is

$$\begin{aligned} &\max_{\{e\}} e_1\pi_1 e_2\pi_2\Delta + e_1\pi_1 u(H_1) + e_2\pi_2 u(H_2) \\ &-\rho\delta(1-s)H_2(e_2^*\pi_2 - e_2\pi_2) - c(e_1) - \alpha c(e_2). \end{aligned} \tag{23}$$

Denoting the agent's objective function as $A(e)$, the relevant FOC for an interior solution are

$$A_1 = \pi_1 u(H_1) + \pi_1 e_2\pi_2\Delta - c'(e_1) = 0, \tag{24}$$

$$A_2 = \pi_2 u(H_2) + e_1\pi_1\pi_2\Delta + \rho\delta(1-s)H_2\pi_2 - \alpha c'(e_2) = 0. \tag{25}$$

The second order conditions are met if $|H^a| = \alpha c_1''(e_1)c_2''(e_2) - (\pi_1\pi_2\Delta)^2 > 0$,which follows from our previous assumption. Employing comparative static techniques, we can show that

$$\frac{\partial e_1}{\partial \rho} = \frac{\pi_1(\pi_2)^2\Delta\delta(1-s)H_2}{|H^a|} < 0, \text{ and} \tag{26}$$

$$\frac{\partial e_2}{\partial \rho} = \frac{c''(e_1)\delta(1-s)\pi_2 H_2}{|H^a|} > 0. \tag{27}$$

A greater discovery probability raises the effort on the biased investment, but, in contrast to the risk neutral case, it lowers effort on the project with no bias.

Let us again denote the solution to the principal's problem e^r. If the principal adjusts ρ to

$$\rho = \frac{(\alpha - 1)c'(e_2^*)}{\delta(1-s)\pi_2 H_2}, \tag{29}$$

then (24) and (25) imply that efficiency obtains, $e^r = e^*$. If the right side of (29) is fractional, $(\alpha - 1)c'(e_2^*) < (1-s)\pi_2 H_2$, then it is feasible for the principal to implement efficiency but not necessarily optimal. In this case, the agent's excess marginal effort cost is less than the principal's share of the marginal payoff from the agent's effort. However, if $\frac{(\alpha-1)c'(e_2^*)}{\delta(1-s)\pi_2 H_2} \geqq 1$, then efficiency through a fractional discovery probability is not feasible. This is the case where bias is severe, $(\alpha - 1)c'(e_2^*) \geqq (1-s)\pi_2 H_2$, and the agent's excess marginal effort cost is at least as great as the principal's share of the marginal payoff from the agent's effort.

If the payment to the agent is fixed then we can set $s = 0$ in the above description and the above analysis applies. Again, the marginal conditions are unaffected by a fixed payment, so that the same results applicable to the percentage of returns payment method will apply to the the fixed payment case. We proceed with the fixed percentage version.

The principal's problem with risk aversion is

$$\begin{aligned} &\max_{\{\rho\}} \; e_1(\rho)\pi_1 e_2(\rho)\pi_2\Delta + e_1(\rho)\pi_1 u(H_1) + e_2(\rho)\pi_2 u(H_2) \\ &+\rho\delta(1-s)H_2(e_2^*\pi_2 - e_2(\rho)\pi_2) - c(e_1(\rho)) - c(e_2(\rho)) - g(\rho) \end{aligned} \tag{30}$$

The FOC to (30) is

$$\begin{aligned} 0 &= [\pi_1 e_2(\rho)\pi_2\Delta + \pi_1 u(H_1) - c'(e_1(\rho))](\partial e_1/\partial \rho) \\ &+[\ e_1(\rho)\pi_1\pi_2\Delta + \pi_2 u(H_2) - c'(e_2)](\partial e_2/\partial \rho) \\ &-\rho\delta(1-s)H_2\pi_2(\partial e_2/\partial \rho) + \delta(1-s)H_2(e_2^*\pi_2 - e_2(\rho)\pi_2) - g'(\rho). \end{aligned}$$

The first term is zero from the agent's problem (24). The FOC then becomes analogous to that of the risk neutral case

$$\begin{aligned} 0 &= [\ e_1(\rho)\pi_1\pi_2\Delta + \pi_2 u(H_2) - c'(e_2)](\partial e_2/\partial \rho) \\ &-\rho\delta(1-s)H_2\pi_2(\partial e_2/\partial \rho) + \delta(1-s)H_2(e_2^*\pi_2 - e_2(\rho)\pi_2) - g'(\rho). \end{aligned} \tag{31}$$

