
BIASED MONITORS: CORPORATE GOVERNANCE WHEN MANAGERIAL ABILITY IS MIS-ASSESSED*

BENJAMIN E. HERMALIN[†]

ABSTRACT

An important dimension to corporate governance is the assessment of managers. When managers vary in ability, determining who is good and who is not is vital. Moreover, knowing they will be assessed can lead those being assessed to behave in ways that make them appear better. Such signal-jamming behavior can be beneficial (*e.g.*, an executive works harder on behalf of shareholders) or harmful (*e.g.*, the behavior is myopic, boosting short-term performance at the expense of long-term success). In standard models of assessment, it is assumed those doing the assessing behave according to Bayes Theorem. But what if the assessors suffer from one of many well-documented cognitive biases that makes them less-than-perfect Bayesians? This paper begins an exploration of that issue by considering the consequence of one such bias, the base-rate fallacy, for two of the canonical assessment models: career-concerns and optimal monitoring and replacement. Although firms can suffer due to the base-rate fallacy, they can also benefit from this bias.

Keywords: Corporate Governance, Career Concerns, Learning and Assessment, Cognitive Biases

JEL Classification: G34, M12, D83, D81

*The author gratefully acknowledges the financial support of the Thomas & Alison Schneider Distinguished Professorship in Finance.

[†]University of California, Berkeley

CONTENTS

1	INTRODUCTION	1
2	MEANS OF UPDATING	2
2.1	Bayesian Learning	2
2.2	Biases in Updating	3
3	CAREER-CONCERNS UNDER THE BASE-RATE FALLACY	4
3.1	The Model without Effort	6
3.2	The Model with Effort	7
4	EMPLOYMENT DECISIONS	15
4.1	The Executive's Behavior	17
4.2	The Firm's Investment in Information	18
4.3	When Compensation Changes with Effort	21
5	DISCUSSION AND CONCLUSIONS	23
	APPENDIX A: PROOFS	24
	REFERENCES	26

1 INTRODUCTION

Since Fama (1980) and Holmstrom (1999) [1982],¹ it has been understood that many key phenomena in corporate governance (agency more generally) derive from the need to assess managerial ability. In particular, such assessment is the source of incentives, both desired (*e.g.*, greater efforts due to career concerns) and perverse (*e.g.*, forgoing profitable investments and other instances of managerial myopia); as well as being key to understanding phenomena around firms' choices of managers (CEOs) and, even, board composition. Much of this literature is surveyed in a forthcoming piece by Hermalin and Weisbach.

As discussed in Hermalin and Weisbach (forthcoming), the models in this literature rely on Bayesian updating (typically, the normal-learning model). A danger in relying on Bayesian updating, as those authors note at the end of their chapter, is that there is a large psychological literature that indicates that most people are, in fact, *not* Bayesian updaters; that is, they revise their beliefs upon receiving new information in ways that are inconsistent with Bayes Law. Hermalin and Weisbach suggest that reexamining assessment models taking into account known biases in how people update beliefs could be a fruitful avenue for future research. This article is a beginning on that research agenda.

After a brief review of Bayesian learning, in particular the so-called normal-learning model, the idea of the base-rate fallacy is introduced. This is a well-documented bias in which those making assessments overweight new information and underweight their prior information (the base rates).

In Section 3, the consequences of the base-rate fallacy for Holmstrom's canonical model of career concerns are considered. The principal findings are that the more employers suffer from the base-rate fallacy, the harder executives work in equilibrium. This follows because how hard an executive works is a function of how much weight employers place on his current performance versus their prior assessment of him. The more weight on the former, which is influenced by his effort, the harder he works. Because the base-rate fallacy means more weight on current performance, an executive's incentives to work hard are greater. An employer that suffers less from the base-rate fallacy than a rival avoids losing money in expectation. The same is not necessarily true of the rival; it can lose money in expectation. On the other hand, there are circumstances in which it too can expect to make money over the course of the game. A firm would like to play against one that is a worse Bayesian than it (suffers more from the base-rate fallacy), but if it has to play against one that is better, then it does better the more Bayesian its rival is; that is, the worst rival is one that is only slightly more Bayesian than you.

Section 4 takes up the other canonical model of assessment: a firm's deciding whether to keep or fire its manager based on its assessment of his ability. As in Section 3, the less Bayesian is the firm, the harder its executive will work. This reflects the weighting effect outlined in the previous paragraph, but also

¹Holmstrom's paper was originally published in 1982 as a chapter in a hard-to-find festschrift for Lars Wahlbeck. In 1999, the *Review of Economic Studies* reprinted it.

the greater monitoring that a less Bayesian firm does. Whether a firm suffers from being non-Bayesian depends on its value for this greater effort and how that compares to (i) over-investing in monitoring; (ii) firing the executive too readily; and (iii) having to pay greater compensation (in the Section 4.3 version only). It turns out that, although a firm could do best if wholly Bayesian, this is not always true; in some circumstances some amount of deviation from being a perfect Bayesian maximizes the firm's value.

As indicated, although firms can lose from failing to be Bayesian, they can also benefit. Moreover, because at least in the Section 3 model, the executive tends to undersupply effort from the perspective of welfare, the base-rate fallacy can be welfare improving (even if not always profit improving).

The last section offers a brief conclusion, in which some of the empirical implications of the results are discussed, as well as next steps.

Some technical details, including proofs not given in the text, can be found in the Appendix.

2 MEANS OF UPDATING

2.1 BAYESIAN LEARNING

It is worth briefly reviewing the normal-learning model, which represents rational (Bayesian) updating of beliefs when the relevant parameters are normally distributed. This review will limit itself to settings relevant for this article, for a more extensive review see Hermalin and Weisbach (forthcoming).

Suppose that an employer's (shareholders') *expected* payoff is a function of the employee's (executive's) ability, $\alpha \in \mathbb{R}$. The employer is assumed not to know the employee's ability, but she knows its relevant statistical properties. Specifically, she knows that α is drawn according to a normal distribution with mean $\hat{\alpha}_0$ and variance $1/\tau_0$; that is, $\alpha \sim N(\hat{\alpha}_0, 1/\tau_0)$. When the variance is written in the form $1/\zeta$, ζ is referred to as the *precision* of the distribution.

Additionally, the employer observes signals that permit her to update her beliefs about the employee's ability. Specifically, let $s_t \in \mathbb{R}$ denote the signal she observes at time t (e.g., s_t is the realization of profit at time t or an indicator of whether the period- t project was successful). The signal in any given period is drawn from a distribution that is conditional on the employee's true ability. Specifically, assume $s_t = \alpha + \varepsilon_t$, where $\varepsilon_t \sim N(0, 1/\eta)$. As the signal could always be redefined as $\tilde{s} = s - \mathbb{E}\{\varepsilon\}$, there is no loss of generality in assuming $\mathbb{E}\{\varepsilon\} = 0$. Assume the ε_t are distributed independently of each other. Note that one can express the conditional distribution of s_t as $N(\alpha, 1/\eta)$.

It can be shown (see, e.g., DeGroot, 1970, p. 166), that the *posterior* distribution of ability given a sequences of signals s_1, \dots, s_t is normal with mean

$$\hat{\alpha}_t = \frac{\tau_0 \hat{\alpha}_0 + \eta \sum_{i=1}^t s_i}{\tau_0 + t\eta} = \frac{\tau_0 \hat{\alpha}_0 + t\eta \bar{s}}{\tau_0 + t\eta} = \frac{\tau_0}{\tau_0 + t\eta} \hat{\alpha}_0 + \frac{t\eta}{\tau_0 + t\eta} \bar{s}, \quad (1)$$

where \bar{s} is the arithmetic average of the t signals, and precision

$$\tau_t = \tau_0 + t\eta. \quad (2)$$

Observe, from the last equality in (1), that the posterior belief about ability is a weighted average of the prior belief and the signals. A generalization of this updating rule is

$$\hat{\alpha}_t = \lambda_t \hat{\alpha}_0 + (1 - \lambda_t) \bar{s}, \quad (3)$$

where $\lambda_t \in [0, 1]$. When

$$\lambda_t = \frac{\tau_0}{\tau_0 + t\eta},$$

the updating is consistent with Bayes Law; otherwise it is inconsistent.

2.2 BIASES IN UPDATING

There is a large body of psychological research that convincingly demonstrates that people often hold beliefs or take actions that are inconsistent with their having properly employed Bayes Law to account for new evidence.² In particular, the psychology literature documents a number of biases or decision-making fallacies that lead individuals to depart from rationality in their decision-making and, critically, to do so in predictable ways. One such departure is especially relevant here: the *base-rate fallacy*.³

The base-rate fallacy is a tendency to *underweight* base rates; that is, when people receive a signal, they revise their beliefs by more than Bayes Law would have them do. In terms of expression (3), the λ_t they use is less than $\tau_0/(\tau_0 + t\eta)$; that is, it violates the normal learning model. Numerous experiments have given test subjects information about the population (the base rate), and then subsequent information that can be used to answer a question. As an example, the experiment might describe a hypothetical diagnostic test for a rare disease: the subjects are told that the prevalence of some disease is, say, one in 10,000 in the population and there is a test for that disease that has only a one-percent false positive rate and a very high (perhaps even perfect) true positive rate. The subjects are then asked how likely is it that a patient who tests positive has the disease. The subjects' guesses are usually very high, often over 90%.⁴ The true answer, however, is less than one percent: if p is the true positive rate, then, utilizing Bayes Law, the posterior probability of having the disease based on a positive test is

$$\frac{p \times \frac{1}{10,000}}{p \times \frac{1}{10,000} + \frac{1}{100} \times \frac{9999}{10,000}} < \frac{1}{1 + 99.99} \approx .0099.$$

In other words, individuals underweight the base rate (the remarkably low prevalence of the disease) and place too much weight on new information (the signal—the test result).

