NOISY AGENTS

Francisco Espinosa ⓡ Debraj Ray[†]

November 2018

Abstract. A principal seeks to retain good agents. Agent type is signaled with some ambient noise. Agents can add or remove noise at a cost. We show that monotone retention strategies, in which the principal keeps the agent above some signal threshold, are generically never equilibria. The main result identifies an equilibrium in which the principal retains the agent if the signal is "moderate" and replaces him otherwise. We consider various extensions: non-normal signal structures, non-binary types, interacting agents, costly mean-shifting, dynamics with term limits, and principal commitment. We discuss applications to risky portfolio management, fundraising, and political risk-taking.

1. INTRODUCTION

We study a model of deliberately noisy signaling. An agent who privately knows his type (good or bad) seeks to be retained by a principal. The principal wishes to retain a good type, and to remove a bad type. The agent generates a noisy but informative signal centered on his type. He can choose to amplify or reduce the precision of this process. But there are restrictions. First, such actions are costly. Second, the signal structure is constrained by the type; specifically, the mean of the signal is given by the type. Third, signals cannot be tampered with ex post. Specifically, the signal realization cannot be augmented *nor* reduced: there is no "free disposal." The principal observes the signal realization (but not the structure), and makes a retention decision.

A study of the equilibria of such a game is the subject matter of our paper.

Minimal though this model might be, it lends itself easily to extensions and has several applications. We describe these in more detail in Section 8. One setting is portfolio management, in which a money manager with career concerns might overload on risk in the hope of scoring big. The principal — his client — might be able to verify the portfolio at any one point of time, so that there is no chance of ex-post "disposal" of financial returns, but may not be aware ex ante when a risky strategy is adopted. Moreover, while the client may be happy with a large return today, her main goal is to find a durable relationship with a competent money manager, and in this sense the (possibly welcome) return today is *also* a signal about the manager's type.

Or consider a non-governmental organization (or NGO) of unknown competence, seeking funding from donors. The NGO could take on a safe but rather humdrum project — e.g., it could become a social-services provider in a well-understood urban setting — with outcomes that fairly accurately signal its competence. Or it could try something risky — perhaps a program of disease eradication in a distant rural setting, with some chance of attention-grabbing success. A

[†]Espinosa: New York University, fje206@nyu.edu; Ray: New York University and University of Warwick, debraj.ray@nyu.edu. Ray acknowledges funding under National Science Foundation grant SES-1629370. We thank Dilip Abreu, Dhruva Bhaskar, Emir Kamenica and Gaute Torvik for useful comments. Names are in random order, following Ray ⓡ Robson (2018).

potential donor cannot assess these ex ante risks and only sees the final outcome. Again, while the outcome is payoff-relevant to the donor, her concerns lie with whether the NGO should be funded for *future* activities, and for this purpose the current outcome is also a signal.

Likewise, a political leader lacking in competence could try something relatively safe (such as moderate and possibly ineffectual tweaks to an existing health care law), or could alternatively attempt something very risky (say, a summit with a rogue power), in the hope of a spectacular success — but simultaneously braving the chances of abject failure. Again, the median voter may not be able to assess the risk ex ante, but would infer the extent of risk-taking, along with the observed attendant outcome, and use these as guidelines for possible re-election.

One could continue: a government under pressure might inject noise into official statistics, an individual might take risky steps to bolster his cv for an upcoming promotion or interview, a less-than-competent lawyer might call a high-risk witness (who could destroy the case or win it), an athlete might engage in doping, and so on.

In this setting, it might seem natural to study equilibria in which the principal uses a "monotone" or threshold strategy; that is, she retains the agent when the signal realization is above some bar and replaces him otherwise. Our first result (Proposition 1) is that such monotone strategies can never be equilibria except for non-generic values of the parameters. Indeed, a generic equilibrium must invariably involve either *bounded retention* or *bounded replacement*; that is, bounded intervals of signals in which retention or replacement occur (Proposition 2). Under the former, bad agent types choose higher noise than their good counterparts. That noise is then more likely to generate very good or very bad signals. The principal therefore treats *both* kinds of excessive signals with suspicion, and retains the agent if and only if the signal falls in some intermediate bounded set. In short, she follows the maxim: "if it's too good to be true, it probably is."

In contrast, under bounded replacement, a principal retains the agent if the returns are extreme. In this equilibrium, a good type chooses higher noise, so a moderate signal is viewed with suspicion. This is a strange outcome, but it could happen; we provide examples. That said, we will argue that of these two types of equilibria, bounded retention is the more reasonable and more likely outcome. Indeed, for a wide range of parameters, a bounded retention equilibrium exists (Proposition 3), and there is also an identifiable range in which it is the *only* kind of equilibrium (Propositions 4 and 5): a bounded replacement equilibrium does not exist.

As already mentioned, there is no "free disposal" of signals; that is, the signal realization cannot be manipulated "upward" nor "downward." Let us return to our examples to illustrate this. First, consider the money manager who looks after your funds, and ends with some realized outcome. If you are aware of his choice of portfolio, then you have access to all realized rates of return, high or low, even though you do not know ex ante what the right decisions are. So that realized rate cannot be manipulated, either upward or downward, by the money manager, though in a setting where the portfolio cannot be observed, a different assumption could be more natural. Or imagine the political leader's attempt to organize a summit with a rogue power. A variety of outcomes are possible, and ex ante, one cannot predict which one it will be. What we *do* know, however, is that no matter what the outcome, it cannot be taken back, or "freely disposed of."

The equilibria we uncover are robust to various extensions: non-normal signal structures satisfying a strong version of the monotone likelihood ratio property (Propositions 6 and 7 in Section 7.1), non-binary agent types (Proposition 8 in Section 7.2), interacting agents with privately known types (Proposition 9 in Section 7.3), or situations in which the agent can shift the *mean* of their signal, presumably at an additional cost (Section 7.4). We also introduce a simpler variant of the model with costless choice of noise (Section 7.5 and Proposition 10), and apply this variant to a dynamic version of the model with agent term limits, in which the principal's outside option from a new agent is endogenously determined (Proposition 11 in Section 7.6). We discuss applications to risky portfolio management (Section 8.1), to funding-raising by organizations (Section 8.2), and to political risk-taking (Section 8.3).

2. Related Literature

While our main results are (to our knowledge) new, we are far from the first to study models of deliberate vagueness or noise.[1] The cheap talk literature beginning with Crawford and Sobel (1983) can be thought of as a leading example of noisy communication. In that example nothing binds the sender, because talk is cheap. In contrast, as explained above, our chosen communication structures must have mean equal to the true state, and the choice of structure is costly. It is central to the analysis that each individual chooses a *distribution* over signals, rather than an announcement, and cannot hide the outcome ex post.

The choice of an information structure is present in the Bayesian persuasion model of Kamenica and Gentzkow (2011). But neither sender not receiver knows agent type ex ante, and the chosen information *structure* is fully observed by the receiver. This last feature — an observed information structure — is shared by Degan and Li (2016), but the type of the agent is privately known, as in our model.[2] In contrast, in our setting, the choice of information structure is not observed, only the signal. These three models are complementary, and generate their own distinctive features, on which more in Section 7.8.[3]

Dewan and Myatt (2008) examine a model of leadership in which an individual's clarity in communication is a virtue, in that it attracts attention and thereby generates influence. But clarity also requires lower processing time from the audience, leaving more time for the audience to listen to others. Therefore zero noise is not chosen, because a leader might wish to hold on to an audience for longer, effectively dissuading them from listening to others.

Edmond (2013) also studies the obfuscation of states (say by a dictatorial regime). While such obfuscation occurs through the shifting of the *mean* signal with the use of a costly action, he also considers the case in which the state is communicated in a deliberately noisy way, with mean unchanged. The noise prevents coordination by receivers against the interests of the regime.

[1] In this brief review we omit discussion of a related but distinct literature with *exogenous* noise, as in the limit pricing game studied by Matthews and Mirman (1983), the choice of mean return by managers of unknown quality who might seek to herd (Zweibel 1995), or inference settings when values have exogenous but unknown precision (Subramanyam 1996).

[2] At the time we wrote our first draft, we were unaware of this paper, but cite it now as relevant to our work.

[3] There are also models of unobserved precision choice with no player types at all; see, e.g., Penno (1996) on financial reporting.

Edmond restricts attention in his analysis (by assumption) to receiver-actions that are monotone in the signal realization. In contrast, in our setting, the *non*-monotonicity of receiver actions is a fundamental and robust outcome of the model.

Harbaugh, Maxwell and Shue (2016) study the inclinations of a sender to distort the news about multiple projects, depending on the overall realization of news.[4] By distorting the news from bad projects when the overall news is good, and by exaggerating the news from good projects when the news is bad, the sender effectively adjusts the *realized* spread of *multidimensional* news over multiple projects in opposite directions, depending on mean realizations. Such distortions are separate from mean-preserving noisy announcements; moreover, the focus is on realized spread. The results we develop are entirely distinct, but they too take note of a different "too-good-to-be-true" inference problem, whereby a posterior update reverts more strongly towards the prior for certain distributions when an extreme signal is received. Such extremeness (relative to the other components) is therefore eschewed by the sender when the mean news is good.

Hvide (2002) studies tournaments with moral hazard where two risk-neutral agents compete for a prize. The contractible variable is output, which is the result of their effort and a random component. A risk-neutral committee wants to ensure that agents exert high costly effort. If agents can costlessly increase noise in the random component of output (assumed to be normally distributed), rewarding the agent with the highest realization of output will lead to an equilibrium with low effort and high noise. If agents are rewarded depending on who gets closer to some pre-stipulated, finite level of output, a high effort low noise equilibrium is achieved. Less related are Palomino and Prat (2003) and Barron, Giorgiadis and Swinkels (2017), who also study situations in which agents can inject noise into a moral hazard setting.[5]

Finally, there is a literature on policy uncertainty (see, for example, Shepsle 1972, Alesina and Cukierman 1990, Glazer 1990, Aragones and Neeman 2000, and Aragones and Postlewaite 2007), often referred to as "strategic ambiguity." Candidates offer policy platforms which can be more or less ambiguous, and this ambiguity generates uncertainty about the policies the candidate could implement were she to win the election. (An empirical analysis of strategic ambiguity can be found in Campbell 1983.) Ambiguity here is the result of the trade-off faced by the candidate between winning the election and implementing a certain policy (either his ideal policy or the most expedient one).

3. The Model

3.1. **Baseline Model.** An agent works for a principal. The agent can be good (g) or bad (b). He knows his type. The principal doesn't. She has a prior probability $q \in (0, 1)$ that the agent is good.[6] At the end of a single round of interaction, to be described below, the principal decides whether or not to retain the agent. Retention of an agent of type $k = g, b$ yields an expected

[4] Footnote 2 applies here as well.

[5] In Palomino and Prat (2003), an agent manages a portfolio for a principal but can hide part of the return, which forces monotonicity of any optimal contract. Barron, Giorgiadis and Swinkels (2017) study contracts that are immune to risk-taking, thereby forcing concavity of agent payoff with respect to produced output before the noise is added. A similar theme is also present in the endogenous risk-taking model studied in Ray and Robson (2012).

[6] On multiple types, see Section 7.1.

payoff of U_k to the principal, with $U_g > U_b$. Non-retention yields the principal $V \in (U_b, U_g)$. The type-k agent gets a payoff equal to 1 if he is retained and 0 otherwise. The agent therefore prefers to be retained regardless of type, while the principal prefers to retain the good agent.

The principal receives a signal from the agent, which is presumably indicative of his type. Based on the realization of that signal, the principal decides whether or not to retain the agent. The agent has some control over the distribution of this signal, but conditional on this, cannot alter in any way the signal realization. Specifically, suppose that the signal is given by

$$x = \theta_k + \sigma_k \epsilon,$$

for $k = g, b$, where θ_k is a type-specific mean with $\theta_g > \theta_b$, $\epsilon \sim N(0,1)$ is zero-mean normal noise, and σ_k is a term that scales the noise. That is, the agent cannot shift the mean of his signal (though see Section 7.4), but he can modulate its precision. The principal does not observe σ_k, but observes the realization of the signal. She then decides whether to retain or replace the agent.

There is a cost to modulating precision. That is, there is some "natural" baseline degree of noise, but deviations from that baseline are costly in either direction. Specifically, we assume that there is a smooth, strictly convex cost function $c(\sigma)$, which reaches its minimum value (normalized to zero) at some positive noise level $\sigma = \underline{\sigma}$. So cost increases as we depart from $\underline{\sigma}$ in either direction. Assume that $c(0) = c(\infty) = \infty$; that is, it is extremely costly at the margin to be fully precise or fully noisy. The former restriction is presumably self-explanatory. To understand the latter, note that σ large implies that very good (and very bad) signals are generated with positive probability, or equivalently, that the public evaluation of agent actions can be excellent or dismal. In effect, we assume that it is costly to disguise one's true characteristics and intentions in an attempt to generate some chance that the evaluation will be positive.

Because V is the payoff to the principal from non-retention, the variable $p \in (0,1)$, defined by

$$pU_g + (1-p)\,U_b \equiv V; \tag{1}$$

is interpretable as an "outside option probability" that leaves the principal indifferent between retaining and replacing. How might this compare with q, the prior probability that the agent is good? A salient benchmark is $p = q$; we refer to this as the *balanced model*. But there may be systematic departures of p from q. Notice that V incorporates the option value of dealing with a new agent, so in a dynamic context, p should not be smaller than q, and may well be strictly larger.[7] Call this a model with an *optimistic future*. If, on the other hand, our current agent is an ongoing hire about whom some (positive) information has already been received, then p could be smaller than q; call this a model with a *pessimistic future*. We allow for all three cases for now, though in a simple dynamic extension of our model with term limits, in which V is endogenous (Section 7.4) we will be able to whittle these alternatives down.

[7] Let V be the equilibrium value of restarting an interaction in an infinite horizon setting, normalized by a discount factor δ. Assume the principal gets utility from the agent in every period, though payoffs cannot be used as signals. Once an agent is replaced, the principal gets V again. By (1), we have $V = pU_g + (1-p)\,U_b$. However, since "replace the agent no matter what" is a feasible move for the principal at any date, we also have $V \geq (1-\delta)[qU_g + (1-q)\,U_b] + \delta V$ when our agent is a new hire, which implies that $p \geq q$.