Using the proof of Lemma 3 with slight modifications, it is easy to show that a solution $\rho \in (0,1)$ exists to equation (31). This interior solution is a maximum if the relevant SOC is met. In the Appendix, we discuss this condition and assume that it is met. At such a solution, both efforts are distorted relative to the social optimum. Effort on the biased project is under supplied $e_2^r < e_2^*$, with

$$\pi_2 u(H_2) + e_1\pi_1\pi_2\Delta - c'(e_2) = (\alpha - 1)c'(e_2) - \rho(1-s)H_2\pi_2 > 0.$$

This is true because otherwise there are no damages.[7] Given this result, consider e_1^r and its relationship to e_1^*. Define the marginal benefit of effort 1 as $MB_1(e_2) = \pi_1 u(H_1) + \pi_1 e_2\pi_2\Delta$, where $MB_1'(e_2) < 0$. At an efficient solution, $MB_1(e_2^*) = c'(e_1^*)$. By the facts that $e_2^r < e_2^*$ and $MB_1'(e_2) < 0$, we have that $MB_1(e_2^r) > MB_1(e_2^*)$. This inequality in turn implies that $MB_1(e_2^r) = c'(e_2^r) > c'(e_1^*)$. It follows that by $c'' > 0$, $e_1^r > e_1^*$. Effort 1 is over supplied. This result is important,

[7] In the proof of Proposition 3 to follow, we rigorously show that e_2 is under supplied.

because it implies that no damages will be assessed with respect to the first project. We summarize our results in

Proposition 3. Suppose that the principal is risk averse and the agent is risk neutral. Let assumptions A.1'-A.3 and A.4 hold. Assume that the SOC to the principal's problem is met. At the principal's equilibrium, the audit mechanism results in $e_1^r > e_1^*$ *being over supplied and effort* $e_2^r < e_2^*$ *being under supplied.*

The risk averse principal will not attain efficiency through the audit mechanism just as in the risk neutral case. As in the latter case, the effort exerted on the project for which there is bias is under supplied. However, the new feature is that the effort for which there is no bias turns out to be oversupplied, due to the fact that risk aversion makes the two efforts substitutes in the agent's objective function.

At this second best characterized by (31), the principal's objective function is not total surplus, S, nor is it total surplus less audit costs, $S - g$. A question of interest is whether an upward perturbation of ρ, in equilibrium, would lead to an increase in welfare. That is, from a policy perspective, would it be good to invoke institutional measures which would encourage the principal to exert more effort in auditing the agent than would be privately optimal? Given the above discussion, it is easy to see that

$$\frac{dS}{d\rho} = [\ e_1(\rho)\pi_1\pi_2\Delta + \pi_2 u(H_2) - c'(e_2)](\partial e_2/\partial\rho) > 0,$$

so that a greater ρ would lead to greater social surplus defined as S. However, if we define total surplus as "net total surplus", $(S - g)$, we can write

$$(\frac{dS}{d\rho} - g'(\rho)) = \rho\delta(1-s)\pi_2 H_2\frac{\partial e_2}{\partial\rho} - \delta(1-s)\pi_2 H_2((e_2^* - e_2(\rho)),$$

using the FOC to the principal's problem. We can show that

$$(\frac{dS}{d\rho} - g'(\rho)) \gtreqless 0 \text{ if } \frac{\partial e_2}{\partial \rho}\frac{\rho}{e_2} \gtreqless \frac{e_2^* - e_2}{e_2}.$$

Thus, net total surplus rises in equilibrium with an increase in ρ if the effort elasticity on the biased project with respect to ρ is greater than the percentage distortion in that effort from the first best effort. The converse also holds. That is, when the percent distortion in effort 2 from the first best is greater than the audit effort elasticity, then more intense monitoring relative to the second best actually decreases net surplus.

An active court can help the principal achieve efficiency in equilibrium. The court could precommit to an adjustment δ per dollar of damages given by

$$\delta = \frac{(\alpha - 1)c'(e_2^*)}{\underline{\rho}(1-s)\pi_2 H_2}, \text{ for } \rho \geqq \underline{\rho}. \tag{32}$$

The fraction $\underline{\rho}$ is, as above, a minimum probability of detection for which the court would issue damages. We have

Proposition 4. Suppose that the principal is risk averse and the agent is risk neutral. Let assumptions A.1'-A.3 and A.4 hold. Assume that the SOC to the principal's problem is met. If the court is active, the principal chooses $\rho = \underline{\rho}$ and the efficient solution $e^r = e^*$ is attained.