²Some good introductions and overviews of this literature are Gilovich (1991), Plous (1993), and Kahneman (2011).

³See Kahneman (2011) for, *inter alia*, an overview of this and other biases.

⁴Having routinely run this experiment in my first-year MBA course, I can attest to such findings.

As suggested, the base-rate fallacy translates into the normal learning model as the λ_t in expression (3) underweighting the prior, $\hat{\alpha}_0$ and overweighting the signal(s). That is,

$$\lambda_t < \frac{\tau_0}{\tau_0 + t\eta}. \quad (4)$$

There are certainly other cognitive biases worth considering (as suggested, *e.g.*, in Hermalin and Weisbach, forthcoming). Some (*e.g.*, the “hot-hand” fallacy and the fundamental attribution bias) are similar in spirit to the base-rate fallacy, insofar as the predictive value of recent individual achievement is over-estimated. Others might arguably point in an opposite direction (*e.g.*, an over-confidence bias that caused someone to ignore or underweight new information); however, as will be evident, much of the analysis below readily translates to situations in which the inequality in (4) is reversed.

3 CAREER-CONCERNS UNDER THE BASE-RATE FALLACY

Holmstrom (1999) devised his career-concerns model to examine a conjecture of Fama’s (1980) that career concerns serve to motivate agents to work harder. Specifically, if the terms of future employment, particularly compensation, are a function of how able employers believe an agent to be and if his efforts affect the signals used by employers in estimating his ability, then the agent could have incentives to work hard (supply effort) in order to boost employers’ estimates of his ability and, thus, his compensation.

A simple, yet useful version of the Holmstrom career-concerns model is the following: a firm employs an executive (the CEO) who has a two-period working life. Each period, his contribution to firm profit (gross of his compensation) in period t is

$$s_t = e_t + \alpha + \varepsilon_t, \quad (5)$$

where α and ε_t are as above, with the same statistical properties, and $e_t \in \mathbb{R}_+$ is the executive’s action (effort) in period t .

Following Holmstrom (1999), assume that the executive has no better information about his ability than do employers at the start of the game; that is, like them, he knows only that his ability is drawn from $N(\hat{\alpha}_0, 1/\tau_0)$. See Holmstrom or Hermalin and Weisbach (forthcoming) for discussions of why an assumption of *ex ante* symmetry of information is reasonable in many situations.

Assume the executive’s utility in period t is $w_t - c(e_t)$, where w_t is his compensation that period and $c : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is a twice differentiable increasing function. To ensure unique interior maxima, assume that $c'(0) = 0$ and $c(\cdot)$ is strictly convex (for future reference, the latter assumption entails first-order conditions are sufficient as well as necessary). Consistent with the usual notion of cost, $c(0) = 0$.

Assume the executive’s action each period is a *hidden action*; that is, known to him, but not observable by anyone else. Although no one but the executive knows his action, e_t , in period t , there is a level of effort, \hat{e}_t , that interested parties (*i.e.*, current and potential employers) anticipate he will take. This

means that the other interested parties translate performance, s_t , in period t into a signal of ability by subtracting \hat{e}_t from s_t ; call this constructed signal \tilde{s}_t and observe

$$\tilde{s}_t \equiv s_t - \hat{e}_t = \alpha + \varepsilon_t + e_t - \hat{e}_t.$$

It follows from (1) and (2) that Bayesian observers should hold the following estimate of his ability after the first-period

$$\hat{\alpha}_1^B = \frac{\tau_0 \hat{\alpha}_0 + \eta(\alpha + \varepsilon_1 + e_1 - \hat{e}_1)}{\tau_0 + \eta} = \frac{\tau_0}{\tau_0 + \eta} \hat{\alpha}_0 + \frac{\eta}{\tau_0 + \eta} (\alpha + \varepsilon_1 + e_1 - \hat{e}_1). \quad (6)$$

The superscript B denotes Bayesian. The generalized version, along the lines of (3) is

$$\hat{\alpha}_1(\lambda) \equiv \lambda \hat{\alpha}_0 + (1 - \lambda)(\alpha + \varepsilon_1 + e_1 - \hat{e}_1). \quad (7)$$

For future reference, define

$$\lambda^B = \frac{\tau_0}{\tau_0 + \eta}.$$

Consistent with the base-rate fallacy, much of the focus will be on principals (firm owners) who update with a $\lambda < \lambda^B$.

Keeping with Holmstrom (1999), attention is limited to very simple employment contracts; to wit, a contract can last only one period and pays a salary (wage) that is *non*-contingent on performance. In essence, in each period the executive's compensation is whatever salary he was offered by that period's employer.

Because competition for managerial talent is a key component of this type of model, suppose that there are two firms that compete for the executive in each of the two periods. This is, to be sure, a somewhat artificial way of introducing competition; in part, because it begs the question of how does the firm that fails to employ the executive function. There are two potential answers. First, one could assume that both firms have access to alternative managers of the same and known ability, who are not as good as the executive in question; hence, all payoffs are relative to having employed one of these alternatives.⁵ Alternatively, the firm that fails to hire the executive in the first period goes out of business, with its "clone" entering the market in the second period. Likewise, the firm that fails to employ the executive in the second period shuts down.

In both periods, the firms compete for the executive. Consider the second period and let λ^1 be the weighting factor used by the executive's first employer when updating its estimate of his ability and let λ^2 be the factor used by the second firm (note both firms are assumed to observe s_1). The expected value of the executive in period 2 (gross of his compensation) to firm j is thus $\hat{\alpha}_1(\lambda^j) + \hat{e}_2$, which is its expected value of s_2 .

⁵Given the use of the normal distribution, the ability of the executive in question and, more critically, estimates of that ability are unbounded below. Hence, given a truly horrible signal, the firms might prefer alternative managers to the executive in question. Extending the analysis to allow for such a truncation below would not have much bearing on the results presented here; but would greatly complicate the analysis.

The executive's remaining lifetime utility in the second (terminal) period of his career is $w_2 - c(e_2)$, which reflects simple employment contracting and the fact that, being at the end of his career, he is indifferent to how employers (including his own) might assess his ability in the future. The effort that maximizes that expression is clearly 0; that is, $e_2 = 0$. Given that $e_2 = 0$ is a dominant strategy for the executive, it seems reasonable—even though they are arguably less than rational—that the employers understand that and, so, $\hat{e}_2 = 0$. Consequently, the value of the executive to firm j in period 2 is $\hat{\alpha}_1(\lambda^j)$.

Assume that the two firms bid for the executive via a second-price sealed bid auction (roughly equivalent to an ascending-bid English auction).⁶ Given that neither employer possesses private information, it follows that each employer should bid its $\hat{\alpha}_1(\lambda)$; that is, as is well known, it's a dominant strategy to bid one's value in a second-price auction under symmetric information).

An issue is how does each employer form its belief about the executive's first-period effort, \hat{e}_1 ; in particular, do the employers hold a common belief or not? With wholly rational actors, the standard notions of equilibrium require that the players correctly anticipate the strategies of other players in equilibrium; hence, if the executive plays the pure strategy e_1 , it should be that $\hat{e}_1 = e_1$. However, such a belief is rationalized via an understanding of the equilibrium and that, in turn, would require that each employer understand that it is potentially using a different λ than the other. Such an understanding, however, seems at odds with the idea that the employers suffer from a fallacy with respect to their updating.

It is worth postponing a resolution of that issue, by considering, as an aside, a variant of the model without effort.

3.1 THE MODEL WITHOUT EFFORT

Without effort,

$$\hat{\alpha}_1(\lambda^j) = \lambda^j \hat{\alpha}_0 + (1 - \lambda^j) \underbrace{(\alpha + \varepsilon_1)}_{s_1} = s_1 + \lambda^j(\hat{\alpha}_0 - s_1).$$

It follows that the employer who most suffers from the base-rate fallacy (*i.e.*, has the lower λ) wins the auction (*i.e.*, hires the executive in the second period) whenever the signal exceeds the prior estimate (*i.e.*, when $s_1 > \hat{\alpha}_0$) and loses the auction whenever the signal is less than the prior (*i.e.*, when $s_1 < \hat{\alpha}_0$).

In expectation, the winning firm's profit, accounting for the executive's compensation, is

$$\hat{\alpha}_1(\lambda^B) - \hat{\alpha}_1(\lambda^o) = (\lambda^B - \lambda^o)(\hat{\alpha}_0 - s_1),$$

where the superscript o refers to the other (losing) firm. Some immediate conclusions:

⁶The equivalence of a second-price sealed bid auction and an ascending-bid English auction does depend on bidders in the latter not learning about their own valuation from the bidding behavior of the other bidders. This could be a slightly problematic assumption when the employers use different weighting factors—would the bidding behavior of one employer lead another to realize she was behaving in a Bayesian manner? For those worried about such issues, it is fine to limit attention to second-price sealed bid auctions.

- If the losing firm is Bayesian (doesn't suffer from the base-rate fallacy), then the winning firm's expected profit is zero.
- If the losing firm suffers from the base-rate fallacy (so $\lambda^o < \lambda^B$), then the winning firm expects to suffer a loss if the signal exceeds the prior (*i.e.*, if $s_1 > \hat{\alpha}_0$); moreover, this loss is greater the less Bayesian is the losing firm (the lower is λ^o).
- If the losing firm suffers from the base-rate fallacy, then the winning firm expects to enjoy a profit if the signal is less than the prior (*i.e.*, if $s_1 < \hat{\alpha}_0$); moreover, this profit is greater the less Bayesian is the losing firm (the lower is λ^o).