3.2. **Equilibrium.** Agent k chooses noise σ_k. The principal does not observe the choice of noise, just some realization or signal x with distribution $N(\theta_k, \sigma_k^2)$. The principal uses Bayes' Rule to retain the agent if (and modulo indifference, only if)

$$\Pr(k = g|x) = \frac{q\frac{1}{\sigma_g}\phi\left(\frac{x-\theta_g}{\sigma_g}\right)}{q\frac{1}{\sigma_g}\phi\left(\frac{x-\theta_g}{\sigma_g}\right) + (1-q)\frac{1}{\sigma_b}\phi\left(\frac{x-\theta_b}{\sigma_b}\right)} \geq p, \tag{2}$$

where ϕ is the pdf of the standard normal. Rearranging, we have retention if and only if

$$\frac{\frac{1}{\sigma_b}\phi\left(\frac{x-\theta_b}{\sigma_b}\right)}{\frac{1}{\sigma_g}\phi\left(\frac{x-\theta_g}{\sigma_g}\right)} \leq \frac{1-p}{p}\frac{q}{1-q} =: \beta \in \mathbb{R}. \tag{3}$$

Simple algebra involving the normal density yields the equivalent expression

$$\left(\sigma_g^2 - \sigma_b^2\right)x^2 + 2\left(\sigma_b^2\theta_g - \sigma_g^2\theta_b\right)x + \left(\sigma_g^2\theta_b^2 - \sigma_b^2\theta_g^2 + 2A\sigma_g^2\sigma_b^2\right) \geq 0, \tag{4}$$

where $A := \ln(\beta\sigma_b/\sigma_g)$. The inequality (4) defines a *retention regime*, a zone X of signals for which the principal will want to retain the agent. An *equilibrium* is a configuration (σ_g, σ_b, X) such that given (σ_g, σ_b), X is the set of "retention signals" x which solve (4), and given X, each type k chooses σ_k to maximize the probability of retention net of noise cost:

$$\sigma_k \in \arg\max_{\sigma} \int_X \frac{1}{\sigma}\phi\left(\frac{x-\theta_k}{\sigma}\right)dx - c(\sigma).$$

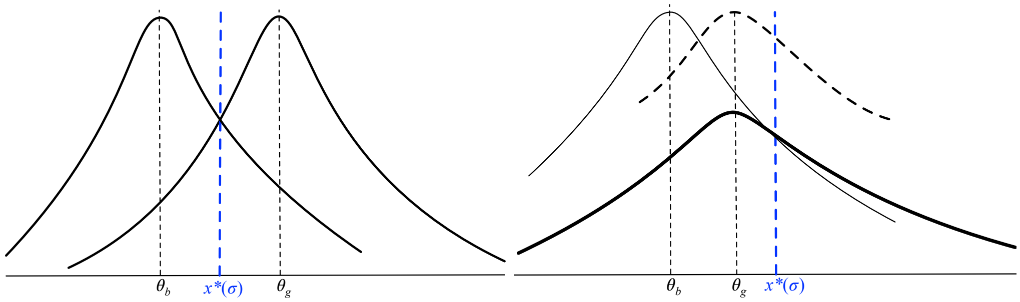
4. Preliminary Remarks on Retention Regimes

Recall that a retention regime is given by a set X of signals for which the principal will want to retain the agent.

4.1. **Trivial Retention Regimes.** Two examples of retention zones are (a) "always retain," so that $X = \mathbb{R}$, and (b) "always replace," that is, $X = \emptyset$. As far as equilibrium regimes are concerned, these are of little interest. Both generate complete indifference across the two types as to the noise regime. If the cost function for noise is strictly increasing away from $\underline{\sigma}$, then $\sigma_g = \sigma_b = \underline{\sigma}$ in such an equilibrium. But then the expression in (4) *must* alter sign over different values of x, knocking out either regime. Thus trivial equilibria do not exist in our setting.

4.2. **Monotone Retention Regimes.** An equilibrium regime is *monotone* if there is a finite threshold x^* such that the principal replaces the agent for signals on one side of x^*, and retains him for signals to the other side of x^*.[8] See Figure 1.

[8] Whether x^* is included or not doesn't matter.



(A) Threshold between types (B) Threshold to one side of types

FIGURE 1. The Symmetric Threshold $x^*(\sigma)$

A monotone retention regime arises (and can *only* arise) when both types transmit with the *same* noise $\sigma_b = \sigma_g = \sigma$.[9] Then (4) reduces to the condition

$$x \geq x^* (\sigma) := \frac{\theta_g + \theta_b}{2} - \frac{\sigma^2}{\theta_g - \theta_b} \ln (\beta) , \tag{5}$$

and in particular, the retention zone in a monotone equilibrium must be of the form $X = [x^*, \infty)$. Loosely, $x^*(\sigma)$ is the threshold above which the principal deduces that a signal from two possible noisy sources of *equal* variance is more likely to be coming from the higher-mean source. In fact, this is the exact interpretation of $x^*(\sigma)$ in the balanced model with $p = q$, for then $\beta = 1$ and

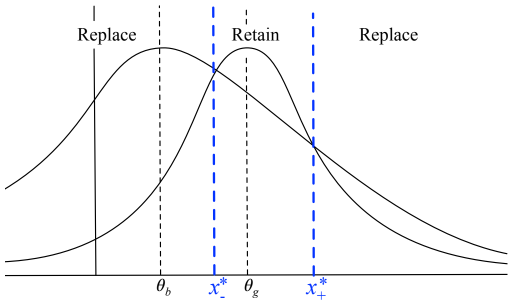
$$x^* (\sigma) = \frac{\theta_g + \theta_b}{2},$$

which is the mid-point between the two means. Notice that x^* is entirely insensitive to σ in the balanced model. With $p = q$, the decision to retain is just a matter of comparing two likelihoods, and Panel A of Figure 1 shows that the likelihood for the good type dominates to the right of $(\theta_g + \theta_b)/2$. However, when $p \neq q$, retention is not simply dependent on relative likelihoods, but also on how pessimistic or optimistic the principal feels about future agents, which is measured by the ratio of q to p, as proxied by β. In the optimistic future setting, we have $\beta < 1$, and better performance is required for the principal to retain the current agent; $x^*(\sigma)$ is higher for each σ as β falls. Panel B of Figure 1 depicts the consequences of an optimistic future, pushing $x^*(\sigma)$ to the right of the midpoint between θ_b and θ_g, and possibly even to the right of θ_g.

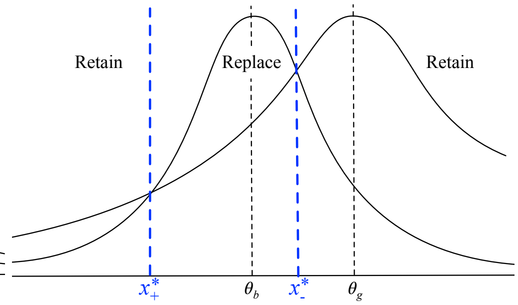
4.3. **Non-Monotone Retention Regimes.**

When agents of different types transmit at different noises, the corresponding best response for the principal is never a monotone regime. For instance, when the bad type chooses higher noise than the good type, there *cannot* be a single threshold for retention. Good news — but only moderately good news — offer the best likelihood ratios in favor of the good type, and will generate retention. But an *extremely* good signal will be regarded as too good to be true: for those signals, the higher chosen variance of the bad

[9]To see this, recall the retention condition (4), and notice that if $\sigma_g \neq \sigma_b$, then the resulting retention regime is either trivial or non-monotone.



(A) Bounded Retention



(B) Bounded Replacement

FIGURE 2. Differential Noise and the Retention Decision

type will dominate the lower mean, leading to a high likelihood that the signal was emitted by the bad type. Panel A of Figure 2 illustrates this (for the balanced case).

On the other hand, if the bad type transmits at lower noise than the good type, the retention rule is flipped. Now replacement occurs in some bounded interval of signal realizations, but elsewhere the principal will actually retain. See Panel B of Figure 2. We make these observations more formal in Proposition 2 of Section 6.3.

5. MONOTONE RETENTION IS (ALMOST) NEVER AN EQUILIBRIUM

Monotone retention is a natural focal point of inquiry. The types in our model are ordered, so that all other things being the same, the good type is more likely to generate larger signals. In this sense larger signals appear to be prima facie evidence that the type emitting them is good.[10]

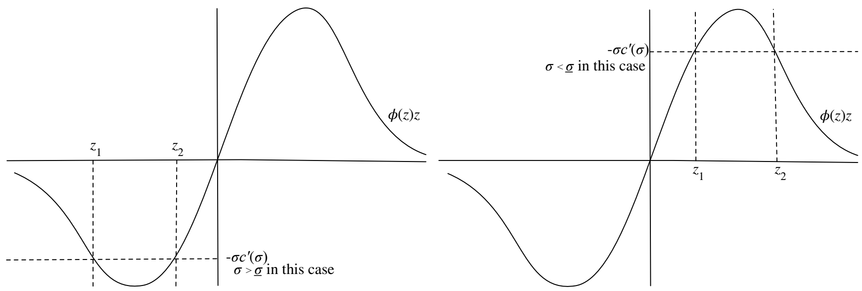
Indeed, in our model, an equilibrium can involve monotone retention; see Online Appendix for a specific example. But the example isn't robust: in "almost all" cases, the answer is no:

Proposition 1. *Generically, a monotone equilibrium can not exist. Specifically, given model parameters, there is at most one common value of σ that both players must choose in any monotone equilibrium, and this value is pinned down independently of the cost function for noise choice.*

For some intuition, consider any single retention threshold as in Figure 1. In this figure, the threshold lies strictly between the types of the two agents. As already discussed, both agent types must choose a common noise σ. But the incentives for each type push in opposite directions away from σ: with the cost of noise disregarded, the good type benefits from lower noise, while the bad type wants to amplify noise. Of course, the noise cost must be factored in, but the cost of the desired move must be non-positive for one of the two types.[11] It follows that at least one

[10] For instance, in the context of a global game in which a sender can manipulate the noise with which signals are emitted and seeks to prevent a coordinated attack against the sender, Edmond (2013) restricts his attention to monotone responses. We should add, though, that noise manipulation is only one of several extensions that Edmond studies in his paper, and it is not his main focus.

[11] The common value of σ is either weakly to the left or to the right of $\underline{\sigma}$, and the cost function is smooth.



(A) Optimistic Future ($\beta < 1$) (B) Pessimistic Future ($\beta > 1$)

FIGURE 3. Conditions for Monotone Retention

of the types will wish to deviate from the presumed equilibrium choice of σ, so no monotone equilibrium can exist in this case.

But it's possible that both types lie on the same side of the threshold. We need to dig deeper to handle such cases. Type k seeks to maximize

$$1 - \Phi\left(\frac{x^* - \theta_k}{\sigma_k}\right) - c\left(\sigma_k\right)$$

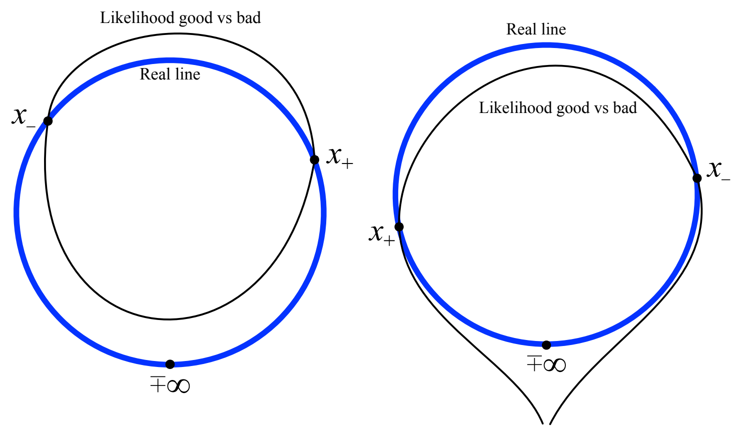
by choosing σ_k, and the corresponding first-order condition is

$$\phi\left(\frac{x^* - \theta_k}{\sigma_k}\right)\frac{x^* - \theta_k}{\sigma_k^2} - c'\left(\sigma_k\right) = 0, \tag{6}$$

where x^* is given by (5). Recall that both types need to choose the same value of σ for a monotone regime to emerge in equilibrium. Therefore, setting $\sigma_g = \sigma_b = \sigma$ and defining $\Delta := \theta_g - \theta_b$, we can rewrite the first-order condition for good and bad types as

$$\begin{aligned}\phi\left(\frac{\sigma}{\Delta}\ln(\beta) + \frac{\Delta}{2\sigma}\right)\left(\frac{\sigma}{\Delta}\ln(\beta) + \frac{\Delta}{2\sigma}\right) &= \phi\left(\frac{\sigma}{\Delta}\ln(\beta) - \frac{\Delta}{2\sigma}\right)\left(\frac{\sigma}{\Delta}\ln(\beta) - \frac{\Delta}{2\sigma}\right) \\ &= -\sigma c'(\sigma).\end{aligned} \tag{7}$$

Equation (7) tells us that we will need to study the function $\phi(z)z$; Figure 3 does so. Denote $\frac{\sigma}{\Delta}\ln(\beta) - \frac{\Delta}{2\sigma}$ by z_1 and $\frac{\sigma}{\Delta}\ln(\beta) + \frac{\Delta}{2\sigma}$ by z_2. Given the shape of $\phi(z)z$, Figure 3 indicates how z_1 and z_2 must be located relative to each other: they must both have the same sign and generate the same "height." With an optimistic future ($\ln\beta < 0$), both z_1 and z_2 are negative; see Panel A. With a pessimistic future, $\ln\beta > 0$, so z_1 and z_2 are both positive as in Panel B. In each case, there is only one value of σ that can solve this requirement; i.e., just one value that fits the *first* equality in (7). It is entirely independent of the cost function for noise, and so the second equality cannot generically hold. (The Appendix formalizes the argument.)



(A) Bounded Retention (B) Bounded Replacement

FIGURE 4. Bounded Retention and Replacement Zones.

6. BOUNDED RETENTION AND REPLACEMENT REGIMES

6.1. **Two Possible Regimes.** With monotonicity out of the way, we are left with equilibria in which the two types choose different noise levels. In Section 4.3 we suggested that there could only be two possibilities:

1. *Bounded Retention.* The good type transmits with lower noise than the bad type, and Panel A of Figure 2 applies. The principal retains the agent if the signal is good but not "*too* good."

2. *Bounded Replacement.* In a bounded replacement equilibrium, the good type transmits with higher noise than the bad type, and Panel B of Figure 2 is relevant. In this case, the principal replaces the agent if the signal has moderate values, and retains him if the signal is extreme.

The reason that there are just these two possibilities, but no more, is evident from (4). Retention or replacement zones are demarcated by values of the signal that solve a quadratic equation, which has at most two real roots. The absence of a real root is indicative of a trivial "always-retain" or "always-replace" regime that we have already ruled out. So there must be two real roots, and therefore one of the two zones of retention or replacement *must* be a bounded interval.

Now, the quadratic criterion for replacement or retention is a feature of the normal distribution, so we won't make too much of it. It is perhaps possible that with more general signal distributions, there is alternation between replacement and retention. But the general point is that one of the two decisions must be guided by a bounded zone of signals (see Section 7.1 for more).