5. Conclusion

The standard audit process used for breach of the duty of care in a fiduciary relationship does not lead to economic efficiency. One remedy for the inefficient outcome is for the court to become active in assessing damages by precommiting to applying a multiplier to expectation damages which is a function of a minimum audit probability and the degree of bias of the agent found guilty of a breach of the duty of care. We have specified such a function and have shown that it can inflate or deflate expectation damages depending on the degree of bias of the agent found guilty of breach. We have characterized this damage multiplier for cases where the principal is risk neutral and risk averse. In the risk neutral case it increases in the degree of the bias and decreases in the share of payoff going to the principal as well as the burden of proof probability. In the risk aversion case, the multiplier is increasing in the agent's excess marginal effort cost (degree of bias) and decreasing in the share of the principal in the marginal payoff from the agent's effort as well as the burden of proof probability.

Appendix

Proof of Lemma 1: For solution (1) to exist, it suffices that the Inada conditions in A.3 imply $\lim_{e_i \to 0} [z_i - c'(0)] > 0$ and $\lim_{e_i \to 1} [z_i - c'(1)] < 0$. For the solution to (2)-(3) to exist, an analogous argument applies. Thus, there exists a $e_i^a \in (0, 1)$ such that (2)-(3) are met for each $\alpha > 0$. Further, each solution is unique because $[z_i - c'(e_i)], [sz_2 - \alpha c'(e_2)]$ are strictly decreasing in $e_i, i = 1, 2$, and e_2, respectively.

We must show e_2^a is strictly decreasing in α, while e_1^a does not vary with α. To see the latter, rewrite (2) as $e_1^a = c'^{-1}(e_2 z_2)$. Next, (3) implies $e_2^a = c'^{-1}(\frac{e_2 z_2}{\alpha})$. Because $c'^{-1}(\frac{e_2 z_2}{\alpha})$, is decreasing in α and $e_2^a \to e_2^*$ as $\alpha \to 1$, the results hold. ■

Proof of Lemma 2: Consider $z_2[1 + \rho\delta(1 - s)] - \alpha c'(e_2) = 0$ and take

$$\lim_{e_2 \to 0} \{z_2[1 + \rho\delta(1 - s)] - \alpha c'(e_2)\} = \{z_2[1 + \rho\delta(1 - s)] - 0\} > 0.$$

Moreover note that

$$\lim_{e_2 \to 1} \{z_2[1 + \rho\delta(1-s)] - \alpha c'(e_2)\} = \{z_2[1 + \rho\delta(1-s)] - \alpha(\infty)\} < 0.$$

By continuity and global satisfaction of the SOC, the result holds. ■

Proof of Lemma 3: We can use the Inada conditions in A.3 to show that there exists a ρ^r solving (13). We have that the limit of (13) as $\rho \to 0$ is

$$[z_2 - c'(e_2^a)]\frac{z_2\delta(1-s)}{\alpha c''(e_2^a)} + \delta(1-s)z_2(e_2^* - e_2^a) - 0 > 0,$$

while the limit of (13) as $\rho \to 1$ is

$$[z_2(1-\delta(1-s)) - c'(e_2(1))]\frac{z_2\delta(1-s)}{\alpha c''(e_2(1))} + \delta(1-s)z_2(e_2^* - e_2(1)) - \infty \to -\infty < 0.$$

It follows that there exists a $\rho^r \in (0,1)$ solving (13). If the SOC is met, ρ^r is unique. ■

The Second Order Condition in the Principal's Problem with Risk Neutrality: The SOC corresponding to the interior solution is

$$\begin{aligned} 0 \;>\; & [z_2(1-\rho\delta(1-s)) - c'(e_2)]\frac{\partial^2 e_2}{\partial \rho^2} - 2\delta(1-s)z_2\frac{\partial e_2}{\partial \rho} \\ & -c''(e_2)(\frac{\partial e_2}{\partial \rho})^2 - g''(\rho), \end{aligned} \tag{a1}$$

where

$$\frac{\partial^2 e_2}{\partial \rho^2} = \frac{z_2\delta(1-s)}{\alpha}(-1)(\alpha c''(e_2))^{-2}\alpha c'''(e_2)\frac{\partial e_2}{\partial \rho} \leqq 0. \tag{a2}$$

Assuming the latter inequality, the SOC (a1) is met if $[z_2(1-\rho(1-s)) - c'(e_2)] \geqq 0$.The latter condition is true at a point where the FOC are satisfied if $[2z_2 - (\alpha+1)c'(e_2)] \geqq 0$. If $[z_2(1-\rho\delta(1-$

$s)) - c'(e_2)] < 0$, then the first term of (a1) is nonnegative and it would have to be dominated by the last three negative terms.