In light of this analysis, in a model without effort, each firm would prefer to play against a rival that was a worse Bayesian than it (had a lower λ). This suggests that there could be pressure on firms to be more Bayesian in their assessments.

On the other hand, if each firm is equally bad (*i.e.*, $\lambda^1 = \lambda^2 < \lambda^B$), then neither firm loses profit in expectation. Because they offer the same bid, they win half the time (assume the executive flips a coin when indifferent), each firm's *ex ante* expected profit is

$$\mathbb{E}_{s_1} \left\{ \frac{1}{2}(\lambda^B - \lambda^o)(\hat{\alpha}_0 - s_1) | \hat{\alpha}_0 \right\} = 0$$

(recall $\mathbb{E}_{s_1}\{s_1 | \hat{\alpha}_0\} = 0$). To summarize:

Proposition 1. *Consider the career-concerns model without effort. If one firm suffers from the base-rate fallacy to a greater extent than another (*i.e.*, has a lower λ), then that firm will lose money on average (unless its rival is Bayesian) and its rival will make money on average. If the firms suffer from the base-rate fallacy to an equal degree (*i.e.*, have the same λ s), then both firms expect to breakeven.*

It is worth noting that a firm that suffers from the base-rate fallacy is in the *least* danger of losing money if its rival is Bayesian and in the most danger (faces greatest expected losses) when its rival suffers almost as badly as it does from that fallacy. In other words, if λ is the parameter for the firm in question, $\lambda < \lambda^B$, then its expected profits are falling as λ^o decreases within the interval $(\lambda, \lambda^B]$.

3.2 THE MODEL WITH EFFORT

To begin an examination of the model with effort, it is worth supposing that both firms have a common λ (which may vary from λ^B) and that this is commonly understood.

The value the firms place on the executive for the second period is given by (7). The outcome of bidding for his services means that will be his wage.

Hence, in choosing his period-one effort, the executive seeks to maximize⁷

$$\hat{\alpha}_1(\lambda) - c(e_1) = \lambda \hat{\alpha}_0 + (1 - \lambda)(\alpha + \varepsilon_1 + e_1 - \hat{e}_1) - c(e_1). \quad (8)$$

Assumptions made earlier about $c(\cdot)$ ensure a unique interior maximum exists and, moreover, that the first-order condition,

$$(1 - \lambda) - c'(e_1) = 0, \quad (9)$$

is necessary and sufficient. Denote the solution by $e^*(\lambda, \lambda)$ (the reason for repeating the argument will become clear later). Via usual comparative statics, the following is readily proved:

Lemma 1. *Assume a common updating rule by firms (i.e., a common λ), then the lower the weight placed on the prior estimate (i.e., the lower is λ), the greater the effort supplied by the executive in the first period (i.e., $de^*(\lambda, \lambda)/d\lambda < 0$).*

Intuitively, because the executive's effort is, effectively, an attempt to boost (jam) the signal, the more weight placed on the signal, the greater the executive's incentive to boost the signal (i.e., to supply effort).

This analysis also establishes that the executive plays a pure strategy in equilibrium. Given that the game is commonly understood—in particular, all players know the updating rule being employed even if it is inconsistent with Bayes Law—the players must correctly anticipate others' strategies in equilibrium. Hence, $\hat{e}_1 = e^*(\lambda, \lambda)$. In other words, in equilibrium, the executive's efforts to boost the signal are wholly anticipated on the equilibrium path. Given *accurate* anticipation, observe, from (7), that the executive's actions don't actually influence the estimate of his ability. This might lead one to wonder why the executive then bothers to supply effort; the answer is that were he to so deviate, then he would suffer because the firms would still subtract \hat{e}_1 in estimating his ability; his second-period wage would be lower than he desires.⁸

Further observe that expected value (gross of compensation) of the executive in the first period is

$$\hat{\alpha}_0 + \hat{e}_1 = \hat{\alpha}_0 + e^*(\lambda, \lambda).$$

The question of first-period compensation will be addressed later. A corollary of Lemma 1 is, therefore,

Corollary 1. *Assume a common updating rule by firms (i.e., a common λ), then the lower the weight placed on the prior estimate (i.e., the lower is λ), the greater the expected profit (gross of executive compensation) of the executive's first-period employer.*

⁷For convenience and without loss, set the intertemporal discount factor to one (i.e., ignore discounting).

⁸Hermalin and Weisbach (forthcoming) refer to this as the Red Queen effect, the idea that the executive has to run quickly just to stay in place.

A related point is the following. The welfare-maximizing level of first-period effort maximizes $e_1 - c(e_1)$ given that marginal social return to the executive's effort is one. Hence, a second corollary of Lemma 1 is

Corollary 2. *Assume a common updating rule by firms (i.e., a common λ), then the lower the weight placed on the prior estimate (i.e., the lower is λ), the closer the executive's first-period effort is to the welfare-maximizing level. In particular, the more the firms suffer from the base-rate fallacy, the greater will be welfare.*

Now consider the situation in which $\lambda^1 \neq \lambda^2$. A critical issue is what understanding do the firms have of the situation and what is the executive's? In what follows, assume that the executive is sophisticated in the sense that he knows (i) what λ^1 and what λ^2 are and (ii) the beliefs of each firm about the other firm's λ . For the firms, two assumptions will be entertained:

1. *Naïve firms:* each firm believes (a) the other firm uses the same λ (updates as it does) and (b) the executive believes the firms are both using that λ ; or
2. *Stubborn firms:* each firm knows its rival's λ , but that does not affect its own λ , and, moreover, it believes the executive knows both λ s.

The stubborn-firms assumption is arguably difficult to justify insofar as it begs the question, if a firm knows its rival is not updating as it is and presumably "rational" updating is desirable, why then doesn't the firm in question rethink its own updating process? On the other hand, to the extent that "updating" is a shorthand for various personnel policies and procedures, it may be possible both that firms differ in those and understand that they differ.

Naïve firms. From expression (7), a firm's value for the executive, given its observation of s_1 , is

$$\hat{\alpha}_1(\lambda) = \lambda \hat{\alpha}_0 + (1 - \lambda) \underbrace{(\alpha + \varepsilon_1 + e_1)}_{s_1} - e^*(\lambda, \lambda), \quad (10)$$

taking into account it believes the executive took action $e^*(\lambda, \lambda)$. Let $\lambda^h > \lambda^\ell$ denote the weights the two firms put on the prior estimate of ability. Given a second-price auction, the firm that values the executive more wins his services and pays him the value the losing firm assigned. The firm that places less weight on the prior (the λ^ℓ firm) wins when $\hat{\alpha}_1(\lambda^\ell) > \hat{\alpha}_1(\lambda^h)$; that is, when

$$\begin{aligned} \lambda^\ell \hat{\alpha}_0 + (1 - \lambda^\ell)(s_1 - e^*(\lambda^\ell, \lambda^\ell)) &> \lambda^h \hat{\alpha}_0 + (1 - \lambda^h)(s_1 - e^*(\lambda^h, \lambda^h)) \iff \\ (\lambda^h - \lambda^\ell)y &> (1 - \lambda^\ell)e^*(\lambda^\ell, \lambda^\ell) - (1 - \lambda^h)e^*(\lambda^h, \lambda^h) - (\lambda^h - \lambda^\ell)e_1, \end{aligned} \quad (11)$$

where

$$y \sim N\left(0, \frac{1}{\tau_0} + \frac{1}{\eta}\right) \equiv N\left(0, \frac{1}{H}\right) \quad (12)$$

(that is, $y = \alpha + \varepsilon_1 - \hat{\alpha}_0$). Note the implicit definition of the variance of y , $1/H$. Observe that (10) can be rewritten as

$$\hat{\alpha}_1(\lambda) = \hat{\alpha}_0 + (1 - \lambda)(y + e_1 - e^*(\lambda, \lambda)) \quad (10')$$

Define

$$\Delta = \frac{(1 - \lambda^\ell)e^*(\lambda^\ell, \lambda^\ell) - (1 - \lambda^h)e^*(\lambda^h, \lambda^h)}{\lambda^h - \lambda^\ell}.$$

From Lemma 1, $\Delta > 0$. Moreover, condition (11) can be rewritten as

$$y > \Delta - e_1. \quad (13)$$

Let Φ and ϕ denote, respectively, the distribution and density functions of the standard normal (*i.e.*, $N(0, 1)$). Using well-known methods,⁹ it follows that

$$\Pr\{y > \Delta - e_1\} = 1 - \Phi((\Delta - e_1)\sqrt{H}) = \Phi((e_1 - \Delta)\sqrt{H}).$$

The executive's expected utility is $\mathbb{E}\{\min\{\hat{\alpha}_1(\lambda^\ell), \hat{\alpha}_1(\lambda^h)\}\} - c(e_1)$ given his period 2 compensation is the losing bidder's value for him. His expected utility can be rewritten as

$$\begin{aligned} & \hat{\alpha}_0 + \int_{-\infty}^{\infty} \min\left\{(1 - \lambda^\ell)(y + e_1 - e^*(\lambda^\ell, \lambda^\ell)), \right. \\ & \quad \left. (1 - \lambda^h)(y + e_1 - e^*(\lambda^h, \lambda^h))\right\} \sqrt{H} \phi(y\sqrt{H}) dy - c(e_1) \\ &= \hat{\alpha}_0 + \int_{-\infty}^{\Delta - e_1} (1 - \lambda^\ell)(y + e_1 - e^*(\lambda^\ell, \lambda^\ell)) \sqrt{H} \phi(y\sqrt{H}) dy \\ & \quad + \int_{\Delta - e_1}^{\infty} (1 - \lambda^h)(y + e_1 - e^*(\lambda^h, \lambda^h)) \sqrt{H} \phi(y\sqrt{H}) dy - c(e_1) \quad (14) \end{aligned}$$