6.2. **Equilibrium Conditions.** It will be convenient to use the notation $[x_-, x_+]$ to denote the relevant interval when bounded retention occurs, and by $[x_+, x_-]$ to denote the interval when bounded replacement occurs. Figure 4 illustrates this by folding the real line on itself in a circle so that the ends $-\infty$ and $+\infty$ are identified with each other. The zone of retention can then always be thought of as the arc of the circle starting from x_- and moving to x_+ in a clockwise direction. The Figure also depicts the weighted relative likelihood of good versus bad types given their strategies; see the irregular ovals. If that likelihood lies "outside" the circle, the good type is more likely; if inside, the bad type is more likely. Any such equilibrium implies the following restrictions. First, because each type k seeks to maximize $\Phi\left(\frac{x_+-\theta_k}{\sigma_k}\right) - \Phi\left(\frac{x_--\theta_k}{\sigma_k}\right) - c\left(\sigma_k\right)$ by choosing σ_k, we have the necessary first-order conditions

$$\phi\left(\frac{x_- - \theta_k}{\sigma_k}\right)\left(\frac{x_- - \theta_k}{\sigma_k^2}\right) - \phi\left(\frac{x_+ - \theta_k}{\sigma_k}\right)\left(\frac{x_+ - \theta_k}{\sigma_k^2}\right) = c'\left(\sigma_k\right) \tag{8}$$

for each type $k = g, b$. Next, for $x = x_-, x_+$,

$$\beta\frac{1}{\sigma_g}\phi\left(\frac{x-\theta_g}{\sigma_g}\right) = \frac{1}{\sigma_b}\phi\left(\frac{x-\theta_b}{\sigma_b}\right) \tag{9}$$

represents the equalization of weighted likelihoods for the good and bad types; see Figure 4 which depicts the relative likelihoods for all realizations x. The principal is indifferent between retaining and replacing at the points x_- and x_+. Third, the weighted likelihood for the good type must have a higher slope in x relative to that for the bad type, evaluated at x_-, so that retention occurs to the right of x_- (again consult Figure 4). That means

$$\beta\frac{1}{\sigma_g^2}\phi'\left(\frac{x_- - \theta_g}{\sigma_g}\right) > \frac{1}{\sigma_b^2}\phi'\left(\frac{x_- - \theta_b}{\sigma_b}.\right),$$

Because $\phi(z) = (1/\sqrt{2\pi})\exp\{-z^2/2\}$ satisfies $\phi'(z) = -z\phi(z)$, this is equivalent to:

$$\beta\phi\left(\frac{x_- - \theta_g}{\sigma_g}\right)\frac{x_- - \theta_g}{\sigma_g^3} - \phi\left(\frac{x_- - \theta_b}{\sigma_b}\right)\frac{x_- - \theta_b}{\sigma_b^3} < 0. \tag{10}$$

Likewise, the weighted likelihood for the good type must have a *lower* slope in x relative to that for the bad type, evaluated at x_+, so that

$$\beta\phi\left(\frac{x_+ - \theta_g}{\sigma_g}\right)\frac{x_+ - \theta_g}{\sigma_g^3} - \phi\left(\frac{x_+ - \theta_b}{\sigma_b}\right)\frac{x_+ - \theta_b}{\sigma_b^3} > 0 \tag{11}$$

This set of equations and inequalities help to narrow down the equilibria of our model.

6.3. **Bounded Retention and the Type-Specific Choice of Noise.** We now use these equilibrium conditions to make a case for bounded retention as the "more natural" outcome. Begin by using (9) for $x = x_-$ in equation (10) to obtain

$$\left(\sigma_b^2 - \sigma_g^2\right)x_- < \sigma_b^2\theta_g - \sigma_g^2\theta_b.$$

In the same way, use (9) for $x = x_+$ in equation (11) to see that

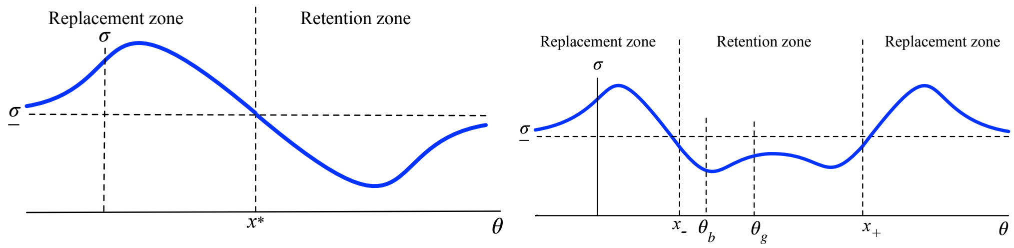
$$\left(\sigma_b^2 - \sigma_g^2\right)x_+ > \sigma_b^2\theta_g - \sigma_g^2\theta_b.$$

Combining these two inequalities, we must conclude that

$$\left(\sigma_b^2 - \sigma_g^2\right)\left(x_+ - x_-\right) > 0 \tag{12}$$

in any non-monotonic equilibrium. This formalizes an earlier informal discussion as:

Proposition 2. *Bounded retention with $x_+ > x_-$ is associated with $\sigma_b > \sigma_g$, while bounded replacement with $x_- > x_+$ is associated with $\sigma_b < \sigma_g$.*



(A) Monotone Retention (B) Bounded Retention

FIGURE 5. How Choice of Noise Varies With Agent Type

In light of this proposition, we ask which type has the incentive to use greater noise. Intuitively, it would seem that this should be the bad type — after all, if the good type could communicate with infinite precision, she would, while the bad type would seek to disguise her characteristics. Proposition 2 states that in that case, a bounded retention equilibrium must obtain. And yet matters are more complex than that. Infinite precision is not available except at infinite cost, and within the realm of positive noise choices, the good and bad types may have marginal preferences for noise that criss-cross each other. An analysis of these two possible equilibrium regimes is therefore closely related to an understanding of noise choices.

Optimally chosen noise moves in a subtle and quite complicated way as a player's type moves relative to the retention zone. Figure 5, Panel A, illustrates this for a monotone retention threshold. When a player's type is outside the retention zone and far away from the threshold, it takes a large amount of noise to create a significant probability that a signal will be generated within the retention zone. That's costly, so noise converges to $\underline{\sigma}$ as the type moves far from the retention zone. Moving closer to the zone, noise increases, but reaches a maximum when the type is still some distance away. The easiest way to understand this is to think of what happens when the type is *on* the edge of the zone, at which point noise makes no difference to the chances of retention, so that the noise level is back to $\underline{\sigma}$ again. Now continue the process by moving the type into the retention zone. In this case, noise can throw the player out of the zone, so she seeks to lower it. Her optimum choice therefore falls below $\underline{\sigma}$. But the downward movement does not continue forever. Deep in the retention zone, the type is confident of remaining there, and so noise goes up again, converging again to $\underline{\sigma}$, but this time from below.

With bounded retention zones, the choice function exhibits even more non-monotonicities.[12] Panel B of Figure 5 shows that there will generally be five turning points. There is one each for either side of the retention zone, for the same reason as in the earlier discussion. There are three more within the retention zone: noise initially falls as an agent with type close to the edge avoids escape from the zone; then rises in the middle of the zone as the risk of escape falls, then falls again as the risk goes up, and finally rises as we approach the edge. (The noise choice at the edges is below $\underline{\sigma}$, because the retention zone is bounded.)

[12] Formal details are available on request from the authors.

6.4. **Existence of Equilibrium With Bounded Retention.** In what follows, we retain the complexities discussed above as they are not merely technical but intrinsic to the economics of the problem. But there are other complications that we did not emphasize. The single-peakedness of the noise distribution generates a non-convexity in the agent's optimization problem, which raises the possibility that an agent's choice could be multi-valued. For monotone or bounded retention regimes, such multivaluedness is more a technical nuisance than a feature of any economic import,[13] and we rule it out by assumption:

[U] For every monotone or bounded retention zone and for each agent type, the optimal choice of noise is unique.

It is possible to deduce [U] by placing alternative primitive restrictions on the parameters of the model. One is that the curvature of the cost function is large enough. The Appendix shows that a sufficient condition for [U] is

$$c''(\sigma) > \frac{\kappa}{\sigma^2} \text{ for all } \sigma \in [\sigma_*, \sigma^*], \tag{13}$$

where $\kappa \approx 0.6626$, and σ_* and σ^* are two distinct lower and upper bounds on noise that straddle $\underline{\sigma}$, such that $c(\sigma_*) = c(\sigma^*) = 1$.

While [U] is a technical restriction of little economic import, the next assumption we impose is substantive. Recalling that we normalized the agent's payoff from retention to equal 1, and from replacement to equal 0, it is obvious that no agent would ever choose a level of noise outside the interval $[\sigma_*, \sigma^*]$. Now imagine that both agents transmit *common* noise equal to the upper limit σ^*. We know already that the principal would respond by choosing a single threshold $x^*(\sigma^*)$ for retention, described by equation (5), reproduced here for convenience:

$$x^*(\sigma^*) = \frac{\theta_g + \theta_b}{2} - \frac{\sigma^{*2}}{\theta_g - \theta_b} \ln(\beta).$$

We ask that this threshold must lie in $[\theta_b, \theta_g]$.

This implies a restriction on the parameters of the model; specifically, on β. The assumption states that when the agent chooses common noise (equal to σ^*), the principal will "start retaining" from a threshold smaller than θ_g, and replace when a realized signal lies below θ_b. This requires the weighted relative likelihood for the type being good or bad to flip sign at some intermediate point between θ_b and θ_g. It should be noted that this condition is *automatically* satisfied in the balanced case with $\beta = 1$, because in that case, as already observed, $x^*(\sigma^*) = (\theta_g + \theta_b)/2$. More formally, we can write this condition as a set of restrictions on the extent to which β can depart from 1 "on either side." That is, we want the future to be neither too optimistic nor too pessimistic. Do this by subtracting the formula for $x^*(\sigma^*)$ from θ_b and then θ_g to obtain

$$-\frac{\Delta^2}{2\sigma^{*2}} \leq \ln(\beta) \leq \frac{\Delta^2}{2\sigma^{*2}}. \tag{14}$$

Proposition 3. *Under Conditions* [U] *and* (14)*, there is an equilibrium with bounded retention.*

[13] For bounded *replacement* regimes, the possibility of multiple solutions is more natural. For instance, an agent located in one of the two retention zones to the side, but close to the replacement zone, could be indifferent between a small and a large choice of noise.

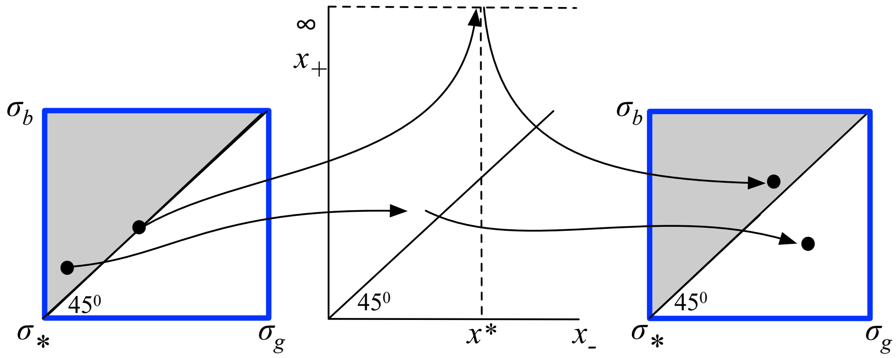


FIGURE 6. Fixed-Point Mapping to Show Existence of Bounded Retention

The proof provides some intuition for the result, so we loosely outline it here. Begin by searching for any equilibrium via a fixed-point mapping. The very first box in Figure 6 delineates the domain of that mapping. No agent will choose noise below σ_* or above σ^*, so we have a compact domain. The image of this mapping is derived as follows: for each (σ_g, σ_b), find the retention decision of the principal, shown in the middle graph (where x_- and x_+ are chosen), and then record the best response to that decision, shown by the continuation mapping into the last box, a replica of the one we started from. A fixed point of this mapping will yield an equilibrium.

The problem is that this fixed point mapping is not well-behaved. For any point (σ_g, σ_b) in the domain with $\sigma_b < \sigma_g$, the planner will best-respond with bounded *replacement*, and the "subsequent" response that completes the mapping is generally not continuous in (σ_g, σ_b). This discontinuity problem is endemic. Given that the retention region (under bounded replacement) is made out of separated zones, the choice of two or more noise levels that maximize retention probabilities is generally unavoidable. With that multiplicity in place, discontinuities in the fixed-point mapping are unavoidable. The simplest fixed-point approach is a dead end.

However, given our specific interest in the existence of a bounded retention equilibrium, we want to start from an even smaller domain, which is the shaded triangle in the left box, over which $\sigma_b \geq \sigma_g$. This subdomain is better-behaved — the principal chooses bounded retention (or a monotone threshold) as a best response, and the best response by the agents to each such retention policy is unique (by Condition U) and therefore continuous. But now the problem is different: it may well be that the mapping slips out of the smaller domain. In general, this slippage cannot be controlled. In Panel B of Figure 5, we have a bounded retention zone that could arise from some "starting" (σ_g, σ_b) with $\sigma_b > \sigma_g$. And yet in response, type g chooses larger noise as illustrated, which propels the system out of the triangle. See the lower pair of arrows in Figure 6.

At the same time, the mapping on the smaller domain has an interesting property. On the *boundary* between the two subdomains, the mapping "points inwards" whenever (14) holds. Look at the upper pair of arrows in Figure 6. The first arrow in the pair maps a point on the principal diagonal of the square (where $\sigma_b = \sigma_g$) to a monotone retention regime; that is, (x_-, x_+) is of the form (x^*, ∞). By our restriction on β in condition (14), x^* must lie between θ_b and θ_g. So the good type wants to reduce noise to remain within the retention zone, while the bad type wants to increase it. That means that the good type must choose noise $\sigma_b < \underline{\sigma}$, while the opposite is true of the bad type. But that implies a best response with $\sigma_b > \sigma_g$, which takes us back into the starting subdomain from its boundary. (It also implies, in passing, that under condition (14), a monotone equilibrium cannot exist, whether generically or otherwise.) A fixed point theorem due to Halpern (1968) and Halpern and Bergman (1968) then completes the argument, establishing the existence of a bounded retention equilibrium when β does not take on "extreme" values.

In summary, we have shown that when the future is neither too optimistic nor too pessimistic — and certainly when it is balanced — a bounded retention equilibrium must exist. Indeed, it could be the only equilibrium, as the following proposition suggests:

Proposition 4. *When $\beta = 1$, every equilibrium involves bounded retention. More generally:*

(i) *It cannot be that $\sigma_b \leq \underline{\sigma} \leq \sigma_g$.*

(ii) *If $\sigma_g < \underline{\sigma}$ and $\beta \leq 1$, then $\sigma_b > \sigma_g$ and there can only be bounded retention.*

(iii) *If $\sigma_g > \underline{\sigma}$ and $\beta \geq 1$, then $\sigma_b > \sigma_g$ and there can only be bounded retention.*

While these propositions are by no means a universal claim for bounded retention, it is true that moderate values of β do appear to be incompatible with bounded replacement. In Section 6.5, we will see that this is indeed the case: we can rule out bounded replacement equilibria for moderate values β. The case $\beta = 1$ in Proposition 4 is a good benchmark: it means that the prior q on the current agent equals the "effective prior" p on future agents.