Proof of Proposition 1: If bias is severe, $\frac{\alpha-1}{\delta(1-s)} \geqq 1$, then it must be that $\rho^r < \frac{\alpha-1}{\delta(1-s)}$, so that the result holds. Next, consider the case where bias is not severe, $\frac{\alpha-1}{\delta(1-s)} < 1$, so that a fractional audit probability producing efficiency is feasible. Let us evaluate FOC (15) at $\rho = \frac{\alpha-1}{\delta(1-s)}$. We have

$$\begin{aligned} & [z_2 - c'(e_2)]\frac{\partial e_2}{\partial \rho} - z_2\rho\delta(1-s)\frac{\partial e_2}{\partial \rho} + \delta(1-s)z_2(e_2^* - e_2) - g'(\rho) \\ = \; & 0 - z_2(\alpha-1)\frac{\partial e_2}{\partial \rho} + 0 - g'(\frac{\alpha-1}{\delta(1-s)}) \\ = \; & -z_2(\alpha-1)\frac{z_2\delta(1-s)}{\alpha c''(e_2^*)} - g'(\frac{\alpha-1}{\delta(1-s)}) < 0. \end{aligned}$$

The negativity of this expression implies that $\rho^r < \frac{\alpha-1}{\delta(1-s)}$, so that from the agent's problem, $z_2 - c'(e(\rho^r)) > 0$, by $z_2 - c'(e(\frac{\alpha-1}{\delta(1-s)})) = 0, \rho^r < \frac{\alpha-1}{\delta(1-s)}$, and $c', e(\rho)$ increasing. Thus, e_2 is under supplied. ■

Proof of Proposition 2: The Lagrangian function for the principal's problem is

$$L = z_2e_2(\rho) - c(e_2(\rho)) + \frac{\rho}{\underline{\rho}}(\alpha-1)z_2(e_2^* - e_2(\rho)) - g(\rho) + \gamma(\rho - \underline{\rho}).$$

Evaluate the derivative of L at $\rho = \underline{\rho}$:

$$\begin{aligned} 0 \; = \; & [z_2e_2(\underline{\rho}) - c(e_2(\underline{\rho}))]\frac{\partial e_2}{\partial \rho} + \frac{1}{\underline{\rho}}(\alpha-1)z_2(e_2^* - e_2(\underline{\rho})) - \frac{\rho}{\underline{\rho}}(\alpha-1)z_2\frac{\partial e_2}{\partial \rho} - g'(\underline{\rho}) + \gamma \\ = \; & -1(\alpha-1)z_2\frac{\partial e_2}{\partial \rho} - g'(\underline{\rho}) + \gamma. \end{aligned}$$

It follows that $\gamma > 0$ and $\rho = \underline{\rho}$. ■

Proof of Proposition 3: First let us show that $\rho^r < \frac{(\alpha-1)c'(e_2^*)}{\delta(1-s)\pi_2H_2}$. If $\frac{(\alpha-1)c'(e_2^*)}{\delta(1-s)\pi_2H_2} \geqq 1$, then $\rho^r < \frac{(\alpha-1)c'(e_2^*)}{\delta(1-s)\pi_2H_2}$ by virtue of feasibility. Let $\frac{(\alpha-1)c'(e_2^*)}{\delta(1-s)\pi_2H_2} < 1$ and evaluate the principal's FOC at $\rho =$

$\frac{(\alpha-1)c'(e_2^*)}{\delta(1-s)\pi_2 H_2}$. We have

$$-(\alpha-1)c'(e_2^*)(\partial e_2(\frac{(\alpha-1)c'(e_2^*)}{\delta(1-s)\pi_2 H_2})/\partial\rho) - g'(\frac{(\alpha-1)c'(e_2^*)}{\delta(1-s)\pi_2 H_2}) < 0.$$

It again follows that $\rho^r < \frac{(\alpha-1)c'(e_2^*)}{\delta(1-s)\pi_2 H_2}$.