The executive chooses his first-period effort to maximize (14). The corresponding first-order condition is

$$\begin{aligned} & \left((1 - \lambda^h)(\Delta - e^*(\lambda^h, \lambda^h)) - (1 - \lambda^\ell)(\Delta - e^*(\lambda^\ell, \lambda^\ell))\right) \sqrt{H} \phi((\Delta - e_1)\sqrt{H}) \\ & + (1 - \lambda^\ell) \Phi((\Delta - e_1)\sqrt{H}) + (1 - \lambda^h) \Phi((e_1 - \Delta)\sqrt{H}) - c'(e_1) = 0. \quad (15) \end{aligned}$$

⁹In particular, if $\xi \sim N(\mu, \sigma^2)$, then the distribution and density functions for ξ can be expressed, respectively, as

$$\Phi\left(\frac{\xi - \mu}{\sigma}\right) \quad \text{and} \quad \frac{1}{\sigma} \phi\left(\frac{\xi - \mu}{\sigma}\right).$$

Because the standard normal is symmetric around 0, the mass to the left of $-\xi$ equals the mass to the right of ξ ; hence, $1 - \Phi(\xi) = \Phi(-\xi)$.

Using the definition of Δ , simple algebra reveals that the top line of (15) is zero; hence, the first-order condition can be rewritten as

$$(1 - \lambda^\ell)\Phi\left((\Delta - e_1)\sqrt{H}\right) + (1 - \lambda^h)\Phi\left((e_1 - \Delta)\sqrt{H}\right) - c'(e_1) = 0. \quad (15')$$

Let $e^*(\lambda^\ell, \lambda^h)$ denote the solution to (15'). Because

$$\Phi\left((\Delta - e_1)\sqrt{H}\right) + \Phi\left((e_1 - \Delta)\sqrt{H}\right) = 1.$$

It follows from (15') that $c'(e^*(\lambda^\ell, \lambda^h))$ equals a weighted average of $1 - \lambda^\ell$ and $1 - \lambda^h$. Given the strict convexity of $c(\cdot)$, it must therefore be that

$$e^*(\lambda^h, \lambda^h) < e^*(\lambda^\ell, \lambda^h) < e^*(\lambda^\ell, \lambda^\ell). \quad (16)$$

This establishes

Proposition 2. *Under the naïve-firms assumption, the executive chooses, in equilibrium, a level of effort that is greater than that anticipated by the firm that suffers less from the base-rate fallacy (i.e., the λ^h firm), but that is less than the level of effort anticipated by the firm that suffers more from the base-rate fallacy (the λ^ℓ firm).*

For which firm is the executive more likely to work in the second period? Given the mean and median of a normal distribution coincide and $\mathbb{E}y = 0$, the answer depends on whether the righthand side of (13) is positive or negative: if positive, then the λ^h firm is the more likely second-period employer; if negative, then the λ^ℓ firm is the more likely second-period employer.

Proposition 3. *Under the naïve-firms assumption, in the second period, the executive is more likely to be employed by the firm that suffers less from the base-rate fallacy (i.e., the λ^h firm) than by the firm that suffers more from the base-rate fallacy (the λ^ℓ firm).*

In terms of intuition, recall that the λ^ℓ firm expects more work from the executive than does the λ^h firm and more than the executive actually supplies. Hence, the λ^ℓ firm constructs a signal, its \tilde{s} , that is biased downward. In contrast, the λ^h firm constructs a signal that is biased upward. Were this biasing not to occur, then, from earlier analysis, it follows that the λ^ℓ firm would outbid its rival when the signal exceeds $\hat{\alpha}_0$ (equivalently, when $y > 0$) and underbid when the signal was less than $\hat{\alpha}_0$ (equivalently, when $y < 0$). Because the unbiased signal is symmetrically distributed about $\hat{\alpha}_0$ (equivalently, y around 0), those two events are equally likely. Because of the biasing, the λ^ℓ firm outbids its rival only when the signal is measurably greater than $\hat{\alpha}_0$ (i.e., when $y > \Delta - e^*(\lambda^\ell, \lambda^h) > 0$); this, then, means the λ^ℓ firm outbids its rival less than half the time.

Turning to the issue of employment in the first period. A firm with weighting factor λ values the executive at $\hat{\alpha}_0 + e^*(\lambda, \lambda)$ given its belief he will provide effort

$e^*(\lambda, \lambda)$. Lemma 1 entails $e^*(\lambda^\ell, \lambda^\ell) > e^*(\lambda^h, \lambda^h)$; hence, the λ^ℓ firm will outbid the λ^h firm. It will hire him at a salary of $\hat{\alpha}_0 + e^*(\lambda^h, \lambda^h)$. Consequently, in *expectation*, the λ^ℓ firm will make a profit of

$$e^*(\lambda^\ell, \lambda^h) - e^*(\lambda^h, \lambda^h) > 0, \quad (17)$$

where the inequality follows from Proposition 2. This establishes

Proposition 4. *Under the naïve-firms assumption, in the first period, the executive is employed by the firm that suffers more from the base-rate fallacy (i.e., the λ^ℓ firm). That firm earns a positive expected profit in the first period (the amount in expression (17)).*

What about overall profits? In light of Proposition 3, the λ^ℓ firm has negative expected profit in the second period (the reasoning is similar to that behind Proposition 1). It makes an expected profit in the first period. Across both periods, is its expected profit positive or negative? What about the λ^h firm? To explore these questions, consider the following example: $\hat{\alpha}_0 = 0$, $\tau_0 = \eta = 2$ (so $\lambda^B = 1/2$ and $H = 1$), and $c(e) = e^2/2$ (so $e^*(\lambda, \lambda) = 1 - \lambda$). It is readily shown that

$$\Delta = \frac{(1 - \lambda^\ell)^2 - (1 - \lambda^h)^2}{\lambda^h - \lambda^\ell}.$$

The true *expected* second-period profit of a firm that successfully bids for the executive's services, conditional on the signal, is

$$\begin{aligned} & (1 - \lambda^B) \underbrace{\left(\alpha + \varepsilon_1 \right)}_y - \underbrace{(1 - \lambda^o) \left(\underbrace{\alpha + \varepsilon_1}_y + e^*(\lambda^\ell, \lambda^h) - e^*(\lambda^o, \lambda^o) \right)}_{\hat{\alpha}_1(\lambda^o)} \\ &= (\lambda^o - \lambda^B)y - (1 - \lambda^o)(e^*(\lambda^\ell, \lambda^h) - e^*(\lambda^o, \lambda^o)) \equiv \mathbb{E}V_2(\lambda^B, \lambda^o, y), \end{aligned} \quad (18)$$

given $\hat{\alpha}_0 = 0$, where o denotes, as before, the other (losing) firm. Because the cutoffs of who wins or loses the bidding for the executive's second-period services don't depend on the realization of y , iterative expectations permits calculating the expected profits of the firms from the second period as

$$\begin{aligned} \mathbb{E}V^\ell &= \int_{\Delta - e^*(\lambda^\ell, \lambda^h)}^{\infty} \mathbb{E}V_2\left(\frac{1}{2}, \lambda^h, y\right) \phi(y) dy \quad \text{and} \\ \mathbb{E}V^h &= \int_{-\infty}^{\Delta - e^*(\lambda^\ell, \lambda^h)} \mathbb{E}V_2\left(\frac{1}{2}, \lambda^\ell, y\right) \phi(y) dy, \end{aligned}$$

where use has been made of the fact that $H = 1$ and $\lambda^B = 1/2$ in this example.

The λ^ℓ firm can have either positive or negative expected *lifetime* profit. This is readily shown by fixing $\lambda^\ell = 1/4$ and considering $\lambda^h = 7/24$ or $5/12$. Table 1 provides the relevant values.¹⁰ Although, as noted, the λ^ℓ firm faces

λ^h	7/24	5/12
$e^*(1/4, \lambda^h)$.740	.706
$\mathbb{E}V^\ell$	-0.070	-0.046
$\mathbb{E}V^h$.083	.106
Expression (17)	.032	.122
Lifetime profit λ^ℓ firm	-0.038	.076

Table 1: How the Firms Do in the Example for Naïve Firms

expected losses in the second period, it may gain enough in the first (expression (17)) to offset those losses.

In Table 1, the λ^ℓ firm, the one that suffers more from the base-rate fallacy, does better the more Bayesian its rival (*i.e.*, the closer λ^h is to λ^B). This reflects two effects: one is the effect documented in Proposition 1 (so $\mathbb{E}V^\ell$ is greater); the second is that the greater is λ^h , the less effort the λ^h firm expects of the executive in the first period, so the less it bids for him, which lowers the salary the λ^ℓ firm must pay. Of course, the actual effort the executive supplies, $e^*(\lambda^\ell, \lambda^h)$ is also falling in λ^h , but that effect is smaller than the effect on salary.