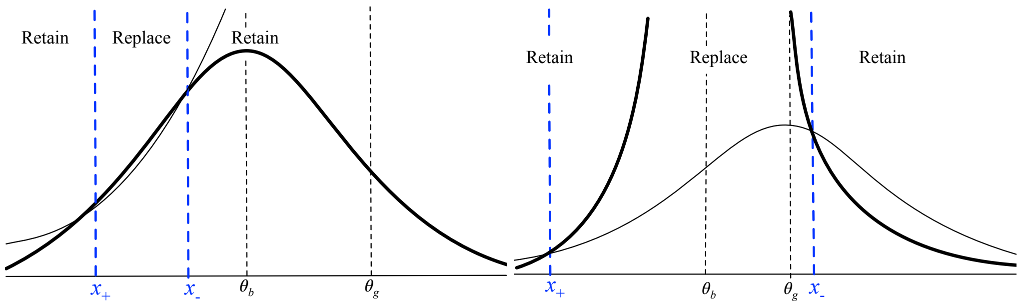
6.5. **Non-Existence of Bounded Replacement Equilibrium for Moderate β.** Moderate degrees of optimism or pessimism about the future are not only conducive to the existence of a bounded retention equilibrium, they push against the existence of a bounded replacement equilibrium. For instance, assume a sizable difference between the two types; specifically, that

$$\theta_g - \theta_b \geq \sigma^*, \tag{15}$$

where recall that σ^* is defined by the larger of the two solutions to $c(\sigma) = 1$.

Proposition 5. *Assume that Condition (14) used in Proposition 3 holds, and so does (15). Then only bounded retention equilibria can exist.*

While the Appendix contains a formal proof, it is easy enough to illustrate the main argument. Consider the same fixed point mapping used to establish the existence of a bounded retention equilibrium. The first component of this mapping takes noise choices $(\sigma_g, \sigma_b) \in [\sigma_*, \sigma^*]^2$ to best responses by the principal of the form (x_-, x_+). These responses, as already noted, could involve bounded retention ($x_- < x_+$), bounded replacement ($x_- > x_+$) or monotone regimes ($x_+ = \infty$). In all these cases, conditions (14) and (15) can be used to show that the bad type must lie *outside* the retention zone, while the good type lies *in* it. Now consider the second component of the fixed point mapping in which the agents react to these retention and replacement zones. The Appendix formally shows that in all such situations, the bad type exerts more noise in a quest to land inside the retention zone, while the good type attempts to reduce noise so as not



(A) $\beta \gg 1$

(B) $\beta \ll 1$

FIGURE 7. Possible Configurations for Bounded Replacement Equilibria

to wander out of it. In short, $\sigma_b > \sigma_g$. But now we've established that starting from *any* $(\sigma_g, \sigma_b) \in [\sigma_*, \sigma^*]^2$, the mapping points into the shaded triangle of Figure 6 in which $\sigma_b > \sigma_g$. Consequently, *every* equilibrium must have $\sigma_b > \sigma_g$, which — as we know already from Proposition 2 — must involve bounded retention.[14]

The heart of the argument for Proposition 5 concerns the location of types relative to replacement and retention zones. Figure 7 illustrates the exceptions. The density for the bad type is the thicker line in both cases. The Figure shows that β must be so large or so small (that is, the future is either super-optimistic or super-pessimistic) so that the intersection points of the two weighted densities are either on one side of both the mean types, or straddle them both.[15] These are the only two possible kinds of bounded replacement equilibria. For completeness, the Appendix provides examples for each of them. In one, both types are embedded in the retention zone as in Panel A of Figure 7, with $x_+ < x_- < \theta_b < \theta_g$. Because they want to remain there, both want noise lower than the ambient level. But the bad type is closer to the edge, so he will make a bigger effort than the good type to stay safe, and $\sigma_b < \sigma_g$. To justify this configuration as an equilibrium, the future must be super-pessimistic: $q \gg p$.

In the second example, as in Panel B of Figure 7, both θ_b and θ_g lie in the replacement zone, with $x_+ < \theta_b < \theta_g < x_-$, and both exert costly effort to escape it. The good type is embedded closer to the edge of the zone and has a high marginal benefit of noise, while the bad type is embedded deep in the zone and has only a low marginal benefit. The good type therefore exerts greater noise. The principal reacts by choosing a bounded replacement zone. To implement this equilibrium, the future must be super-optimistic: $p \gg q$.

[14]In particular, the careful reader will have noticed that under the additional restriction imposed by (15), the Halpern-Bergman theorem no longer needs to be invoked to prove Proposition 3; Brouwer will suffice.

[15]This argument shows, in particular, that Panel B of Figure 2 — which we put forward as a possible candidate for a bounded replacement equilibrium — cannot ever be a full equilibrium satisfying both best response conditions.

7. EXTENSIONS

In this Section, we describe several variations on the model. Section 7.1 replaces the normality restriction by signal structures that satisfy a strong version of the monotone likelihood ratio property. Section 7.2 studies non-binary agent types. Section 7.3 considers more than one agent, each with privately known type. Section 7.4 studies situations in which the agent can shift the *mean* of their signal, presumably at an additional cost. Section 7.5 introduces a simpler variant of the model with costless choice of noise. Section 7.6 applies this variant to a dynamic version of the model with agent term limits, in which the principal's outside option from a new agent is endogenously determined. Section 7.7 we evaluate whether the principal can benefit from committing ex ante to retention rules. Finally, Section 7.8 compares our setting with two related but distinct alternative formulations.

7.1. Non-Normal Signal Structures.

Consider the following generalization of our model: the signal x is given by:

$$x = \theta_k + \sigma_k \varepsilon, \tag{16}$$

where ε is distributed according to some differentiable density function f, which is positive on all of $\mathbb{R}$, with mean normalized to 0. The density for x given type k is

$$g\left(x|k\right) = \frac{1}{\sigma_k} f\left(\frac{x-\theta_k}{\sigma_k}\right)$$

Assume that f satisfies the monotone likelihood ratio property (MLRP) so that when two types transmit with the *same* noise, larger signals are increasingly likely to be associated with the higher type. Indeed, we assume that the relative likelihood for the good type climbs without bound as $x \to \infty$, while the opposite is true as $x \to -\infty$. Formally, we assume

Strong MRLP. $f(z-a)/f(z)$ is increasing in z whenever $a > 0$, with

$$\lim_{z\to\infty} \frac{f(z-a)}{f(z)} = \infty \text{ and } \lim_{z\to-\infty} \frac{f(z-a)}{f(z)} = 0. \tag{17}$$

In particular, the limit conditions ensure that a monotone regime is possible for any value of $\beta \in (0, \infty)$, provided both types use the same noise. The normal density satisfies (17).

Proposition 6. *Assume the signal structure is the one in* (16)*, and satisfies strong MLRP. Then:*

(i) *Generically, a monotone equilibrium can not exist. Specifically, given model parameters, there is at most one common value of σ that both players must choose in any monotone equilibrium, and this value is pinned down independently of the cost function for noise choice.*

(ii) *All other equilibria will have $\sigma_b \neq \sigma_g$, and will involve either a bounded retention zone or a bounded replacement zone.*

This proposition follows the same argument as in the basic model. Strong MLRP delivers the observation that "spreads dominate means," which is the argument that sends likelihood ratios for extreme signals in favor of the type using the higher spread. Therefore, a monotone equilibrium can only arise if both types are choosing the same amount of noise. The non-genericity of identical choices then follows lines similar to that for normal noise. The boundedness of either retention or replacement zones in equilibrium is an easy though not logically immediate consequence.[16]

We end this section with two observations. First, while we have not emphasized this so far, the boundedness of retention (or replacement) zones does not imply that such zones are necessarily *intervals*. Second, it would be useful to establish an analogue of Proposition 3: that bounded retention equilibria do exist for a intermediate interval of β values. The following proposition goes some way towards answering both questions.

Proposition 7. *Assume that the model is balanced or has a pessimistic future, so* $\beta \geq 1$. *Then, every bounded retention equilibrium must employ a bounded interval for retention.*

It is easy to combine this Proposition with analogues of Conditions U and (14) to obtain an existence theorem for bounded retention equilibrium. Specifically, if agents make unique choices of noise for every bounded or monotone retention interval, and if the future is not *too* pessimistic, then a bounded interval retention equilibrium must exist.

7.2. **Multiple Types.** We extend Proposition 1 to many types. We can do so at a level of generality that nests the two-type case, but it is expositionally easiest to assume that there is a prior on types given by some density $q(\theta)$ on $\mathbb{R}$. Let $\mathcal{Q}$ be the space of all such densities and give it any reasonable topology; for concreteness, think of $\mathcal{Q}$ as a subset of the space of all probability measures on $\mathbb{R}$ with the topology of weak convergence. A subset $\mathcal{Q}^0$ of $\mathcal{Q}$ is *degenerate* (relative to $\mathcal{Q}$) if its complement $\mathcal{Q} - \mathcal{Q}^0$ is (relatively) open and dense in $\mathcal{Q}$.

Given $q \in \mathcal{Q}$, each agent of type θ chooses noise $\sigma(\theta)$ as in the baseline model. Following the choice of noise, a signal is generated. The principal obtains payoff $u(\theta)$ from type θ, where u is some nondecreasing, bounded, continuous function. There is some given continuation payoff — V — from replacing an agent, which reasonably lies somewhere in between the retention utilities: $\lim_{\theta \to -\infty} u(\theta) < V < \lim_{\theta \to \infty} u(\theta)$. We also make the generic assumption that $u(\theta)$ is not locally flat exactly at V. As before, the principal maximizes expected payoff by deciding whether or not to retain the agent after each signal realization, and agents do their best to get retained, with the cost of noise factored in.

Proposition 8. *Fix all the parameters of the model except for the type distribution. Then, under Condition U, an equilibrium with a monotone retention regime can exist only for a degenerate subset of density functions over types.*

We outline the argument here (see Appendix for details). Think of a monotone retention regime of the form $[x^*, \infty)$. Figure 5 describes the optimal noise response; it attains a maximum at some distinguished value $\theta^* < x^*$. This picture translates perfectly as we move x^* around: θ^* moves with x^* staying at a fixed distance t^* from it, and this distance t^* is completely independent of the underlying density of types $q(\theta)$.

Now, we've already seen the sender with the highest noise enjoys the highest likelihood of having transmitted signals at the extreme ends of the line. That must mean that *conditional* on such

[16] In principle, a non-monotone equilibria could involve perennially alternating zones of retention and replacement.

extreme signals, the expected utility to the receiver must approximate $u(x^*-t^*)$. But then $u(x^*-t^*) \geq V$, for if not, a very large positive signal would be met with replacement, contradicting our presumption that the retention zone is $[x^*, \infty)$. But it can't be that strict inequality holds, for if it did, a very large *negative* signal would be met with *retention*, again contradicting our presumption. In short, $u(x^* - t^*) = V$. Because t^* is fixed (as already argued), and because u is not locally constant at V, this argument *fully pins down the retention threshold x^* independent of the density of types.*

But that points rather straightforwardly to a non-generic situation. After all, because x^* is the retention threshold, the receiver must be indifferent between replacement and retention at x^*; that is, the expected utility *at* x^* must be exactly V. But only a non-generic choice of density can guarantee that happy coincidence.

7.3. **Multiple Agents.** We've assumed that there is a single agent of unknown type. Suppose there are two agents, 1 and 2, who simultaneously signal their types, and the principal must decide which agent to retain. She wants to retain the better agent — or one of them, if she is indifferent. This sort of structure brings us closer to a model of political campaigns.

Assume that it is common knowledge that only one of the two agents is good. The agents know their own types and therefore both types. But they look identical ex ante to the principal, so her prior places equal probability on the two. The communication technology is unchanged:

$$x_i = \theta_{k(i)} + \sigma_{k(i)}\varepsilon_i, \tag{18}$$

where $i = 1, 2$, and $k(i)$ denotes i's type. The errors are independent and identically distributed standard normal random variables. In this game, by symmetry, a strategy for agent i is a function $\sigma : g, b \to \mathbb{R}_+$. As for the principal, a strategy is a function $r : \mathbb{R}^2 \to \{1, 2\}$, which indicates for every possible pair of signals (x_1, x_2) the agent she wants to retain. After observing (x_1, x_2) the principal retains agent 1 if (and, modulo indifference, only if)

$$\frac{\frac{1}{\sigma_g}\phi\left(\frac{x_1-\theta_g}{\sigma_g}\right)}{\frac{1}{\sigma_b}\phi\left(\frac{x_1-\theta_b}{\sigma_b}\right)} \geq \frac{\frac{1}{\sigma_g}\phi\left(\frac{x_2-\theta_g}{\sigma_g}\right)}{\frac{1}{\sigma_b}\phi\left(\frac{x_2-\theta_b}{\sigma_b}\right)}. \tag{19}$$

In this setting, a *monotone equilibrium* is defined as one where the principal retains the agent with the higher signal value. Once again, monotonicity can only be achieved if both *types* of agent play the same σ, but that won't happen.

Proposition 9. *If an equilibrium exists, it can only be the case that $\sigma_b > \sigma_g$, and the principal retains agent 1 if and only if $|x_1 - \hat{x}| \leq |x_2 - \hat{x}|$, where $\hat{x} = (\sigma_b^2\theta_g - \sigma_g^2\theta_b)/(\sigma_b^2 - \sigma_g^2)$ is the signal value that maximizes the likelihood ratio $\frac{1}{\sigma_g}\phi\left(\frac{x-\theta_g}{\sigma_g}\right)/\frac{1}{\sigma_b}\phi\left(\frac{x-\theta_b}{\sigma_b}\right)$. In particular, monotone equilibria do not exist.*

The proof of this proposition is long and involved, and we relegate it to the Online Appendix. Intuitively, when both types choose the same level of noise, the principal retains the one with the higher signal realization. But the bad type then wants to inject additional noise, since the good type has a lot of probability mass around his (higher) mean. At the same time, and for the same reason, the good type wants to decrease noise.

Next, assume an equilibrium features $\sigma_b < \sigma_g$. If this is the case, the principal will respond by retaining the agent whose signal is further away from $\hat{x} = (\sigma_b^2\theta_g - \sigma_g^2\theta_b)/(\sigma_b^2 - \sigma_g^2)$, which is the value that minimizes the likelihood ratio $\frac{1}{\sigma_g}\phi\left(\frac{x-\theta_g}{\sigma_g}\right)/\frac{1}{\sigma_b}\phi\left(\frac{x-\theta_b}{\sigma_b}\right)$, and it is to the left of θ_b. Then, it turns out that the best way for the bad type to escape from defeat is to inject additional noise (whereas it is unclear whether the good type wants to increase noise or precision), so $\sigma_b > \underline{\sigma}$. This, together with the fact that the conjectured equilibrium features $c'(\sigma_b)\sigma_b > c'(\sigma_g)\sigma_g$ (see the proof), implies that $\sigma_b > \sigma_g$, and hence a contradiction.

Our result bears a broad resemblance to Hvide (2012), who studies tournaments with moral hazard, when agents can influence both the mean and spread of their output. In equilibrium, there is excessive risk taking. By setting an intermediate value for output and rewarding the agent who gets closer to this threshold, the principal can do better.

7.4. **Mean-Shifting Effort.** We can easily augment the baseline model to include effort to shift the mean value of one's type. For instance, suppose that each agent k is endowed with some baseline value (or type) $\underline{\theta}_k$ (with $\underline{\theta}_g > \underline{\theta}_b$). He can augment θ using a cost function $d(\theta_k - \underline{\theta}_k)$, common to both types, where d defined on $\mathbb{R}_+$ is increasing, strictly convex and differentiable, with $d'(0) = d(0) = 0$. The signal sent is then given by $x_k = \theta_k + \sigma_k\varepsilon$. Finally, the principal makes a decision to retain or replace.