Next consider the agent's FOC for e_2 evaluated at $\rho = \frac{(\alpha-1)c'(e_2^*)}{\delta(1-s)\pi_2 H_2}$

$$\begin{aligned} & \pi_2 u(H_2) + e_1(\frac{(\alpha-1)c'(e_2^*)}{\delta(1-s)\pi_2 H_2})\pi_1\pi_2\Delta - c'(e_2(\frac{(\alpha-1)c'(e_2^*)}{\delta(1-s)\pi_2 H_2})) \\ = \quad & (\alpha-1)[c'(e_2(\frac{(\alpha-1)c'(e_2^*)}{\delta(1-s)\pi_2 H_2})) - c'(e_2^*)] = 0. \end{aligned}$$

Because $\rho^r < \frac{(\alpha-1)c'(e_2^*)}{\delta(1-s)\pi_2 H_2}$, lowering ρ from a level of $\frac{(\alpha-1)c'(e_2^*)}{\delta(1-s)\pi_2 H_2}$ results in $\pi_2 u(H_2) + e_1\pi_1\pi_2\Delta - c'(e_2)$ taking on a greater magnitude than zero. This follows from the fact that a lower ρ raises $\pi_2 u(H_2) + e_1\pi_1\pi_2\Delta - c'(e_2)$. To see this note that

$$\begin{aligned} \frac{d[\pi_2 u(H_2) + e_1\pi_1\pi_2\Delta - c'(e_2)]}{d\rho} &= \frac{\pi_1(\pi_2)^2\Delta\delta(1-s)H_2\pi_1\pi_2\Delta - c''(e_1)\delta(1-s)\pi_2 H_2(c''(e_2))}{|H^a|} \\ &= \frac{(\delta(1-s)\pi_2 H_2)(-|H|)}{|H^a|} < 0. \end{aligned}$$

Thus, as ρ is lowered below $\frac{(\alpha-1)c'(e_2^*)}{\delta(1-s)\pi_2 H_2}$, the net marginal benefit $\pi_2 u(H_2) + e_1\pi_1\pi_2\Delta - c'(e_2)$ rises from zero to be positive. We have under supply of e_2. In the text, we show that e_1 is over supplied.

■

The Second Order Condition in the Principal's Problem with Risk Aversion: The SOC is given by

$$\begin{aligned} 0 &> [\pi_1 e_2(\rho)\pi_2\Delta + \pi_1 u(H_1) - c'(e_1(\rho)) - \rho(1-s)\pi_2 H_2](\partial^2 e_2/\partial\rho^2) \\ &\quad +\pi_1\pi_2\Delta(\partial e_1/\partial\rho)(\partial e_2/\partial\rho) - c''(e_2)(\partial e_2/\partial\rho)^2 - (1-s)\pi_2 H_2(\partial e_2/\partial\rho) - g''(\rho), \quad \text{(a3)} \end{aligned}$$

where

$$\partial^2 e_2/\partial\rho^2 = \frac{\pi_2 H_2(1-s)}{(|H_A|)^3}[c'''(e)\pi_1\pi_2\Delta(1-\alpha c''(e_1)c''(e_2)) - (c''(e_1))^2(c''(e_2))^2].$$

The last three terms of (a3) are negative, but the first is generally indeterminate. The first term of (a.3) contains expression

$$\begin{aligned} & [\pi_1 e_2(\rho)\pi_2\Delta + \pi_1 u(H_1) - c'(e_1(\rho)) - \rho(1-s)\pi_2 H_2] \\ = \ & [2(\pi_1 e_2(\rho)\pi_2\Delta + \pi_1 u(H_1)) - (\alpha+1)c_2'(e_2)] \end{aligned}$$

where the latter equality holds from the agent's problem. The expression

$$[2(\pi_1 e_2(\rho)\pi_2\Delta + \pi_1 u(H_1)) - (\alpha+1)c_2'(e_2)] \geqq 0, \quad \text{(a4)}$$

if the bias is not too great. The second derivative $\partial^2 e_2/\partial\rho^2$ is negative if $(1-\alpha c''(e_1)c''(e_2)) > 0$. If this is true, then assuming (a4), (a3) is negative. Otherwise, the first term of (a3) is not generally signable. If it is positive, then the last three terms would have to dominate.