Table 1 and connected analysis suggest that being a “bad Bayesian” need not prove fatal, insofar as even when competing with more Bayesian firms, a less Bayesian firm can still make a profit. Indeed, somewhat paradoxically, a firm that is a bad Bayesian does better the more Bayesian its rival is.¹¹

Stubborn firms. Recall that the stubborn-firms assumption is that although each firm updates using its weighting factor, λ^ℓ or λ^h , which differs from the Bayesian factor, it nonetheless knows that (i) its weighting factor differs from its rival; (ii) its rival’s weighting factor; and (iii) that the executive knows both firms’ weighting factors. From expression (7), a firm’s value for the executive after the first period is

$$\hat{\alpha}_1(\lambda) = \lambda \hat{\alpha}_0 + (1 - \lambda)(\alpha + \varepsilon_1 + e_1 - e^*(\lambda^\ell, \lambda^h)). \quad (19)$$

Recall that the λ^ℓ firm wins the bidding for the executive’s services in the second period if and only if $\hat{\alpha}_1(\lambda^\ell) > \hat{\alpha}_1(\lambda^h)$; that is, when

$$\begin{aligned} (1 - \lambda^\ell)y - \hat{\alpha}_0 + (1 - \lambda^\ell)(e_1 - e^*(\lambda^\ell, \lambda^h)) &> (1 - \lambda^h)y - \hat{\alpha}_0 + (1 - \lambda^h)(e_1 - e^*(\lambda^\ell, \lambda^h)) \\ \iff (\lambda^h - \lambda^\ell)y &> (\lambda^h - \lambda^\ell)(e^*(\lambda^\ell, \lambda^h) - e_1); \end{aligned}$$

¹⁰Mathematica program used for calculations available from the author upon request.

¹¹For this example, if $\lambda^h = \lambda^B$, then the λ^ℓ firm’s lifetime expected profit would be 0.154. It is important that being a better Bayesian in this context does not mean the more Bayesian firm isn’t naïve—the assumption remains that it is naïve insofar as it acts as if both firms using the weighting factor λ^B .

hence, the λ^ℓ firm wins the bidding if $y > e^*(\lambda^\ell, \lambda^h) - e_1$. The same logic that led to expression (14) entails that the executive's expected utility is

$$\begin{aligned} \hat{\alpha}_0 + \int_{-\infty}^{\infty} \min \left\{ (1 - \lambda^\ell)(y + e_1 - e^*(\lambda^\ell, \lambda^h)), \right. \\ \left. (1 - \lambda^\ell)(y + e_1 - e^*(\lambda^\ell, \lambda^h)) \right\} \sqrt{H} \phi(y\sqrt{H}) dy - c(e_1) \\ = \hat{\alpha}_0 + \int_{-\infty}^{e^*(\lambda^\ell, \lambda^h) - e_1} (1 - \lambda^\ell)(y + e_1 - e^*(\lambda^\ell, \lambda^h)) \sqrt{H} \phi(y\sqrt{H}) dy \\ + \int_{e^*(\lambda^\ell, \lambda^h) - e_1}^{\infty} (1 - \lambda^h)(y + e_1 - e^*(\lambda^\ell, \lambda^h)) \sqrt{H} \phi(y\sqrt{H}) dy - c(e_1). \quad (20) \end{aligned}$$

The first-order condition for maximizing (20) with respect to e_1 is readily seen to be

$$\begin{aligned} (1 - \lambda^\ell) \Phi(e^*(\lambda^\ell, \lambda^h) - e_1) \sqrt{H} \\ + (1 - \lambda^h) \Phi(e_1 - e^*(\lambda^\ell, \lambda^h)) \sqrt{H} - c'(e_1) = 0. \quad (21) \end{aligned}$$

Lemma 2. *For any $e^*(\lambda^\ell, \lambda^h)$, expression (20) is globally concave in e_1 .*

In light of Lemma 2, the solution to (21) is unique and defines a global maximum.

In equilibrium, under the stubborn-firms assumption, the solution to (21) must be $e^*(\lambda^\ell, \lambda^h)$; that is, the firms must correctly forecast the executive's choice of effort. Making that substitution yields

$$(1 - \lambda^\ell) \frac{1}{2} + (1 - \lambda^h) \frac{1}{2} - c'(e^*(\lambda^\ell, \lambda^h)) = 0. \quad (22)$$

As with (15'), $e^*(\lambda^\ell, \lambda^h)$ is determined by a weighted average of $1 - \lambda^\ell$ and $1 - \lambda^h$; hence, (16) continues to hold—although, note, the $e^*(\lambda^\ell, \lambda^h)$ with naïve firms is different than with stubborn firms. In fact, the former is larger given that $\Delta - e_n^*(\lambda^\ell, \lambda^h) > 0$ (the subscript n for naïve).¹²

$$\Phi((\Delta - e_n^*(\lambda^\ell, \lambda^h))\sqrt{H}) > \Phi(0) = \frac{1}{2};$$

hence, with naïve firms, more weight is put on $1 - \lambda^\ell$ than with stubborn firms. To summarize:

Proposition 5. *Under the stubborn-firms assumption, there exists an equilibrium in which the executive chooses a first-period effort of $e_s^*(\lambda^\ell, \lambda^h)$ (the subscript s for stubborn), which satisfies*

$$e^*(\lambda^\ell, \lambda^\ell) > e_n^*(\lambda^\ell, \lambda^h) > e_s^*(\lambda^\ell, \lambda^h) > e^*(\lambda^h, \lambda^h).$$

¹²That $\Delta - e_n^*(\lambda^\ell, \lambda^h) > 0$ was proved as part of establishing Proposition 3.

The intuition behind this Proposition is that because the executive knows he is more likely to work for the λ^ℓ firm in the stubborn-firms setting than in the naïve-firms setting, he knows his wage is more often a function of $1 - \lambda^h$ because the λ^h firm is more likely to be the losing bidder and his compensation is set by the losing bidder's valuation for him. Given $1 - \lambda^h < 1 - \lambda^\ell$, this means he has less incentive to exert effort (jam the signal) the more likely it is that the λ^ℓ firm will be his employer.

As just noted, in the stubborn-firms setting, the λ^ℓ firm is more likely to employ the executive in the second period than it would be in the naïve-firms setting. In fact now, as should be clear from the analysis, with stubborn firms, the λ^ℓ firm employs the executive if and only if $y > 0$. The situation is similar to Proposition 1. Moreover, because, now, both firms hold the same expectation of the executive's first-period value, bidding will result in his first-period salary equaling that value; hence, the firm that employs him in the first period breaks even in expectation. Consequently, the expected lifetime profit of the λ^ℓ firm is negative; unless $\lambda^h = \lambda^B$, in which case it is zero (Proposition 1). The expected lifetime profit of the λ^h firm is positive. To summarize:

Proposition 6. *Under the stubborn-firms assumption, the firm employing the executive in the first period earns an expected profit of zero. If the firm that suffers more from the base-rate fallacy (i.e., the λ^ℓ firm) employs the executive in the second period, it will lose money in expectation, unless its rival is perfectly Bayesian (i.e., $\lambda^h = \lambda^B$). If the firm that suffers less from the base-rate fallacy (i.e., the λ^h firm) employs the executive in the second period, it will make money in expectation.*

4 EMPLOYMENT DECISIONS

As discussed in Hermalin and Weisbach (forthcoming), another application of learning (assessment) models has concerned the question of retaining or replacing an executive. In most of these models, there is a single firm that seeks to assess the ability of its executive. If his ability is judged to fall below the expected value of a replacement, he is fired; otherwise, he retains his position.

Assume the following timing:

1. The firm hires an executive from a large pool of *ex ante* identical executives. The ability, α , of any one executive is an independent draw from $N(\hat{\alpha}_0, 1/\tau_0)$.
2. The firm (its owner) invests in receiving a signal, s , about the executive. Let $p \in [0, 1]$ denote the investment, where p is the probability that it observes s . Let $\gamma(p)$ denote the cost to the firm.
3. The executive chooses effort e that affects the signal, as detailed previously, and that may have an effect on the firm's ultimate payoff, x (gross of its investment in information and the executive's compensation). The executive does not observe p (but, of course, infers it correctly in equilibrium). As in Section 3, his action is hidden from the firm (its owner).

4. The firm observes the signal s with probability p . If it receives the signal, it forms a revised estimate of the executive's ability. Its revised estimate is given by (7). Based on this revised estimate, the firm keeps or fires the executive. A replacement comes from the same pool from which the original executive was drawn. Dismissing the executive costs the firm f , an exogenously set firing cost.
5. The firm realizes $x = \alpha + \beta(e)$, where α is the ability of the executive in place at this stage (original or replacement) and e is the original executive's effort (*i.e.*, his effort may be some form of investment). The properties of $\beta : \mathbb{R}_+ \rightarrow \mathbb{R}$ will be considered later.

Additionally, assume that any executive has a reservation utility of 0 and is protected by limited liability insofar as the firm cannot collect any payment from an executive (*i.e.*, compensation must be non-negative). Assume that the executive in place at the end enjoys a control benefit $b > 0$. Assume that firing the executive in stage 4 does not excuse the firm from paying the salary promised the originally hired executive. It should be noted that these assumptions are fairly standard in this literature (see Hermalin and Weisbach, forthcoming; also Hermalin, 2005, and Hermalin and Weisbach, 1998, 2012). Last, restrict attention to the case in which the executive knows the firm's (owner's) weighting factor, λ .

Given a large pool of executives from which to choose, the firm has all the bargaining power and can be presumed to make take-it-or-leave-it (TIOLI) offers to executives. Hence, an executive receives the minimum compensation consistent with his reservation utility and limited liability. For a replacement executive, this bargaining results in a salary of zero.¹³ The expected value of a replacement executive, including his compensation, is thus $\hat{\alpha}_0$.