Parts of this model fully parallel our setting. The principal makes her decisions on the basis of conjectured means and variances chosen by each type, leading to the familiar conditions (9)–(11) for the retention edge-points x_- and x_+. Similarly, an agent of type k maximizes the probability of retention net of cost. Whether or not x_- is smaller or larger than x_+ (and even when $x_+ = \infty$ as it will be with monotone retention), the agent always maximizes $\Phi([x_+ - \theta_k]/\sigma_k) - \Phi([x_- - \theta_k]/\sigma_k) - c(\sigma_k)$, but this time by choosing *both* σ_k and θ_k. The first-order condition for σ_k is unchanged; what this extension adds is a first-order condition for θ_k, given by

$$\frac{1}{\sigma_k}\phi\left(\frac{x_- - \theta_k}{\sigma_k}\right) - \frac{1}{\sigma_k}\phi\left(\frac{x_+ - \theta_k}{\sigma_k}\right) \leq d'(\theta_k - \underline{\theta}_k), \tag{20}$$

with equality holding if $\theta_k > \underline{\theta}_k$. This additional condition can be used to show that the extension *fully* mimics the original model: we must have $\theta_b < \theta_g$, with other choices of noise and principal decisions just as in our baseline setting; see Online Appendix for details.

This extension is also useful for understanding other aspects of the noisy relationship between principal and agent. For instance, mean-shifting effort for the sake of retention could be directly valuable to the principal, apart from providing information about type.[17] If neither that effort nor the payoff-relevant "output" from it is contractible, then the principal could want to structure her environment to keep agent effort high. Of particular interest is the case in which the background noise $\underline{\sigma}$ is close to zero, so that the agents can communicate their types with very high precision.

[17] For other models of relational contracts in which effort provides both current output and information about match quality, see, Kuvalekar and Lipnowski (2018), Kostadinov and Kuvalekar (2018) and Bhaskar (2017).

In general, this limit model has several equilibria, some pooling and some separating. To see the issue that arises, let's concentrate on a particular parametric configuration in which $\underline{\theta}_g$ and $\underline{\theta}_b$ are sufficiently separated from each other so that

$$d\left(\underline{\theta}_g - \underline{\theta}_b\right) > 1. \tag{21}$$

In this case it is easy to see that there can be only separating equilibria in zero-ambient-noise limit. In each such equilibrium, the bad type exerts no effort whatsoever. *The principal cannot incentivize the agent because there is no noise in the signal.* Both types reveal themselves perfectly. There are still many equilibria possible in which the good type is forced to exert effort to raise θ_g beyond $\underline{\theta}_g$, simply because the principal's retention set is some singleton $\{\theta_g\}$ with $\theta_g > \underline{\theta}_g$. But these equilibria are shored up by the "absurd belief" that observations between $\underline{\theta}_g$ and θ_g are attributable to the bad type. These configurations can be eliminated by standard refinements, leaving only the least-cost separating equilibrium in which retention occurs if $x = \underline{\theta}_g$, and no agent exerts any effort at all. Condition (21) guarantees that the bad type will not want to mimic the good type in this case.

If mean-shifting effort is separately valuable to the principal, this outcome is undesirable to her. The solution will therefore involve the *principal* adding noise, thereby ensuring that the bad type has some chance of being retained, and so incentivizing him. In any equilibrium of such an extended model in which the principal can move first, the principal will choose $\underline{\sigma} > 0$, endogenously injecting noise into the system.

7.5. **A Variant With Costless Noise.** Our results assume a smoothly convex noise cost function. Of course, this assumption is consistent with the choice of noise being essentially costless over a wide range, as long as the cost mounts up at "either end." In this section, we consider a simple but attractive variant of our model: suppose that any level of noise can be costlessly chosen, as long as it is no smaller than $\underline{\sigma}$. Any choice smaller than $\underline{\sigma}$ is impossible. The condition $\underline{\sigma} > 0$ is a minimal requirement for the problem to have any interest: otherwise, the high type can always reveal himself by choosing $\sigma_g = 0$, and there is nothing to discuss. Let's call this the *costless noise* model.

This costless noise variant admits a particularly sharp solution. Define a function $\alpha\left(\beta\right)$ by

$$\beta \equiv \frac{1}{\alpha\left(\beta\right) + \sqrt{1 + \alpha\left(\beta\right)^2}} \exp\left[-\frac{\alpha\left(\beta\right)}{\alpha\left(\beta\right) + \sqrt{1 + \alpha\left(\beta\right)^2}}\right]. \tag{22}$$

Notice that $\alpha(\beta)$ is well-defined, that $\alpha(\beta) > 0$ for all $\beta < 1$ and $\alpha(\beta) \to 0$ as $\beta \to 1$. We will assume that $\underline{\sigma}$ is small enough so that:

$$\frac{\underline{\sigma}}{\theta_g - \theta_b} < \frac{1}{2}\alpha(\beta)^{-1} \text{ if } 0 < \beta < 1. \tag{23}$$

and

$$\frac{\underline{\sigma}}{\theta_g - \theta_b} < \left[\sqrt{2\ln\left(\beta\right)}\right]^{-1} \text{ if } \beta > 1 \tag{24}$$

Notice how these conditions become progressively weaker as we converge to the balanced case from either direction (i.e., as p and q get close to each other). *At* or near the balanced case, no restrictions are imposed at all; both right-hand side terms in (23) and (24) diverge to infinity.

One annoying price to pay for this simpler model is that without an upper bound to noise, there could be a fully uninformative in which both types babble with infinite noise and the principal always retains or always replaces. We ignore such equilibria. Thus say that an equilibrium is *nontrivial* if the principal retains the agent for some signals and replaces for others.

Proposition 10. (i) *A nontrivial equilibrium exists if and only if* (23) *is satisfied, and when it exists, it is unique.*

(ii) *If* (24) *is also satisfied, then the nontrivial equilibrium involves bounded retention. In it, the good type chooses* $\sigma_g = \underline{\sigma}$, *the bad type chooses higher but finite noise* $\sigma_b > \sigma_g$, *and the principal employs a strategy of the form: retain if and only if the signal* x *lies in some bounded interval* $[x_-, x_+]$.

(iii) *In particular, in the balanced case or with an optimistic future,* (24) *trivially holds and the equilibrium must involve bounded retention.*

(iv) *If* (24) *happens to fail, then the nontrivial equilibrium involves a monotone retention regime, with both types choosing noise equal to* $\underline{\sigma}$. *The principal retains if and only if* $x \geq x^*(\underline{\sigma})$.

We relegate a formal proof to the Online Appendix.

In particular, Proposition 10 asserts that in the balanced case, there is a unique equilibrium with no restrictions at all on $\underline{\sigma}$; both (23) and (24) are vacuous. As in our baseline model, when $p = q$, only bounded retention equilibria are possible. More generally, suppose that $p \geq q$, which means that the situation is either balanced or has an optimistic future. Then Condition (24) imposes no restriction at all, and we will now argue that a nontrivial equilibrium must use bounded retention. If this assertion is false, then a nontrivial equilibrium must involve either bounded replacement or a monotone threshold. The former is easily dispensed with — with bounded replacement, either type would want to inject unboundedly high noise to minimize the chances of landing in the replacement zone.[18] As for the latter, suppose that the principal employs a single retention threshold given by $x^* \in (\theta_b, \theta_g)$. Then the good type wants to *minimize* noise in order to pull more probability mass into the retention region, whereas the bad type wants to increase noise. This is incompatible with a monotone retention regime. On the other hand, if $x^* > \theta_g$, both agents will react by wanting to inject *additional* noise. We are therefore left only with bounded retention, and the formal proof shows that such an equilibrium must exist under Condition (23).[19]

[18] The non-existence of bounded replacement survives more robust arguments which allow for a finite upper bound to the choice of noise. See Online Appendix for more details.

[19] It is possible that in equilibrium, the principal discards both types of agents irrespective of signal, simply because the option value of a new agent is too high and the minimum noise $\underline{\sigma}$ in the current environment too large. Condition (23), which bounds $\underline{\sigma}$, is necessary and sufficient for eliminating this possibility. Moreover, in any reasonable "general-equilibrium closure" of this model, the failure of (23) is absurd: if both types are let go, where would the optimism regarding a new agent come from in the first place? We formalize this argument in Section 7.6, when we endogenize p.

Finally, with a pessimistic future ($q > p$), the principal is wary of new hires and inclined to retain the current agent. Now Condition (23) is empty, and a nontrivial equilibrium always exists. Under a symmetric choice of noise, it is entirely possible that the retention threshold falls below θ_b. Faced with that low threshold, *both* types will want to reduce noise to the minimum possible, and now there is scope for a nontrivial equilibrium with monotone retention regime, in which both types choose noise $\underline{\sigma}$, while the receiver employs a single threshold $x^*(\underline{\sigma})$. That scope dwindles, however, when $\underline{\sigma}$ is small: the smaller it is, the more sharply is the receiver able to distinguish between good and bad types. Condition (24) on $\underline{\sigma}$ is necessary and sufficient for eliminating the monotone equilibrium.

7.6. **Dynamics With Term Limits.** So far we have studied a static setting, but at the same time we've hinted more than once that the "outside option probability" p could, in principle, be solved for in a dynamic setting. We study the case in which the agent has a two-period "term limit," after which he must be replaced. This is useful for applications to politics, and also — but perhaps in a more limited way — to situations in which the agent is an employee or a contracted expert, such as a fund manager. In what follows we study stationary equilibrium, in which every new agent of a given type takes the same action independent of history.

For noise σ_k for each player of type k, and for each realization x, the Bayes' update on q is

$$q(x) := \frac{q\pi_g(x)}{\pi(x)}, \tag{25}$$

where for each k, the density of signal x is given by $\pi_k(x) = (1/\sigma_k)\phi\left([x-\theta_k]/\sigma_k\right)$, and where $\pi(x) = q\pi_g(x) + (1-q)\pi_b(x)$ is the overall density of signal x.

We can use this information to calculate the lifetime payoff to the principal at the start of any new interaction. To this end, let $M(q') := q'U_g + (1-q')U_b$ be the expected payoff to the principal in any period when her prior (for that period) is given by q'. This prior equals q for a fresh draw from the pool at any date. At the end of the first term, a signal x is generated, and the prior q is updated to $q(x)$. At this stage, the principal decides whether or not to retain for one more period, after which the term limit kicks in.

If V denotes the normalized lifetime payoff to the principal starting from a fresh agent, we can define a *retention zone* X as the set of all x for which $(1-\delta)M(q(x)) + \delta V \geq V$. The lifetime value to the principal can then be expressed as

$$\begin{aligned} V &= (1-\delta)M(q) + \delta\int_X \left[(1-\delta)M(q(x)) + \delta V\right]\pi(x)dx + \delta\int_{X^c} V\pi(x)dx \\ &= (1-\delta)\left[q(1+\delta\Pi_g)U_g + (1-q)(1+\delta\Pi_b)U_b\right] + \delta\left[1-(1-\delta)\Pi\right]V, \end{aligned}$$

where $\Pi_k := \int_X \pi_k(x)dx$ is the type-dependent probability of retention, and $\Pi := q\Pi_g + (1-q)\Pi_b$ is the overall probability of retention. (The second equality above follows from the definition of M and (25).) Transposing terms, we see that V is a convex combination of baseline utilities U_g and U_b; i.e., $V = pU_g + (1-p)\,U_b$, where

$$p = \frac{q\,(1+\delta\Pi_g)}{1+\delta\left[q\Pi_g + (1-q)\,\Pi_b\right]}.$$

We can rewrite this expression to obtain a "general equilibrium formula" for the ratio β:

$$\beta = \frac{q}{1-q}\frac{1-p}{p} = \frac{1+\delta\Pi_b}{1+\delta\Pi_g}. \tag{26}$$

Now observe that in any equilibrium, $\Pi_g \geq \Pi_b$. That has to be the case, because the principal can — and will — choose a retention zone that retains the high type at least as often than the low type. Indeed, it is not even possible to have β *equal* to 1 in any equilibrium.[20]

This setup reveals a clear strategy to solve the two-term dynamic extension of our model. For some (provisionally given) value of β, we obtain the baseline static model. Solve for the equilibrium there. That equilibrium will generate retention probabilities Π_g and Π_b. The circle is closed by the additional condition that (β, Π_g, Π_b) must solve (26).

Our costless noise variant in Section 7.5 is particularly amenable to solving for the details. In that model, noise is costless but bounded below by some number $\underline{\sigma} > 0$. In this setting, we have:

Proposition 11. *When agents can be hired for up to two terms, and the principal always has the option to replace agents with a new draw from a stationary pool, there is a unique equilibrium which has all the properties of the non-trivial equilibrium identified in Proposition 10. In particular, there are no trivial equilibria. Moreover, this unique equilibrium must endogenously display an optimistic future and conditions (23) and (24) do not need to be assumed.*

Proposition 11 says that in a dynamic extension of the model in Section 7.5 with a two-term limit, the equilibrium picks out precisely the two-threshold equilibrium with bounded retention regime, as described in Proposition 10 of the static model. Observe that that equilibrium in the static model does not always exist; after all, $\underline{\sigma}$ needs to be small enough as described in conditions (23) and (24). Those conditions are automatically satisfied here. So Proposition 11 is not just a mere refinement of the static equilibrium that eliminates all monotone and trivial equilibria. It does that, to be sure, but in addition it guarantees that for *any* value of $\underline{\sigma} > 0$, the dynamically determined value of p must adjust itself so that conditions (23) and (24) are automatically met.

7.7. **A Remark on Commitment.** To what degree are the results altered if the principal can commit ex ante to a retention zone? We do not have a full answer to this question, though it appears that the main findings would be unaffected. In the simple model of costless noise developed in Section 7.5, it turns out that the results are not affected *at all*.

Suppose that the realization x is contractible, and that the principal announces an incentive-compatible mechanism that specifies the retention probability for each value of x, and for each (declared) type of agent. The agent can then choose one of the rules — revealing his type — and then a noise level. We assume that the rule, given by $r_k(x) \in [0, 1]$ is piece-wise continuous.[21]

[20]Suppose $\beta = 1$. Then $p = q$, and we know that in the static model only bounded retention equilibria are possible. But in that situation the principal can *strictly* discriminate in favor of the good type, since there will always exist two distinct real roots to (4). But now $\Pi_g > \Pi_b$, which contradicts our starting point that $\beta = 1$.

[21]We conjecture that Proposition 12 below is true for all measurable functions.