Proof of Proposition 4: If γ is the multiplier for the constraint $\rho - \underline{\rho} \geqq 0$, the principal's

Lagrangian is

$$
\begin{aligned}
L &= e_1(\rho)\pi_1 e_2(\rho)\pi_2\Delta + e_1(\rho)\pi_1 u(H_1) + e_2(\rho)\pi_2 u(H_2) \\
&\quad +\rho\frac{(\alpha-1)c'(e_2^*)}{\underline{\rho}\pi_2}(e_2^*\pi_2 - e_2(\rho)\pi_2) - c(e_1(\rho)) - c(e_2(\rho)) - g(\rho) + \gamma(\rho-\underline{\rho})
\end{aligned}
$$

The principal's FOC for the choice of ρ is

$$
\begin{aligned}
0 &= [\, e_1(\rho)\pi_1\pi_2\Delta + \pi_2 u(H_2) - c'(e_2)](\partial e_2/\partial\rho) \\
&\quad -\rho\frac{(\alpha-1)c'(e_2^*)}{\underline{\rho}}(\partial e_2/\partial\rho) + \frac{(\alpha-1)c'(e_2^*)}{\underline{\rho}\pi_2}(e_2^*\pi_2 - e_2(\rho)\pi_2) - g'(\rho) + \gamma.
\end{aligned}
$$

Evaluating this FOC at $\rho = \underline{\rho}$, we obtain

$$
-(\alpha-1)c'(e_2^*)(\partial e_2/\partial\rho) - g'(\underline{\rho}) = -\gamma,
$$

so that $\gamma > 0$ and $\rho = \underline{\rho}$. ■

References

[1] Aronowitz, D. (2023). “The First ESG Breach of Fiduciary Duty Lawsuits.” https://encorefiduciary.com/the-first-esg-breach-of-fiduciary-duty-lawsuits/

[2] Cooter, R. and B.J. Freedman. (1991). "The Fiduciary Relationship: Its Economic Character and Legal Consequences." *New York University Law Review* 66, 1045-1076.

[3] Criscuolo, A. and M. Odom. (2022). “An ESG Investing Hazard For Fiduciaries.” https://www.palisadeshudson.com/2022/07/an-esg-investing-hazard-for-fiduciaries/

[4] Easterbrook, F.H. and D.R. Fischel. (1993). Contract and Fiduciary Duty. *Journal of Law and Economics* 36,425-446.

[5] Havens, J.J. and P.G. Schervish. (2003). Why the $41 trillion Wealth Transfer Estimate is Still Valid: A Review of Challenges and Questions. *The Journal of Gift Planning* 7,11-15.

[6] Hupart, J. (2025). "Federal Court Rules ESG-Guided Investing of 401(k) Plan Is a Breach of Fiduciary Duty." *National Law Review*: 15. https://natlawreview.com/article/federal-court-rules-esg-guided-investing-401k-plan-breach-fiduciary-duty

[7] Kelly, D.B. (2022). On Disgorgement and Punitive Damages in Trust Law. *Iowa Law Review* 107, 2079-2031.

[8] McGowan, J. (2022). "The Trouble with Tibble: Environmental, Social, and Governance (ESG) and Fiduciary Duty." *The University of Chicago Business Law Review: Online Edition.* https://businesslawreview.uchicago.edu/online-archive/trouble-tibble-environmental-social-and-governance-esg-and-fiduciary-duty

[9] Mulvaney, E. (2025). "Judge Rules American Airlines Violated Retirement Plan Duties by Encouraging ESG Investing." *The Wall Street Journal.* https://www.wsj.com/finance/investing/judge-rules-american-airlines-violated-retirement-plan-duties-by-encouraging-esg-investing-5518d752

[10] Polinski, A.M. and S. Shavell. (1998). "Punitive Damages: An Economic Analysis." *Harvard Law Review* 111, 869-957.

[11] Sitkoff, R.H. (2011). "The Economic Structure of Fiduciary Law." *Boston University Law Review* 91, 1039-1049.

[12] Talgo, C. (2023). "A Lawsuit Waiting to Happen: ESG Violates Fiduciary Duty." *The Hill*. https://thehill.com/opinion/4059114-a-lawsuit-waiting-to-happen-esg-violates-fiduciary-duty-to-investors/