If the firm fails to observe s , then it cannot update its beliefs. It believes, therefore, the originally hired executive to be as able in expectation as any replacement. His value at this point, recognizing that any salary promised him is, at this point sunk, is $\hat{\alpha}_0$. Given the firing cost f , the firm would never replace the executive in this case.

If the firm does observe s , then it will dismiss the original executive if and only if

$$\hat{\alpha}_1(\lambda) < \hat{\alpha}_0 - f. \quad (23)$$

Using (7), this means it will dismiss the original executive if and only if

$$y + e - \hat{e} < -\frac{f}{1 - \lambda}, \quad (24)$$

where y is as before (see expression (12)) and \hat{e} is the level of effort expected by the firm. The righthand side of (24) is decreasing in λ , which establishes:

¹³Of course, a literally zero salary is unrealistic. This should simply be understood as a normalization.

Lemma 3. *Ceteris paribus, the more a firm suffers from the base-rate fallacy (i.e., the lower is λ), the more likely it is to fire its originally hired executive.*

For future reference, define

$$F(\lambda) = -\frac{f}{1-\lambda}.$$

4.1 THE EXECUTIVE'S BEHAVIOR

The original executive chooses his effort to maximize his expected utility:

$$\begin{aligned} (1-p)b + pb \left(1 - \Phi \left((F(\lambda) + \hat{e} - e) \sqrt{H} \right) \right) - c(e) \\ = (1-p)b + pb \Phi \left((e - \hat{e} - F(\lambda)) \sqrt{H} \right) - c(e). \end{aligned} \quad (25)$$

The corresponding first-order condition is

$$pb\phi \left((e - \hat{e} - F(\lambda)) \sqrt{H} \right) \sqrt{H} - c'(e) = 0.$$

In equilibrium, $\hat{e} = e$; hence, his equilibrium choice of effort must satisfy¹⁴

$$pb\phi \left(\frac{f}{1-\lambda} \sqrt{H} \right) \sqrt{H} = c'(e^E(\lambda, p)). \quad (26)$$

Recall that $\phi(\xi)$ decreases in ξ for $\xi > 0$. Consequently, the greater is λ , the less will be the lefthand side of (26). Given the convexity of $c(\cdot)$, this helps to establish:

Lemma 4. *Holding p fixed, the original executive's effort in equilibrium is greater the more the firm suffers from the base-rate fallacy (i.e., the lower is λ). Holding λ fixed, his equilibrium effort is also greater the more likely it is that the firm will receive the signal (i.e., the greater is p).*

The second half of Lemma 4 is immediate given the convexity of $c(\cdot)$.

The original executive's compensation must satisfy both his participation constraint and limited liability. The former, in equilibrium, is

$$\underbrace{w + (1-p)b + pb\Phi \left(\frac{f}{1-\lambda} \sqrt{H} \right) - c(e^E(\lambda, p))}_{\mathbf{A}} \geq 0. \quad (27)$$

Because $c(0) = 0$ and $\Phi > 0$, it is evident that the term labeled **A** must be positive by revealed preference given that the executive could have chosen $e = 0$ in (25). It follows that the constraint (27) is slack (a strict inequality); hence $w \equiv 0$ (i.e., the limited-liability constraint binds). To summarize:

¹⁴The first-order condition (26) is necessary, but without additional assumptions it may not be sufficient, as noted by Hermalin (2005). However, as that article shows, reasonable assumptions exist that ensure (26) is sufficient. Going forward, it should be assumed that (26) is sufficient.

Lemma 5. *Regardless of the degree to which it suffers from the base-rate fallacy (i.e., λ) or its equilibrium investment in information (i.e., p), the original executive's compensation is zero in equilibrium.*

4.2 THE FIRM'S INVESTMENT IN INFORMATION

If the firm dismisses the executive, its expected profit is $\hat{\alpha}_0 - f$. If it retains him its expected profit is

$$\hat{\alpha}_1(\lambda) = \hat{\alpha}_0 + (1 - \lambda)y, \quad (28)$$

where the definition of y and the fact that $e = e^E(\lambda, p)$ in equilibrium have been used to rewrite (7). Hence, *ex ante*, before the firm observes the signal, its expected value (ignoring β) is

$$\begin{aligned} (1-p)\hat{\alpha}_0 + p \times & \left((\hat{\alpha}_0 - f)\Phi(F(\lambda)\sqrt{H}) + \int_{F(\lambda)}^{\infty} (\hat{\alpha}_0 + (1 - \lambda)y) \phi(y\sqrt{H})\sqrt{H} dy \right) - \gamma(p) \\ & = \hat{\alpha}_0 + p \times \left((1 - \lambda)\phi(F(\lambda)\sqrt{H}) - f\Phi(F(\lambda)\sqrt{H}) \right) - \gamma(p), \end{aligned} \quad (29)$$

where the second line follows because $\phi(\xi) = \exp(-\xi^2/2)/\sqrt{2\pi}$, so

$$\begin{aligned} \int_{F(\lambda)}^{\infty} y\phi(y\sqrt{H})\sqrt{H} dy &= \int_{F(\lambda)}^{\infty} y\sqrt{H} \exp\left(-\frac{y^2 H}{2}\right) \frac{1}{\sqrt{2\pi}} dy \\ &= -\exp\left(-\frac{y^2 H}{2}\right) \frac{1}{\sqrt{2\pi}} \Big|_{F(\lambda)}^{\infty} = \phi(F(\lambda)\sqrt{H}). \end{aligned}$$

Differentiating the firm's perceived marginal return to gaining information—the term in large parentheses in the second line of expression (29)—with respect to λ yields, using the definition of $F(\lambda)$ to cancel like terms,

$$-\phi(F(\lambda)\sqrt{H}) < 0. \quad (30)$$

Given the sign of (30), it follows that an increase in λ reduces a firm's marginal return to information (as it perceives that return). This, in turn, means a firm that suffers less from the base-rate fallacy (a higher λ firm) will invest less than a firm that suffers more (at least assuming interior solutions to the problem of maximizing (29) with respect to p).¹⁵ To summarize:

Proposition 7. *Maintain assumptions so that (29) has an interior solution (see footnote 15). Then the more the firm suffers from the base-rate fallacy (i.e., the lower is λ), the more it invests in getting a signal about its executive (i.e., the higher will be p). Consequently, an executive employed by a lower- λ firm works harder than an executive employed by a higher- λ firm.*

¹⁵An interior solution is guaranteed if, taking $\gamma(\cdot)$ to be twice differentiable, $\gamma'(0) = 0$, $\lim_{p \rightarrow 1} \gamma'(p) \geq \phi(0)$, $\gamma''(p) > 0$ for all $p \in [0, 1]$, and the marginal return (the expression in parentheses that p multiplies in (29)) is positive. That it is positive is demonstrated in Lemma A.1 in the Appendix.

The “consequently” part of the proposition follows from Lemma 4.

Whether or not the firm wishes the executive to work harder depends on the benefit function, $\beta(\cdot)$. If it is increasing, more work from the executive benefits the firm; if it is decreasing, it harms the firm. It might seem natural to assume that executive effort is beneficial, but there is a significant literature that notes the possibility of signal-jamming effort harming the firm (its shareholders). One example of this when the effort is “myopic”—any short-run benefit is outweighed by long-run cost (see, in particular, Stein, 1989; Stein, 2003, and Hermalin and Weisbach, forthcoming, provide short surveys of this literature).

Corollary 3. *If managerial effort is myopic or otherwise non-beneficial to the firm (i.e., $\beta(\cdot)$ is a decreasing function), then a firm suffers more from such effort the more it suffers from the base-rate fallacy.*

On the other hand, it could be that managerial effort is beneficial (as in Section 3). In that case:

Corollary 4. *If managerial effort is beneficial to the firm (i.e., $\beta(\cdot)$ is an increasing function), then a firm benefits more from such effort the more it suffers from the base-rate fallacy.*

It is important to understand that Corollary 4 does *not* necessarily entail that, when $\beta(\cdot)$ is an increasing function, it behooves a firm to suffer from the base-rate fallacy. The reason for this is two fold: first, recall, a fallacy-suffering firm over-invests in gaining information relative to what a Bayesian firm would do; and, second, it fires the original executive when a Bayesian firm would not; that is, it fires the executive too often and in a non-profit-maximizing way (relative to Bayesian updating). To see this second point, recall that the true value of the executive after observing the signal is $\hat{\alpha}_1(\lambda^B)$; hence, from (28),

$$\hat{\alpha}_0 + (1 - \lambda^B)y.$$

Substituting that into (29) appropriately, the true value of the firm is

$$\hat{\alpha}_0 + p \times \left((1 - \lambda^B)\phi(F(\lambda)\sqrt{H}) - f\Phi(F(\lambda)\sqrt{H}) \right) - \gamma(p). \quad (31)$$

Differentiating (31) with respect to λ , holding p fixed, yields a derivative that has the same sign as

$$\phi(F(\lambda)\sqrt{H}) \underbrace{H}_{<0} \underbrace{F'(\lambda)(F(\lambda^B) - F(\lambda))}_{<0} > 0. \quad (32)$$

Expression (32) confirms that, holding all else equal, the firm’s expected profit increases as its weighting factor approaches the Bayesian value (i.e., as $\lambda \uparrow \lambda^B$).