For any type k, rule r and chosen noise σ, define

$$\rho_k(r,\sigma) := \frac{1}{\sigma}\int_{-\infty}^{\infty} r(x)\phi\left(\frac{x-\theta_k}{\sigma}\right)dx,$$

which is to be interpreted as the overall retention probability for type k when the retention function is r and he chooses noise σ. The principal seeks to maximize her surplus

$$q\rho_g\left(r_g,\sigma_g\right)\left(U_g - V\right) - (1-q)\,\rho_b\left(r_b,\sigma_b\right)\left(V - U_b\right) \tag{27}$$

by "choosing" r_k and σ_k for $k = g, b$, subject to

$$\sigma_k \in \arg\max_{\widetilde{\sigma}\geq\underline{\sigma}} \rho_k\left(r_k,\widetilde{\sigma}\right) \tag{28}$$

and

$$\rho_k\left(r_k,\sigma_k\right) \geq \max_{\widetilde{\sigma}\geq\underline{\sigma}} \rho_k\left(r_\ell,\widetilde{\sigma}\right). \tag{29}$$

for each k and $\ell \neq k$. The first of these constraints is the familiar choice of noise, and the latter comes from truthful revelation of type. But notice that this latter constraint cannot be slack for type b at the optimum. If it were, the principal could simply reduce the retention probability r_b[22] — which makes her happier (the expression in (27) goes up), continues to respect (29) for type b, and does no damage to (28) and (29) for type g.

We must conclude, therefore, that (29) binds for type b; that is, $\rho_b(r_b,\sigma_b) = \rho_b(r_g,\sigma_b')$, where σ_b' maximizes $\rho_b(r_g,\tilde{\sigma})$. Using (27), this further implies that the principal is completely indifferent between type b reporting his type and facing r_b, or misreporting his type and facing r_g. So, without any loss of generality, the principal may as well offer the agent a *single* retention function $r(x)$. That gives rise to a new problem with just one rule, no self-selection constraint (29) for either type, and just payoff maximization (28) for each type. To summarize this new problem, note that by definition of p, $V - U_b = (U_g - U_b)\,p$ and $U_g - V = (U_g - U_b)(1-p)$. Using these in (27), the principal equivalently maximizes

$$\beta\rho_g\left(r,\sigma_g\right) - \rho_b\left(r,\sigma_b\right), \tag{30}$$

where β is $q(1-p)/p(1-q)$ as defined earlier, and where for each $k = g, b$,

$$\sigma_k \in \arg\max_{\widetilde{\sigma}} \rho_k\left(r,\widetilde{\sigma}\right). \tag{31}$$

A single retention rule notwithstanding, there is still room for commitment, because the principal can influence the choice of noise. Yet in the context at hand, the principal has no use for it:

Proposition 12. *Assume condition (23), so that a nontrivial equilibrium exists in the costless noise model. Then an optimal contract involves the same retention function (a.e.) and the same values σ_b^* and σ_g^* as in the nontrivial equilibrium of Proposition 10.*

That is, the solution to the principal's problem with commitment is the same as the no-commitment or equilibrium solution, in the special case of the model with costless noise. For a similar result in a different context (and for distinct reasons), see Glazer and Rubinstein (2004, 2006) and Hart, Kremer and Perry (2016).

[22] She can judiciously remove intervals where $r_b(x) > 0$ to drive retention probability continuously from ρ_b to 0.

7.8. **Two Related Formulations.** Under Bayesian persuasion, a theme pursued in Kamenica and Gentzkow (2011), a principal observes a signal sent by the agent and uses Bayes' rule to update her prior. The agent wants to choose a signaling structure to maximize the chances that the receiver's posterior will cross a certain threshold (in our case, a threshold probability that the sender is of an acceptable type). In this setting, and in contrast to ours, it is presumed that the sender does *not* know his type before he chooses the signal structure, and he cannot re-optimize after knowing his type. In addition, the sender's choice of structure is observed by the receiver.

A second variant of the model is one in which the agent does possess private information about his type, but the principal can directly observe the agent's choice (of signal structure); see Degan and Li (2016).[23] In this variant, the agent's observed choice of risk can directly reveal information about his type, over and above the realization of that risk. In a separating equilibrium, then, signal realizations convey no additional information, because type separation has already been achieved via the observed choice of noise. When agents pool, signal realizations do matter, but retention is indeed monotonic in the signal. This is obviously a very different setting from that of the model studied here. See our comments on Dasgupta and Prat (2006) in Section 8.1 for more detail in the context of a specific application.[24]

Our model generates distinct behavior, by virtue of the fact that *both* the type and the signal structure are unobserved by the principal. These three formulations all have distinct applications — we discuss some applications of our model in Section 8 to follow — but overall, the three models apply to different situations, so the preference for any one over the others would depend on the real-world situation at hand.

8. Applications

Our model separates three distinct features: the action (or the choice of risk), the realization of the signal, and the subsequent inference and decision of the principal. A central implication of the model is that the realizations may be "good" — even in the sense of generating high payoffs for the principal today. At the same time, they could serve as a cautionary indicator for a great deal of risk-taking by the agent, which may well generate a negative inference in agent ability. This may sound contradictory, but as long as we properly separate the current payoff-relevance of a signal realization from its role *qua* signal, there is no inconsistency here.

We discuss three applications: to risky portfolio management (Section 8.1), to the behavior of organizations that seek donor funding (Section 8.2), and to the actions of political leaders (Section 8.3). In each case it should be noted that (a) the choice of action by the agent corresponds to a choice of risk, (b) it is reasonable to suppose that such risk cannot be fully understood (i.e., observed) ex ante by the principal, and (c) the outcome, apart from being intrinsically good or

[23] They also study situations in which signal precision is chosen before agent type is realized; this is closer to the Kamenica-Gentzkow setting, though precision is constrained to be type-independent.

[24] For a related exercise, see Titman and Trueman (1986), in which observed auditor quality is used to signal firm valuation during an initial public offering. (Higher-quality auditors provide more precise information, by assumption.) An entrepreneur with more favorable private information about the value of his firm will choose a higher-quality auditor than will an entrepreneur with less favorable private information.

bad, serves as an indicator for the extent of risk-taking, thereby leading to some form of inference about the agent's competence.

8.1. **Risky Portfolio Managers.** An investment advisor (the agent) manages your (the principal's) portfolio, which yields an uncertain return. Assume you have no idea which stocks are likely to yield good returns: if you did, presumably you would be managing your money yourself. You do have an overall idea of the *distribution* of returns, though, perhaps calibrated to the ex post performance of stock indices or various mutual funds. Managers have varying knowhow about the prospects of various stocks, summarized in some mean return (their type).[25] But they can also take (mean-preserving) risks. You observe just the net return x on the portfolio built for you by the manager. You might observe the portfolio as well, but if you know very little about individual stocks this will mean little or nothing to you. In short, *you cannot use the portfolio to judge the wisdom of the agent's strategy*. That means that risk-taking is unobservable.[26]

Our model then tells us that bad managerial types will endogenously load up on financial risk so as to try and achieve good returns. So you, the principal, should be suspicious not just when returns are low, but *also when returns are excessively high*. It is not because you fear risk-taking *per se*; in this setting you only care about expected payoffs. The central point is that there are career concerns at work which manifest itself through excessive risk-taking.[27]

The specific point that uninformed managers might actively trade is not new. Dasgupta and Prat (2006) consider a setting in which a fund manager faces career concerns. He trades — or passively holds — an asset for a principal who decides whether to retain the manager or not. The good managers know the precise value of the asset, to be later revealed to all. The bad manager is uninformed. The principal wants to retain the good manager, and replace the bad manager. Since the principal observes whether the manager sold, bought, or did nothing, there is an incentive for the bad manager to trade. Because that manager doesn't know what the real value of the asset is, he randomizes between selling and buying, or "churns."

In this setting, the manager's *action* is observed and assessed by the principal. In ours, we effectively assume that the action (risk-taking) isn't observed, for the reasons discussed above. This distinction is important. If actions are observed and can be interpreted — e.g., if buying or selling is known to be generically optimal — then a separating equilibrium cannot exist in which the bad type does nothing. That consideration, by itself, is enough to deliver churning.

That is why a good outcome is always a reason for retention in the Dasgupta-Prat analysis. A principal does not discard a manager who performs "too well," under the suspicion that he has

[25] So our model fits well but not perfectly. Ideally, we would allow managers to choose both the mean and the variance of the portfolio. The latter could well be costless — imagine loading pure risk on by the use of options, for instance. The former would require costly effort that would vary with manager type. This extended model — in a dynamic setting — is pursued in McClellan and Ray (2018). The use of risky gambles by managers with career concerns is studied in a dynamic setting by Makarov and Plantin (2015).

[26] Of course this is an exaggeration, as some excessively risky ventures may be commonly understood to be risky. Eliminate these from consideration here.

[27] It is possible to argue that a manager has access to a wider set of distributions that those achieved through some risk ordering. For instance, the manager could sell call options to truncate his returns at the top. But even with this wider range there cannot exist a monotone equilibrium as long as the manager can *feasibly* load up on risk.

been engaged in excessive risk-taking. We do not claim any empirical support for this prediction, though anecdotal evidence on financial ventures or Ponzi schemes that promise (and initially deliver) high rates of return suggests that careful individuals often stay away from such ventures — to be sure, others don't. We simply put this forward as a necessary corollary of our result, one that perhaps deserves some empirical scrutiny.

8.2. **Fundraising.** Consider a non-governmental organization (or NGO) of unknown competence, seeking funding from donors. An NGO of low competence could take on a safe project. For instance, if it is a microfinance organization, it could move into tested localities, or it could simply emulate what other pioneer microfinance entities have done: for instance, only providing working capital loans, which maintains the ongoing dependence of the borrower on the NGO and serves as an incentive device for repayment. Or the NGO could move into uncharted locations (and in new ways), where knowledge of ground realities is more dispersed and the risks are far greater. For instance, it could start making fixed capital loans, where the payoffs could be large (unlike in the case of working capital, fixed-capital loans can start new businesses). At the same time, such a project could bomb quite spectacularly if there is widespread default.

The donor has no local knowledge, and so cannot observe these risks ex ante. As per our model, it can form expectations about what NGOs with different competencies might try out, and react accordingly after reading the description of NGO achievements in its funding proposal. Those achievements may well be payoff-relevant to the donor, but her concerns lie with whether the NGO should be funded for *future* activities, and for this purpose those achievements serve only as signals, and a very unusual success may be an indicator of excessive risk-taking.

Again, the choice of action — do something standard, versus attempt something "creative" — can be proxied by different levels of risk-taking. Donors, unaware of ground realities, may not be aware of all the attendant risks, and so do not observe the action per se. The signal realization is what the NGO achieves. The inference attempts to separate current success from what it says about future performance. While not pretending that our equilibrium describes how NGOs and donors interact (though by all means not suggesting that it doesn't either), there is some evidence that these considerations do motivate NGOs. In their study of locational choices among NGOs in Bangladesh, Fruttero and Gauri (2005, p. 778–779) write:

"The regressions showed that brand NGO programmes were moving to the same places as new government programmes, and that non-brand NGO programmes were not. Brand NGOs, in other words, were not substituting for government programmes but instead following them. This is consistent with a reputation-building story for new NGOs. Brand NGOs that have established a reputation of high ability have more to lose in choosing areas without other programmes. Hence their decision may be driven by the desire to minimise the risk. On the other hand, non-brand NGOs, which need to single themselves out, go to areas in which they can show their ability."

8.3. **Risky Politics.** Think of a political leader, the assessment of whose competence is currently important, and who seeks to be "retained" by the median voter (who plays here the role of the principal). If that leader is competent, he can attempt to play it safe by implementing unambitious policies, and so the better will be the fix that the public obtains about his true type — though convergence to that understanding may be far from total. In contrast, the incompetent leader can entertain a risky policy: for instance — and only speaking hypothetically — he might attempt to conduct a denuclearization summit with the authoritarian leader of a rogue state. Such a summit is obviously fraught with risk, though it is may be impossible for the public (or the median voter) to assess, ex ante, just how much risk there is. If that project is successful, an otherwise incompetent politician might stand a chance for re-election. Barring that, there is no chance.

Our model suggests that a striking success from such a policy — if, continuing the hypothetical streak, one were to occur — should be treated with a certain degree of reticence by the median voter. It could be a sign of extreme competence. It could also be sign of a desperate move by a largely incompetent individual, which happened to pay off. That outcome, if it occurs, may be good for society. But it may not be a good signal on which to base re-election.

The same is generally true of ambiguity and obfuscation in the political arena. A bad politician cannot fully imitate a good politician. What he can do is to try to look like a good politician to the extent possible. In addition to a concrete policy as described above, that could also imply being ambiguous, inconsistent and generally obscure about attributes, intentions, and policy agenda, or encouraging the spread of fake news. A potential voter might interpret the vague messages in many different ways, both positive and negative.

Under this interpretation, the agent's action is the choice of ambiguity and obfuscation, which we model as a choice of risk (regarding the extent of public approval of those messages, or regarding the verification or debunking of claims made). So in this example, a particular signal realization can be identified with the public's interpretation of the politician's messages.[28] It might appear that there is a paradox between that interpretation being a positive one and the subsequent reticence that the model calls for on the part of the median voter. But — as in the other applications as well — there is no paradox: a good current outcome might still be a reliable indicator for excessive risk-taking by the agent.

9. Summary

We've studied a model in which an agent who seeks to be retained by a principal might deliberately inject noise into a process that signals his type. Possible equilibrium regimes include monotone retention, in which a principal retains if an agent's signal is high enough, and various non-monotone regimes. Of these, we argue that *bounded retention* is the salient equilibrium regime. In it, different types of agents choose different degrees of noise, with worse agents behaving more noisily. The resulting equilibrium has a "double-threshold" property: the principal retains the agent if the signal is good, but neither too good nor too bad. We discuss extensions to non-normal signal structures, non-binary agent types, multiple agents each with privately known types, situations in which the agent can shift the means of their signals, at an additional cost, a dynamic version of the model with agent term limits, and environments in which the principal can ex ante commit to retention rules.

We believe that the deliberate injection of ambiguity or noise is a central feature of many principal-agent interactions. We have discussed some of them: risky portfolio management,

[28] To the extent that such interpretations are themselves stochastic, this also justifies the assumption of some ambient noise, that is, $\underline{\sigma} > 0$.

fundraising by new NGOs, and politics. We make the central assumption that the extent of noise cannot be directly observed by the principal, and must be inferred. Even though this assumption might be violated in some settings, there are many other situations where the receiver does not ex ante fully understand the risk or the full set of all possible options available to the agent. By modeling these situations as constraints on the observability of risk, our framework makes a contribution towards the understanding of such environments.

Appendix: Some Proofs and Missing Details

Details in the Proof of Proposition 1. Recall (7) from the main text; this is the equation that σ must satisfy if it commonly chosen by both types:

$$\phi(z_1)z_1 = \phi(z_2)z_2 = -\sigma c'(\sigma), \tag{32}$$

where $z_1 = (\sigma/\Delta)\ln(\beta) - (\Delta/2\sigma)$ and $z_2 = (\sigma/\Delta)\ln(\beta) + (\Delta/2\sigma)$. The function $\phi(z)z$ has the shape shown in Figure 3, reaching maxima and minima at $z = 1$ and $z = -1$ respectively, and exhibiting "negative symmetry" around 0. Using (7), this tells us that there are two exclusive possibilities: (i) either $\beta > 1$ and $\sigma < \underline{\sigma}$, or (ii) either $\beta < 1$ and $\sigma > \underline{\sigma}$. We study (i); Case (ii) is dealt with in the same way.