To get a sense of how all these factors might affect the firm’s expected value, extend the example of Table 1 by assuming that $b = 1/4$, $f = 1/10$, $\gamma(p) = p^2/2$,

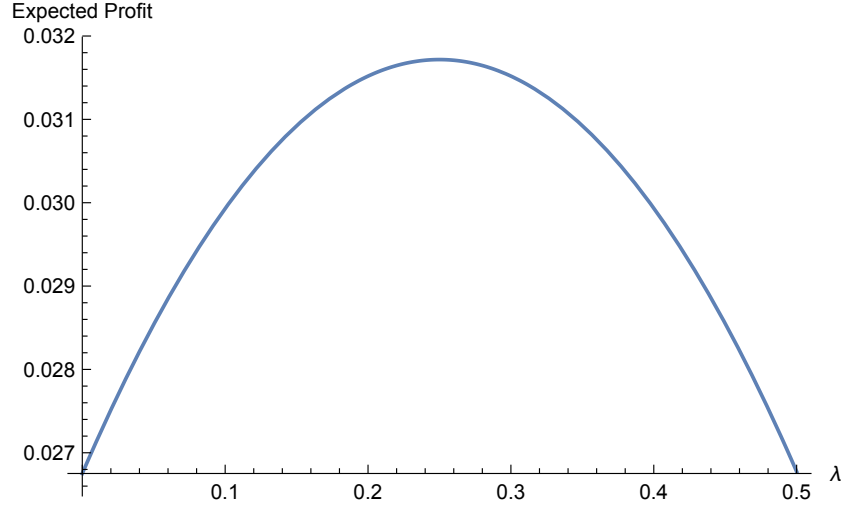


Figure 1: A plot of expected firm value as a function of the weighting factor, λ , that it uses. The Bayesian weighting factor for this example is $1/2$.

and $\beta(e) = e$. Note in this example, the firm benefits from managerial effort. It can be shown, for this example, that

$$e^E(\lambda, p) = p \frac{1}{4} \phi \left(-\frac{1}{10(1-\lambda)} \right) \quad \text{and}$$

$$p^*(\lambda) = (1-\lambda) \phi \left(-\frac{1}{10(1-\lambda)} \right) - \frac{1}{10} \Phi \left(-\frac{1}{10(1-\lambda)} \right),$$

where $p^*(\lambda)$ maximizes (29). Substituting those values into (31) (recalling that $\lambda^B = 1/2$ and $\hat{\alpha}_0 = 0$) and adding

$$\beta(e^E(\lambda, p^*(\lambda))) = e^E(\lambda, p^*(\lambda)),$$

yields expected firm value as a function of λ . This is plotted in Figure 1.¹⁶

In Figure 1, there is some “balance” between the benefit that a low- λ firm gets from greater managerial effort and the costs it incurs from over-investing in information and firing the executive too often. Consequently, the ideal weighting factor in this example is less than the Bayesian factor (the ideal factor can be shown to equal $1/4$, as the figure suggests). This is by no means a universal result: if $\beta(e) = 0$ (*i.e.*, the executive’s effort has no effect on the firm’s payoff), then, consistent with analysis above expected value would be maximized by $\lambda = \lambda^B$. On the other hand, if $\beta(e) = \kappa e$, $\kappa > 0$, and leaving b and f unspecified

¹⁶The Mathematica program used to plot Figure 1 is available from the author upon request.

(but such that all optimization programs yield interior solutions) then it can be shown, *for this example*, that the weighting factor that maximizes expected firm value is

$$\lambda^* = \frac{1}{2}(1 - 2b\kappa) < \frac{1}{2} = \lambda^B. \quad (33)$$

See the Lemma A.2 in the Appendix for more details. To summarize:

Proposition 8. *Conditions exist such that a firm does better in expectation if it suffers to a limited degree from the base-rate fallacy than if it does not (i.e., if its $\lambda < \lambda^B$). The optimal degree of bias can lie between complete bias (disregarding the base rate altogether, $\lambda = 0$) and no bias ($\lambda = \lambda^B$).*

4.3 WHEN COMPENSATION CHANGES WITH EFFORT

In the analysis so far, the original executive's compensation, w , does not depend on his equilibrium effort level. This is a consequence of assuming that the control benefit, b , goes to whomever is in charge at the end (original or replacement executive) and limited liability. To unpack that: if there were no limited liability, then a replacement executive would be paid $-b$; that is, the firm would capture the control benefit back via a negative wage. This, in turn, would give the firm an additional incentive to fire the incumbent—by firing him, it can capture the control benefit. Allowing this complicates the analysis of the firing decision; it was to avoid that complication that the assumption that executives are protected by limited liability was made. The downside of that assumption is the following: as seen above, the *original* executive earning a rent in expectation (Lemma 5). The firm seeks to minimize that rent by setting $w = 0$, but it cannot fully avoid it. Given, though, that $w = 0$ is invariant with respect to the executive's effort, there is no scope in the original analysis to consider a relation between the employer's bias (i.e., degree to which λ departs from λ^B) and executive compensation.

An alternative assumption would be that there is no limited liability, but that only the original executive gets the control benefit and only if he survives to the end.¹⁷ The replacement executive gains no benefit, now, from employment, so holding him to his reservation utility, zero, the firm again pays him zero. The firing decision is, thus, as before. If, however, executives are not protected by limited liability, then the firm will extract the original executive's rent back via a $w < 0$. The actual w will be such that (27) is an equality. It follows, therefore, that if $e^E(\lambda, p)$ is greater, so too must be the executive's compensation. Keeping in mind, too, that the executive does not observe p , only anticipates it in equilibrium, the firm will not be able to reduce compensation by adjusting p ; any promise to do so would be incredible. This means the firm will choose the same p as in the original analysis. Putting all this together, a corollary of Lemma 4 and Proposition 7 is

¹⁷A perhaps more natural way to frame this is he suffers an idiosyncratic cost if dismissed (e.g., a loss of status or due to a need to relocate).

Proposition 9. *Assume, now, that only the original executive gains b if he survives to the end and drop the limited-liability assumption. Then, in equilibrium, the compensation of an executive hired at the initial stage is greater the more his employer suffers from the base-rate fallacy (i.e., the lower is λ).*

In relation to the previous analysis, this becomes a third factor why firms can lose from the base-rate fallacy: the more the firm suffers from the fallacy, the more it must pay its original executive. This does not, however, invalidate Proposition 8. For example, using the same assumptions behind Figure 1, the firm's expected value falls because it must now bear the cost

$$c(e^E(\lambda, p)) = \frac{1}{2} \left(p \frac{1}{4} \phi \left(-\frac{1}{10(1-\lambda)} \right) \right)^2.$$

This, however, has a minor effect: the optimal weighting factor rises from 0.250 in the earlier example to 0.256 for one in which the firm bears this additional cost.

Data on trends in executive compensation and executive tenure show that the former has been rising, while the latter has been falling (see Hermalin, 2005, for a survey of those empirical results). As noted, the probability that the executive loses his job increases the more his employer suffers from the base-rate fallacy (this both because such an employer fires the executive for a greater range of signals and because the employer is more likely to acquire the signal). Hence, if there were a reason to expect that firms (shareholders or boards) were becoming more prone to the base-rate fallacy, then this model would offer a way to tie those two trends together. This is not implausible, at least in the following sense: as has also been true, the composition of boards of directors has trended toward more and more outside directors (i.e., directors from outside the firm). Being outsiders, such directors may hold less precise prior estimates of the executive's ability or what to expect more generally.¹⁸ In terms of expression (1), they have a lower τ_0 than inside directors. Given that even the Bayesian λ is given by

$$\lambda^B = \frac{\tau_0}{\tau_0 + \eta},$$

it follows that having directors with lower τ_0 s is, in essence, equivalent to having directors who suffer more from the base-rate fallacy. This discussion can be summarized as¹⁹

Proposition 10. *Assume rational Bayesian actors, but suppose that an evolution in boards of directors leads to boards with less precise prior estimates of managerial (CEO) ability. Then this same evolution would lead to:*

¹⁸In a different context, Güner et al. (2008) demonstrate that differences in directors' expertise can have important implications for firm behavior.

¹⁹Hermalin (2005) derives a similar result, but via a different channel. In that article, it is assumed that outside directors have the same precision in their estimates as other directors, but that their independence lowers their effective cost of collecting a signal (in the notation of this paper, they have a lower $\gamma(p)$). See Hermalin for details.

- (i) *Greater monitoring of CEOs;*
- (ii) *Shorter average tenures for CEOs (greater firing probabilities); and*
- (iii) *Greater compensation for CEOs.*

5 DISCUSSION AND CONCLUSIONS

As noted at the end of the last section in connection to Proposition 10, an alternative interpretation of at least some of these models is not that the employers (boards, shareholders) suffer from a cognitive bias, but instead differ (perhaps relative to past situations) in the precision of their prior information. If their prior knowledge is less precise, then necessarily more weight is put on current signals. As suggested in connection to Proposition 10, this can be useful for explaining various trends in corporate governance.

The analysis could also arguably be used cross-sectionally or to explain differences across countries. For example, in a new industry, in which no one has much of a track record, priors are going to be imprecise and more weight placed on current signals. All else equal, this suggests that executives in new industries might work harder, face greater scrutiny, and be dismissed with higher probabilities than executives in established industries. Likewise, if in one country the business community is very tight (perhaps because most of its leaders attended the same few universities or institutions), then priors will be precise, executives will correspondingly feel less pressure to work hard, face less scrutiny, and enjoy more secure positions, at least relative to executives in a country with a less tight business community. Of course, if the work that executives do is myopic signal jamming, then firms in old industries or with leaders from tight business communities might have an advantage *vis-à-vis* those in new industries or with leaders from more open business communities.

This notion of Bayesian decision makers with different precisions concerning prior information is fine in settings with a single decision maker, such as Section 4, but does run into problems with competing decisions makers. The analysis in Section 3 supposes a common value for the prior (*i.e.*, both employers know $\hat{\alpha}_0$). If one presumes that precision is a function of the quality and amount of information previously received, then it should be, before they interact, that the employers possess different priors with probability one. To see this, recall expression (1). Suppose that one employer received, just before play, an additional signal; her estimate of ability at time 0—effectively her prior—would be

$$\hat{\alpha}'_0 = \frac{\tau_0}{\tau_0 + \eta} \hat{\alpha}_0 + \frac{\eta}{\tau_0 + \eta} s,$$

which she holds with precision $\tau'_0 = \tau_0 + \eta$. Her precision is greater, but her “prior,” $\hat{\alpha}'_0$ differs almost surely from a rival who just knows the true prior $\hat{\alpha}_0$ (*i.e.*, who missed the additional signal). The analysis in Section 3 does not allow for that. Hence, the parallel between bias and imprecise priors that works for a single decision maker does not obviously apply with competing decision makers.

In those settings, some bounds on rationality are required (*e.g.*, the naïve-firms assumption) for the situation to make sense.

The result that being a bad Bayesian need not result in losses (and can even be beneficial) challenges the usual argument for rational actors—namely that irrational actors will be driven from the market. To be sure, the analysis presumes either a particular form of competition among firms or no competition (Sections 3 and 4, respectively), but these results suggest that a research program of examining cognitive biases in learning models of governance is not doomed due to some form of Darwinian selection.²⁰

As noted earlier, there are many documented cognitive biases (see Malmendier and Taylor, 2015, or Hermalin and Weisbach, forthcoming, for partial lists). While such biases may not alter, in a fundamental sense, the results of earlier rational-actor models (*e.g.*, the principle thrusts of career concerns *à la* Holmstrom or models of CEO dismissal/retention, as surveyed in Hermalin and Weisbach, forthcoming, are not reversed), it is to be hoped this agenda can enrich such modeling, help explain seeming anomalies (*e.g.*, apparent overreacting to new information—see Hermalin and Weisbach, forthcoming, for a survey of some of the relevant empirical literature), and generate useful predictions.

APPENDIX A: PROOFS NOT GIVEN IN TEXT AND OTHER DETAILS

Proof of Proposition 3: As established in the text, the result follows if $\Delta - e^*(\lambda^\ell, \lambda^h)$ is positive. Observe

$$\Delta - e^*(\lambda^\ell, \lambda^h) = \frac{(1 - \lambda^\ell)e^*(\lambda^\ell, \lambda^\ell) - (1 - \lambda^h)e^*(\lambda^h, \lambda^h)}{\lambda^h - \lambda^\ell} - e^*(\lambda^\ell, \lambda^h).$$

That difference has the same sign as

$$(1 - \lambda^\ell)e^*(\lambda^\ell, \lambda^\ell) + \lambda^\ell e^*(\lambda^\ell, \lambda^h) - (1 - \lambda^h)e^*(\lambda^h, \lambda^h) - \lambda^h e^*(\lambda^\ell, \lambda^h). \quad (34)$$

By Proposition 2, expression (34) is greater than

$$\begin{aligned} e^*(\lambda^\ell, \lambda^h) - (1 - \lambda^h)e^*(\lambda^h, \lambda^h) - \lambda^h e^*(\lambda^\ell, \lambda^h) \\ = (1 - \lambda^h)(e^*(\lambda^\ell, \lambda^h) - e^*(\lambda^h, \lambda^h)) > 0, \end{aligned}$$

where the inequality also follows from Proposition 2. The result follows. ■

Proof of Lemma 2: Differentiating the lefthand side of (21) with respect to

²⁰See also Malmendier and Taylor (2015), which, *inter alia*, discusses why market pressures may not eliminate cognitive biases.

e_1 yields

$$\begin{aligned} & - (1 - \lambda^\ell) \phi \left((e^*(\lambda^\ell, \lambda^h) - e_1) \sqrt{H} \right) \sqrt{H} \\ & + (1 - \lambda^h) \phi \left((e_1 - e^*(\lambda^\ell, \lambda^h)) \sqrt{H} \right) \sqrt{H} - c''(e_1) \\ & = -(\lambda^h - \lambda^\ell) \phi \left((e_1 - e^*(\lambda^\ell, \lambda^h)) \sqrt{H} \right) \sqrt{H} - c''(e_1) < 0, \end{aligned}$$

where the equality follows because the standard normal is symmetric about zero (*i.e.*, $\phi(\xi) = \phi(-\xi)$). This establishes that the second derivative of (20) is everywhere negative, which establishes the result. ■

Lemma A.1. *The firm's marginal return to investing in information in expression (29) is positive.*

Proof: Observe the marginal return can be written

$$\begin{aligned} & \int_{-\infty}^{F(\lambda)} (-f) \phi(y \sqrt{H}) \sqrt{H} dy + \int_{F(\lambda)}^{\infty} (1 - \lambda) y \phi(y \sqrt{H}) \sqrt{H} dy \\ & = \int_{-\infty}^{\infty} \max \{ -f, (1 - \lambda) y \} \phi(y \sqrt{H}) \sqrt{H} dy, \quad (35) \end{aligned}$$

where the equality follows from the definition of $F(\lambda)$. Because

$$\max \{ -f, (1 - \lambda) y \} > (1 - \lambda) y$$

for a set of y s with positive measure, it follows the second line of (35) must strictly exceed

$$\int_{-\infty}^{\infty} (1 - \lambda) y \phi(y \sqrt{H}) \sqrt{H} dy = (1 - \lambda) \mathbb{E} y = 0$$

(recall (12)). The result follows. ■

Lemma A.2. *Maintain the assumptions of the example used to generate Figure 1. Let, however, $\beta(e) = \kappa e$, $\kappa > 0$, $b \in [0, 1/\phi(1))$, and f be unspecified. Then the weighting factor, λ , that maximizes firm value is given by (33).*

Proof: Given all the other assumptions, if $b \in [0, 1/\phi(1))$, the executive's choice of effort problem has an interior solution (see Hermalin, 2005, Lemma 1). It can be shown that

$$\begin{aligned} e^E(\lambda, p) &= p b \phi \left(-\frac{f}{1 - \lambda} \right) \quad \text{and} \\ p^*(\lambda) &= (1 - \lambda) \phi \phi \left(-\frac{f}{1 - \lambda} \right) - f \Phi \left(-\frac{f}{1 - \lambda} \right). \end{aligned}$$

Hence, the firm's true expected value is

$$\begin{aligned} \kappa p b \phi \left(-\frac{f}{1-\lambda} \right) - \left(\frac{1}{2} - \lambda \right) \left((1-\lambda) \phi \left(-\frac{f}{1-\lambda} \right) - f \Phi \left(-\frac{f}{1-\lambda} \right) \right) \\ + \frac{1}{2} \left((1-\lambda) \phi \left(-\frac{f}{1-\lambda} \right) - f \Phi \left(-\frac{f}{1-\lambda} \right) \right)^2. \end{aligned} \quad (36)$$

Tedious calculations reveal that the derivative of (36) with respect to λ equals

$$(1 - 2b\kappa - 2\lambda)Z(f, \lambda), \quad (37)$$

where $Z(f, \lambda) > 0$ is a complicated function of f and λ . Setting (37) equal to zero and solving for λ yields (33). ■

REFERENCES

- DeGroot, Morris H.**, *Optimal Statistical Decisions*, New York: McGraw-Hill, 1970.
- Fama, Eugene F.**, “Agency Problems and the Theory of the Firm,” *Journal of Political Economy*, 1980, 88 (2), 288–307.
- Gilovich, Thomas**, *How We Know What Isn't So: The Fallibility of Human Reason in Everyday Life*, New York: The Free Press, 1991.
- Güner, A. Burak, Ulrike Malmendier, and Geoffrey Tate**, “Financial Expertise of Directors,” *Journal of Financial Economics*, May 2008, 88 (2), 323–354.
- Hermalin, Benjamin E.**, “Trends in Corporate Governance,” *Journal of Finance*, 2005, 60 (5), 2351–2384.
- **and Michael S. Weisbach**, “Endogenously Chosen Boards of Directors and Their Monitoring of the CEO,” *American Economic Review*, March 1998, 88 (1), 96–118.
- **and —**, “Information Disclosure and Corporate Governance,” *Journal of Finance*, February 2012, 67 (1), 195–233.
- **and —**, “Assessing Managerial Ability: Implications for Corporate Governance,” in Benjamin E. Hermalin and Michael S. Weisbach, eds., *Handbook of the Economics of Corporate Governance*, Amsterdam: Elsevier, forthcoming.
- Holmstrom, Bengt**, “Managerial Incentive Problems—A Dynamic Perspective,” *Review of Economic Studies*, January 1999, 66 (226), 169–182.
- Kahneman, Daniel**, *Thinking, Fast and Slow*, New York: Farrar, Strauss, and Giroux, 2011.

-
- Malmendier, Ulrike and Timothy Taylor**, “On the Verges of Overconfidence,” *Journal of Economic Perspectives*, Fall 2015, 29 (4), 3–8.
- Plous, Scott**, *The Psychology of Judgment and Decision Making*, New York: McGraw-Hill, 1993.
- Stein, Jeremy C.**, “Efficient Capital Markets, Inefficient Firms: A Model of Myopic Corporate Behavior,” *Quarterly Journal of Economics*, November 1989, 104 (4), 655–669.
- , “Agency, Information and Corporate Investment,” in George M. Constantinides, Milton Harris, and René Stulz, eds., *Handbook of the Economics of Finance*, Amsterdam: North-Holland, 2003.