In Case (i), elementary computation shows that z_2, viewed as a function of σ (holding all other terms constant) starts from infinity as $\sigma = 0$, declines to a minimum of $\sqrt{2\ln(\beta)}$, and then climbs monotonically again to ∞ as $\sigma \to \infty$. Meanwhile, z_1 is always increasing in σ, and is exactly zero when z_2 reaches its minimum. From this point on, $\phi(z_1)z_1$ climbs from 0 to its maximum value of $\phi(1)$ and then falls, while $\phi(z_2)z_2$ falls monotonically from a positive value to zero. Finally, we note that in the phase where $\phi(z_1)z_1$ falls, we have $\phi(z_1)z_1 > \phi(z_2)z_2$ throughout. Putting these observations together, we must conclude that there is a *unique* value of σ such that the *first* equality in (32) holds, and it is independent of the cost function c. ■

Proof of Proposition 3. Recall that $\sigma_* < \underline{\sigma}$ and $\sigma^* > \underline{\sigma}$ are the two solutions to $c(\sigma) = 1$. Let $\Sigma := [\sigma_*, \sigma^*]^2$, and define

$$\Sigma^+ := \{(\sigma_g, \sigma_b) \in \Sigma | \sigma_b \geq \sigma_g\}.$$

For each $\boldsymbol{\sigma} \in \Sigma^+$, define x_- and x^+ by the distinct lower and upper roots to (4) if $\sigma_b > \sigma_g$; otherwise, if $\sigma_b = \sigma_g = \sigma$, set $x_- = x^*(\sigma)$ as defined in (5) and $x_+ = \infty$. Interpret $[x_-, x_+]$ as the retention zone. Call this map Ψ_1. As discussed in the main text, this map is well-defined when $\sigma_b = \sigma_g$. To check that Ψ_1 is also well-defined when $\sigma_b > \sigma_g$, we must show that there are two distinct real roots to the quadratic in (4), or equivalently, using the elementary formula for quadratic roots, that the expression

$$\Delta^2 + \left(\sigma_b^2 - \sigma_g^2\right) 2\ln\left(\beta\frac{\sigma_b}{\sigma_g}\right)$$

is strictly positive. But (14) tells us that $\ln(\beta) \geq -[\Delta^2]/2\sigma^{*2}$, and so

$$\begin{aligned}\Delta^2 + \left(\sigma_b^2 - \sigma_g^2\right) 2\ln\left(\beta\frac{\sigma_b}{\sigma_g}\right) &= \Delta^2 + \left(\sigma_b^2 - \sigma_g^2\right) 2\ln\left(\beta\frac{\sigma_b}{\sigma_g}\right) \\ &\geq \Delta^2 + \left(\sigma_b^2 - \sigma_g^2\right) 2\ln\left(\beta\right) \\ &\geq \Delta^2\left[1 - \frac{\sigma_b^2 - \sigma_g^2}{\sigma^{*2}}\right] > 0,\end{aligned}$$

where the very last inequality uses $\sigma^* \geq \sigma_b > \sigma_g$. So there are distinct roots $x_- < x_+$, and by exactly the same logic as for Proposition 2, the zone $[x_-, x_+]$ must involve retention.

Next, for each pair (x_-, x_+) with $x_+ > x_-$ and with x_+ possibly infinite, define (σ_b', σ_g') to be the best-response choices of noise by the bad and good types who face the retention zone $[x_-, x_+]$. By condition [U], these choices are well-defined and unique. Call this map Ψ_2.

Finally, define a map Ψ with domain Σ^+ and range Σ by $\Psi := \Psi_2 \circ \Psi_1$. We claim that Ψ is continuous. We first argue that Ψ_1 is continuous in the extended reals. That is:

(i) if $(\sigma_g^n, \sigma_b^n) \to (\sigma_g, \sigma_b)$ with $\sigma_b > \sigma_g$, then $\Psi_1(\sigma_g, \sigma_b) = (x_-, x_+)$ with $x_- < x_+ < \infty$, and it is obvious that $\Psi_1(\sigma_g^n, \sigma_b^n) \to \Psi_1(\sigma_g, \sigma_b)$.

(ii) if $(\sigma_g^n, \sigma_b^n) \to (\sigma_g, \sigma_b)$ with $\sigma_b = \sigma_g$, then $\Psi_1(\sigma_g, \sigma_b) = (x_-, \infty)$. In this case, an inspection of the quadratic condition (4) (the roots of which yield x_- and x_+) reveals that $\Psi_1(\sigma_g^n, \sigma_b^n) = (x_-^n, x_+^n)$ must satisfy $x_+^n \to \infty$.

Now we turn to the map Ψ_2. As already mentioned, condition [U] guarantees that best-response noise choices are unique, as long as $x_+ > x_-$. They are fully characterized by the first-order condition (8), which we reproduce here for convenience:

$$\phi\left(\frac{x_- - \theta_k}{\sigma_k}\right)\left(\frac{x_- - \theta_k}{\sigma_k}\right) - \phi\left(\frac{x_+ - \theta_k}{\sigma_k}\right)\left(\frac{x_+ - \theta_k}{\sigma_k}\right) = \sigma_k c'\left(\sigma_k\right) \tag{33}$$

where we include the possibility that $x_+ = \infty$ by setting $\phi(z)z = 0$ when $z = \infty$.

Pick any sequence (x_-^n, x_+^n) that converges in the extended reals. That is, either the sequence converges to (x_-, x_+) with $x_+ < \infty$, or it converges to a limit of the form (x_-, ∞). Let σ_k^n be the best responses for an agent of type k, and let σ_k be the best response at the limit value (x_-, x_+). When $x_+ < \infty$, it is obvious from (33) that $\sigma_k^n \to \sigma_k$. In the latter case, the fact that $\sigma_k^n \to \sigma_k$ follows from the additional observation that $\phi(z^n)z^n \to 0$ for any sequence $z^n \to \infty$.

We claim that Ψ is *inward pointing*; that is, for every $(\sigma_g, \sigma_b) \in \Sigma^+$, there exists $a > 0$ such that

$$(\sigma_g, \sigma_b) + a[\Psi(\sigma_g, \sigma_b) - (\sigma_g, \sigma_b)] \in \Sigma^+. \tag{34}$$

First observe that for every $(\sigma_g, \sigma_b) \in \Sigma^+$, we have $(\sigma_*, \sigma_*) \leq \Psi(\sigma_g, \sigma_b) \leq (\sigma^*, \sigma^*)$. Therefore, if $(\sigma_g, \sigma_b) \in \Sigma^+$ with $\sigma_b > \sigma_g$, (34) is easily seen to hold: for $a > 0$ and small, it must be that both components of the vector

$$(\sigma_g, \sigma_b) + a[\Psi(\sigma_g, \sigma_b) - (\sigma_g, \sigma_b)]$$

lie in $[\sigma_*, \sigma^*]$, and the second component is larger than the first. The remaining case is one in which $(\sigma_g, \sigma_b) \in \Sigma^+$ with $\sigma_b = \sigma_g$. In this case, we know from condition (14) that $\Psi_1(\sigma_g, \sigma_b)$ is of the form $(x_-, x_+) = (x^*, \infty)$, where $x^* \in [\theta_b, \theta_g]$. From the first-order conditions that describe each type — see (6) — it is easy to see that $\sigma_k \gtreqless \underline{\sigma}$ when $x^* \gtreqless \theta_k$. Therefore $\Psi_2(x^*, \infty) = (\sigma'_g, \sigma'_b)$ must have the property that $\sigma'_b > \sigma'_g$ (and of course each component lies between σ_* and σ^*). It follows that for every $a \in (0, 1)$, (34) holds, and the claim is proved.

To summarize, we have: Σ^+ is a nonempty, compact, convex subset of Euclidean space, and Ψ is continuous on Σ^+. In general, however, Ψ will fail to map from from Σ^+ to Σ^+. However, the map is *inward pointing* in the sense of Halpern (1968) and Halpern and Bergman (1968); for an exposition, see Aliprantis and Border (2006, Definition 17.53). By the Halpern-Bergman fixed point theorem (see Aliprantis and Border 2006, Theorem 17.54), there exists $(\sigma_g, \sigma_b) \in \Sigma^+$ such that $\Psi(\sigma_g, \sigma_b) = (\sigma_g, \sigma_b)$. It is easy to see that (σ_g, σ_b), along with the associated bounded retention zone $\Psi_1(\sigma_g, \sigma_b)$, forms an equilibrium. ∎

Lemma 1. *In any equilibrium:* (i) *if* $\sigma_b > \sigma_g$ *then* $x_+ > \frac{x_+ + x_-}{2} > \theta_g$*, and* (ii) *if* $\sigma_b < \sigma_g$ *then* $x_+ < \frac{x_+ + x_-}{2} < \theta_b$.

Proof. When $\sigma_b \neq \sigma_g$, x_- and x_+ are both finite and given by (4). It is easy to check that

$$\frac{x_+ + x_-}{2} = \frac{\sigma_b^2 \theta_g - \sigma_g^2 \theta_b}{\sigma_b^2 - \sigma_g^2}.$$

So if $\sigma_b > \sigma_g$ then $x_+ > \frac{x_+ + x_-}{2} > \theta_g$ and if $\sigma_b < \sigma_g$ then $x_+ < \frac{x_+ + x_-}{2} < \theta_b$. ∎

Lemma 2. *In any equilibrium with finite values for* x_- *and* x_+ *and for either type* k*,*

$$\phi\left(\frac{x_- - \theta_k}{\sigma_k}\right) > \phi\left(\frac{x_+ - \theta_k}{\sigma_k}\right). \tag{35}$$

Proof. Suppose first that $\sigma_b > \sigma_g$. By Lemma 1(i), $(x_+ + x_-)/2 > \theta_k$ and so

$$\frac{x_+ - \theta_k}{\sigma_k} > \frac{\theta_k - x_-}{\sigma_k}$$

which implies, using the single-peakedness and symmetry of ϕ around 0, along with the fact that $x_+ > x_-$ in this case, that

$$\phi\left(\frac{x_+ - \theta_k}{\sigma_k}\right) < \phi\left(\frac{\theta_k - x_-}{\sigma_k}\right) = \phi\left(\frac{x_- - \theta_k}{\sigma_k}\right),$$

which establishes (35) for $\sigma_b > \sigma_g$. On the other hand, if $\sigma_b < \sigma_g$, then $(x_+ + x_-)/2 < \theta_k$ by Lemma 1(ii), so that

$$\frac{x_+ - \theta_k}{\sigma_k} < \frac{\theta_k - x_-}{\sigma_k}.$$

Once again, using the single-peakedness and symmetry of ϕ around 0, but this time the fact that $x_+ < x_-$, we must conclude that

$$\phi\left(\frac{x_+ - \theta_k}{\sigma_k}\right) < \phi\left(\frac{\theta_k - x_-}{\sigma_k}\right) = \phi\left(\frac{x_- - \theta_k}{\sigma_k}\right),$$

which establishes (35) for $\sigma_b < \sigma_g$, and so completes the proof. ■

Lemma 3. *In any equilibrium with bounded retention or replacement, so that $\sigma_b \neq \sigma_g$,*

$$\frac{1}{\sigma_b}c'(\sigma_b) > \beta\frac{1}{\sigma_g}c'(\sigma_g) \text{ and } \sigma_b c'(\sigma_b) > \beta\sigma_g c'(\sigma_g).$$

Proof. To prove the first assertion, combine the inequalities in (10) and (11), while invoking the two first-order conditions in (8), to conclude that

$$\begin{aligned}\frac{1}{\sigma_b}c'(\sigma_b) &= \phi\left(\frac{x_- - \theta_b}{\sigma_b}\right)\frac{x_- - \theta_b}{\sigma_b^3} - \phi\left(\frac{x_+ - \theta_b}{\sigma_b}\right)\frac{x_+ - \theta_b}{\sigma_b^3} \\ &> \beta\left[\phi\left(\frac{x_- - \theta_g}{\sigma_g}\right)\frac{x_- - \theta_g}{\sigma_g^3} - \phi\left(\frac{x_+ - \theta_g}{\sigma_g}\right)\frac{x_+ - \theta_g}{\sigma_g^3}\right] = \frac{\beta}{\sigma_g}c'(\sigma_g).\end{aligned}$$

To prove the second assertion, use (9) in (8) for the good type to obtain

$$\phi\left(\frac{x_- - \theta_b}{\sigma_b}\right)\left(\frac{x_- - \theta_g}{\sigma_b}\right) - \phi\left(\frac{x_+ - \theta_b}{\sigma_b}\right)\left(\frac{x_+ - \theta_g}{\sigma_b}\right) = \beta\sigma_g c'(\sigma_g), \tag{36}$$

and compare this to the first-order condition for the bad type, which is given by:

$$\phi\left(\frac{x_- - \theta_b}{\sigma_b}\right)\left(\frac{x_- - \theta_b}{\sigma_b}\right) - \phi\left(\frac{x_+ - \theta_b}{\sigma_b}\right)\left(\frac{x_+ - \theta_b}{\sigma_b}\right) = \sigma_b c'(\sigma_b) \tag{37}$$

Invoking (35) of Lemma 2, we see that the expression

$$\left[\phi\left(\frac{x_- - \theta_b}{\sigma_b}\right) - \phi\left(\frac{x_+ - \theta_b}{\sigma_b}\right)\right]\left[\frac{\theta_g - \theta_b}{\sigma_b}\right]$$

is strictly positive. But adding this term to the left-hand side of (36) yields the left-hand side of (37). We must therefore conclude that $\sigma_b c'(\sigma_b) > \beta\sigma_g c'(\sigma_g)$, and our proof is complete. ■

Proof of Proposition 4. The first assertion of the proposition is a simple consequence of (i)–(iii), to which we now turn. If (i) is false, then $c'(\sigma_b) \leq 0$ and $c'(\sigma_g) \geq 0$, which contradicts Lemma 3. For (ii), if $\sigma_g < \underline{\sigma}$ then $c'(\sigma_g) < 0$. Then Lemma 3 implies

$$\frac{c'(\sigma_b)}{c'(\sigma_g)} < \beta\min\left\{\frac{\sigma_b}{\sigma_g}, \frac{\sigma_g}{\sigma_b}\right\} < \beta,$$

so that $c'(\sigma_b)/c'(\sigma_g) < 1$ when $\beta \leq 1$. Rearranging (and keeping in mind that $c'(\sigma_g) < 0$), we have $c'(\sigma_b) > c'(\sigma_g)$, or $\sigma_b > \sigma_g$.

To prove (iii), assume $\sigma_g > \underline{\sigma}$. Then $c'(\sigma_g) > 0$, and Lemma 3 implies that

$$\frac{c'(\sigma_b)}{c'(\sigma_g)} > \beta\max\left\{\frac{\sigma_b}{\sigma_g}, \frac{\sigma_g}{\sigma_b}\right\} > \beta.$$

If $\beta \geq 1$, this inequality implies $\sigma_b > \sigma_g$. ■

Lemma 4. *Under (14) and (15), $x_+ < \theta_b < x_- < \theta_g$ in a bounded replacement equilibrium.*

Proof. Consider a bounded replacement equilibrium. Then $\sigma_g > \sigma_b$. Recall (4), which states that retention is strictly optimal if

$$(38) \qquad \left(\sigma_g^2 - \sigma_b^2\right) x^2 + 2\left(\sigma_b^2\theta_g - \sigma_g^2\theta_b\right) x + \left(\sigma_g^2\theta_b^2 - \sigma_b^2\theta_g^2 + 2A\sigma_g^2\sigma_b^2\right) > 0,$$

(where $A = \ln\left(\beta\sigma_b/\sigma_g\right)$), and replacement is strictly optimal if the opposite inequality holds. Putting $x = \theta_b$ in (38) and simplifying, we see that replacement is strictly optimal at θ_b if

$$\beta < \frac{\sigma_g}{\sigma_b} \exp \frac{\Delta^2}{2\sigma_g^2},$$

but this is guaranteed by the right hand inequality of (14), because $\sigma_g > \sigma_b$ and $\sigma_g \leq \sigma^*$. Therefore θ_b lies in the interior of the replacement zone, or put another way, $x_+ < \theta_b < x_-$.

Now putting $x = \theta_g$ in (38) and simplifying, we see that retention is strictly optimal at θ_g if

$$(39) \qquad \frac{\Delta^2}{2\sigma_b^2} + \ln\left(\sigma_b\right) - \ln(\sigma_g) > -\ln(\beta).$$

The derivative of the left hand side of (39) with respect to σ_b is given by

$$\frac{1}{\sigma_b}\left(1 - \frac{\Delta^2}{\sigma_b^2}\right)$$

which is strictly negative given (15) and $\sigma_b \leq \sigma^*$, so it follows that the left hand side of (39) is minimized by setting $\sigma_b = \sigma_g = \sigma^*$. To establish (39), then, it is sufficient to have

$$\frac{\Delta^2}{2\sigma^{*2}} > -\ln(\beta),$$

but this is guaranteed by the left hand inequality of (14). Consequently, the principal strictly prefers to retain the agent if she observes $x = \theta_g$. Given $x_+ < \theta_b < x_-$, this can only mean that $x_- < \theta_g$, and the proof is complete. ∎

Proof of Proposition 5. Suppose that a bounded replacement equilibrium exists. Then we have $\sigma_g > \sigma_b$ and $x_- > x_+$. By Lemma 4, we have $\theta_g \geq x_- \geq \theta_b > x_+$.

Define $B_k(\sigma)$ to be type-k's marginal benefit of noise:

$$(40) \qquad B_k(\sigma) := \phi\left(\frac{x_- - \theta_k}{\sigma}\right)\frac{x_- - \theta_k}{\sigma^2} - \phi\left(\frac{x_+ - \theta_k}{\sigma}\right)\frac{x_+ - \theta_k}{\sigma^2}.$$

That this is indeed the marginal benefit can be seen easily by recalling (8), which sets this expression equal to marginal cost. Observe that for *every* σ,

$$\begin{aligned}
B_b(\sigma) &= \phi\left(\frac{x_- - \theta_b}{\sigma}\right)\frac{x_- - \theta_b}{\sigma^2} - \phi\left(\frac{x_+ - \theta_b}{\sigma}\right)\frac{x_+ - \theta_b}{\sigma^2} \\
&\geq \phi\left(\frac{x_+ - \theta_b}{\sigma}\right)\frac{x_- - \theta_b}{\sigma^2} - \phi\left(\frac{x_+ - \theta_b}{\sigma}\right)\frac{x_+ - \theta_b}{\sigma^2} \\
&= \phi\left(\frac{x_+ - \theta_b}{\sigma}\right)\frac{x_- - x_+}{\sigma^2} \\
&> \phi\left(\frac{x_+ - \theta_g}{\sigma}\right)\frac{x_- - x_+}{\sigma^2} \\
&= \phi\left(\frac{x_+ - \theta_g}{\sigma}\right)\frac{x_- - \theta_g}{\sigma^2} - \phi\left(\frac{x_+ - \theta_g}{\sigma}\right)\frac{x_+ - \theta_g}{\sigma^2} \\
&\geq \phi\left(\frac{x_- - \theta_g}{\sigma}\right)\frac{x_- - \theta_g}{\sigma^2} - \phi\left(\frac{x_+ - \theta_g}{\sigma}\right)\frac{x_+ - \theta_g}{\sigma^2} \\
&= B_g(\sigma), \qquad (41)
\end{aligned}$$

where the first weak inequality follows from $x_- \geq \theta_b$ and inequality (35) of Lemma 2, the first strict inequality follows from ϕ single-peaked around zero and $x_+ - \theta_g < x_+ - \theta_b < 0$, and the last weak inequality follows from $x_- \leq \theta_g$ and (again) inequality (35) of Lemma 2.

But (41) leads to the following contradiction: if the marginal benefit of noise for the bad type strictly exceeds that for the good type at *every* noise level, then by a simple single-crossing argument, we must have $\sigma_b > \sigma_g$. But by Proposition 2, this contradicts the fact that we are in a bounded replacement equilibrium. ■

Proof of the Propositions in Section 7.1. We begin with a summary of some properties for densities f satisfying the strong MLRP.

Lemma 5. *Suppose that f satisfies strong MLRP. Then $f'(z)/f(z)$ is decreasing in z. In particular, f must be single-peaked, first strictly increasing and then strictly decreasing.*

Proof. Elementary differentiation of $f(z-a)/f(z)$ with respect to z establishes the result. ■

Lemma 6. *Pick any θ and θ', and any positive σ and σ'. Define for any x:*

$$h(x) \equiv \frac{f\left(\frac{x-\theta}{\sigma}\right)}{f\left(\frac{x-\theta'}{\sigma'}\right)}$$

(i) *If $\sigma = \sigma'$ and $\theta > \theta'$, then $h(x)$ is strictly increasing in x with $\lim_{x\to-\infty} h(x) = 0$ and $\lim_{x\to\infty} h(x) = \infty$.*

(ii) *If $\sigma > \sigma'$, then $\lim_{x\to\infty} h(x) = \lim_{x\to-\infty} h(x) = \infty$.*

Proof. Part (i): Define $z(x) \equiv (x - \theta')/\sigma$ and $a \equiv (\theta - \theta')/\sigma$. Then, because $\sigma = \sigma'$, we have

$$h(x) = \frac{f\left(z(x) - a\right)}{f\left(z(x)\right)}.$$

Because $z(x)$ is affine and increasing in x, the result follows directly from strong MLRP.

Part (ii): Observe that there exists $\epsilon > 0$ such that for x sufficiently large, $(x-\theta)/\sigma \leq (x - [\theta' + \epsilon])/\sigma'$. Moreover, for x large enough, f is decreasing (by Lemma 5). It follows that for all x so that both these conditions are satisfied,

$$\frac{f\left(\frac{x-\theta}{\sigma}\right)}{f\left(\frac{x-\theta'}{\sigma'}\right)} \geq \frac{f\left(\frac{x-[\theta'+\epsilon]}{\sigma'}\right)}{f\left(\frac{x-\theta'}{\sigma'}\right)},$$

and now, using part (i), the right hand side of this inequality goes to infinity as $x \to \infty$. The case $x \to -\infty$ follows parallel lines: switch (θ, σ) and (θ', σ') in the argument above, notice that f is increasing for x sufficiently negative (Lemma 5), and use part (i) again. ■

Proof of Proposition 6. Note that an equilibrium is monotone if and only if $\sigma_b = \sigma_g$. For if $\sigma_b = \sigma_g$, then Lemma 6(i) tells us that there exists x^* with

$$\beta f\left(\frac{x-\theta_g}{\sigma}\right) \geq f\left(\frac{x-\theta_b}{\sigma}\right) \tag{42}$$

if and only if $x \geq x^*$ (with strict inequality when $x > x^*$). So the principal retains whenever $x \geq x^*$. Conversely, if $\sigma_b \neq \sigma_g$, then the equilibrium cannot be monotone. Indeed, if $\sigma_g > \sigma_b$, then by Lemma 6(ii),

$$\beta f\left(\frac{x-\theta_g}{\sigma}\right) > f\left(\frac{x-\theta_b}{\sigma}\right)$$

for all x sufficiently large and positive, or sufficiently large and negative. But that means retention must occur for all such x, which proves that the replacement zone must be bounded. In similar vein, if $\sigma_b > \sigma_g$, then the retention zone must be bounded. This argument establishes not only that monotonicity is characterized by $\sigma_b = \sigma_g$, it also proves the boundedness of at least one zone of retention and replacement, and proves part (ii) of the Proposition.

To prove part (i), suppose that an equilibrium is monotone. Then $\sigma_b = \sigma_g$ as just proved. Now, an agent of type k seeks to maximize

$$1 - F\left(\frac{x^*-\theta_k}{\sigma_k}\right) - c\left(\sigma_k\right),$$

so that the corresponding first-order condition is given by

$$f\left(\frac{x^*-\theta_k}{\sigma_k}\right)\frac{x^*-\theta_k}{\sigma_k^2} - c'\left(\sigma_k\right) = 0. \tag{43}$$

Because $\sigma_g = \sigma_b = \sigma$ and c' is injective, the two first-order conditions together imply that

$$f\left(\frac{x^*-\theta_g}{\sigma}\right)(x^*-\theta_g) = f\left(\frac{x^*-\theta_b}{\sigma}\right)(x^*-\theta_b). \tag{44}$$

Furthermore, (42) tells us that

$$\beta f\left(\frac{x^*-\theta_g}{\sigma}\right) = f\left(\frac{x^*-\theta_b}{\sigma}\right),$$

and combining this with (44), we must conclude that

$$x^* = \frac{\theta_g - \beta\theta_b}{1-\beta}. \tag{45}$$

The value of x^* is thus completely pinned down by the system parameters. At the same time, going back to equation (44), we have

$$\frac{f\left(\frac{x^*-\theta_g}{\sigma}\right)}{f\left(\frac{x^*-\theta_b}{\sigma}\right)} = \frac{x^*-\theta_b}{x^*-\theta_g}. \tag{46}$$

By the strong MLRP there exists a unique value of σ that satisfies (46). It follows that (45) and (46) fully determine the values of σ and x^* *without paying any attention to the first-order condition* (43), which must also be satisfied. But that imposes an independent condition on the cost function $c(\sigma)$. ■

Proof of Propositions 7–12. See the Online Appendix.

REFERENCES

Alesina, A. and A. Cukierman (1990), "The Politics of Ambiguity," *The Quarterly Journal of Economics* **105**, 829–850.

Aliprantis, C. and K. Border (2006), *Infinite Dimensional Analysis: A Hitchhiker's Guide*, Springer, Third Edition.

Aragones, E. and Z. Neeman (2000), "Strategic Ambiguity in Electoral Competition," *Journal of Theoretical Politics* **12**, 183-204.

Aragones, E., Palfrey, T. and A. Postlewaite (2007), "Political Reputations and Campaign Promises," *Journal of the European Economic Assosiation* **5**, 846-884.

Aragones, E. and A. Postlewaite (2002), "Ambiguity in Election Games," *Review of Economic Design* **7**, 233-255.

Barron, D., Giorgiadis, G., and J. Swinkels (2017), "Optimal Contracts with a Risk-Taking Agent," mimeo.

Bhaskar, D. (2017), "The Value of Monitoring in Dynamic Screening," mimeo., Department of Economics, New York University.

Blume, A. and O. Board (2014), "Intentional Vagueness," *Erkenntnis* **79**, 855–899.

Blume, A., Board, O. and K. Kawamura (2007), "Noisy Talk," *Theoretical Economics* **2**, 395-440.

Campbell, J. (1983), "Ambiguity in the Issue Positions of Presidential Candidates: A Causal Analysis," *American Journal of Political Science* **27**, 284-293.

Crawford, V. and J. Sobel (1982), “Strategic Information Transmission,” *Econometrica* **50**, 1431–1451.

Dasgupta, A. and A. Prat (2006), “Financial Equilibrium with Career Concerns,” *Theoretical Economics* **1**, 67–93.

Degan, A. and M. Li (2016), “Persuasion with Costly Precision,” mimeo., Department of Economics, Concordia University.

Dewan, T. and D. Myatt (2008), “The Qualities of Leadership: Direction, Communication and Obfuscation,” *The American Political Science Review* **102**, 352-368.

Edmond, C. (2013), ”Information Manipulation, Coordination, and Regime Change,” *Review of Economic Studies* **80**, 1422-1458.

Fruttero, A. and V. Gauri (2005), “The Strategic Choices of NGOs: Location Decisions in Rural Bangladesh,” *Journal of Development Studies* **41**, 759–787.

Glazer, A. (1990), “The Strategy of Candidate Ambiguity,” *The American Political Science Review* **84**, 237-241.

Glazer, J. and A. Rubinstein (2004), “On Optimal Rules of Persuasion,” *Econometrica* **72**, 1715–1736.

Glazer, J. and A. Rubinstein (2006), “A Study in the Pragmatics of Persuasion: A Game Theoretical Approach,” *Theoretical Economics* **1**, 395–410.

Harbaugh, R., Maxwell, J. and K. Shue (2016), “Consistent Good News and Inconsistent Bad News,” mimeo., Indiana University.

Hart, S., Kremer, I. and M. Perry (2016), “Evidence Games: Truth and Commitment,” mimeo.

Hvide, H. (2002), “Tournament Rewards and Risk Taking,” *Journal of Labor Economics* **20**, 877-898.

Kamenica, E. and M. Gentzkow (2011), “Bayesian Persuasion,” *American Economic Review* **101**, 2590-2615.

Kostadinov, R. and A. Kuvalekar (2018), “Learning in Relational Contracts,” mimeo., Department of Economics, New York University.

Kuvalekar, A. and E. Lipnowski (2018), “Job Insecurity,” mimeo., Department of Economics, University of Chicago.

Makarov, I. and G. Plantin (2015), “Rewarding Trading Skills without Inducing Gambling,” *Journal of Finance* **70**, 925–962.

Matthews, A. and L. Mirman (1983), “Equilibrium Limit Pricing: The Effects of Private Information and Stochastic Demand,” *Econometrica* **51**, 981–996.

McClellan, A. and D. Ray (2018), “Contracts for Financial Managers Who Can Take on Risk,” mimeo., Department of Economics, New York University.

Palomino, F. and Prat, A. (2003), “Risk Taking and Optimal Contracts for Money Managers,” *RAND Journal of Economics* **34**, 113–137.

Penno, M. (1996), “Unobservable Precision Choices in Financial Reporting,” *Journal of Accounting Research* **34**, 141–149.

Ray, D. ⓡ A. Robson (2018), “Certified Random: A New Order for Coauthorship,” *American Economic Review*, **108**, 489–520.

Shepsle, K. (1972), “The Strategy of Ambiguity,” *American Journal of Political Science* **66**, 555–568.

Stein, J. (1989), “Cheap Talk and the Fed: A Theory of Imprecise Policy Announcements,” *The American Economic Review* **79**, 32–42.

Subramanyam, K. (1996), “Uncertain Precision and Price Reactions to Information,” *The Accounting Review* **71**, 207–219.

Titman, S. and B. Trueman (1986), “Information Quality and the Valuation of New Issues,” *Journal of Accounting and Economics* **8**, 159–172.

Zwiebel, J. (1995), “Corporate Conservatism and Relative Compensation,” *Journal of Political Economy* **103**, 1–25.