Costly Transparency

Justin Fox[*] and Richard Van Weelden[†]

June 13, 2011

Abstract

We consider whether a career-minded expert would make better decisions if the principal could observe the consequences of the expert's action. The previous literature has found that this "transparency of consequence" can only improve the efficacy of the expert's decision making. We show, however, that this conclusion is very sensitive to the specified cost structure: if learning the consequences of the expert's action makes the expert more likely to choose the action most likely to correspond to the true state of the world, when costs are asymmetric, this can be associated with a decrease in the principal's expected welfare. In addition, we show that, when the prior on the state of the world is sufficiently strong, if the principal benefits from learning the consequences of the expert's action, her utility is higher if she observes only the consequences and not the action taken. As such, the optimal transparency regime will involve either the principal observing only the expert's action or only the consequences of the expert's action; it will never be optimal to observe both. We illustrate these results with examples from finance and public policymaking.

Key Words: transparency; reputation; herding; principal-agent; welfare

[*] Assistant Professor of Political Science, Yale University. 77 Prospect St., New Haven, CT 06520. Phone: (203) 606-2608 Email: justin.fox@yale.edu

[†] Corresponding Author: Max Weber Fellow, European University Institute and Assistant Professor of Economics, University of Chicago. Via delle Fontanelle 10, I-50014 San Domenico, Italy. Phone: +39-346-8354-093 Email: richard.vanweelden@gmail.com

1 Introduction

In many situations, individuals delegate authority to an informed expert: investors give money to a fund manager in the hopes that it will be be invested wisely; voters elect politicians to make decisions for them; judges are charged with interpreting the law and upholding the constitution on behalf of the people. Of course, whenever the principal delegates to an expert, there is always the concern that the expert may not act in the principal's interest. It is usually argued that such concerns are lessened, however, if the principal can observe the consequences of the expert's action: learning the consequences of the action increases the incentive for the expert to take actions which are likely to benefit the principal.[1] While observing the consequences of the expert's action is certainly beneficial in many situations, we identify conditions under which it can decrease the principal's welfare.

We analyze a model in which the principal's preferred course of action varies with the underlying state of the world. There is a strong prior about this state, but some mistakes may be more costly than others. We show that, when the expert's primary concern is to promote his own career, and it is very costly if the expert takes the incorrect action in the less likely state of the world, principal welfare is decreased by observing the consequences of the expert's action. There are many such environments in which the ex-ante less likely event is associated with a greater potential cost: stock market crashes are relatively rare events but can lead to large losses or even bankruptcy; the defendant in a trial is more likely to be guilty but the costs of convicting the innocent are greater than the costs of acquitting the guilty; a Senator who believed that Iraq more likely than not had weapons of mass destruction might also believe the costs of invading if they didn't have weapons were greater than the costs of not invading if they did. Our results then show that observing the consequences of the expert's action can be socially harmful in many economically relevant environments.

Consider an expert, who may either be the high type, in which case he observes a perfectly accurate signal of the state, or the low type, in which case he observes a noisy but informative signal, who makes a decision on behalf of a principal. Further, assume that rather than seeking to promote the principal's welfare, the expert's objective is to maximize the principal's ex-post belief that the expert is the high type. Now suppose there is a strong prior about the state of world, and consider the decision of a low type expert who observes a signal contrary to that prior. If only the high type expert were to ever go against the prior, and the state of the world is not learned, the low type expert could mimic the high type expert by choosing the ex-ante less-likely alternative. So, in equilibrium, some low type experts will follow their signal. Now, suppose the state of the world will be learned with certainty before the expert's reputation is assigned. The low type expert can

[1] See Fearon (1999, 83): "Almost surely, politicians are most inclined to choose policies and other actions that the pubic desires when the probability of exposure for failure to do so is highest. The standard liberal observations about the importance of effective media and an informed, interested public follows immediately."

no longer mimic the high type by going against the prior – if he does, and the state matches the prior, the expert will reveal himself to be the low type. Hence, if the prior is high enough, the low type expert will ignore his private information and take the action favored by the prior – even if it's in the principal's interest for the expert to follow his signal.

So, when the prior on the state of the world is sufficiently strong, transparency about the consequences of the expert's action increases the probability a low type expert takes the action recommended by the prior rather than his signal, which increases the probability the expert's action matches the state. Whether this is socially harmful or beneficial depends on the costs associated with different mistakes. When costs are symmetric – that is, the costs of different mistakes are the same – transparency of consquence can only increase principal welfare. When costs are asymmetric, however, it may be socially optimal for the expert to go against the prior even if, conditional on his private information, the state of the world more likely than not matches the prior. The principal's welfare will then be decreased if she observes the consequences of the expert's action.

One extremely natural application of our model is to the decision faced by a fund-manager. First, the manager's reputation for competence is supremely important, so it makes sense to model the manager's objective as maximizing his reputation for competence. Second, a market crash is a relatively rare event but caries with it extremely large costs for investors. Third, there is extremely fast feedback about the consequences of the manager's decisions. Finally, there is much concern in the financial literature about herd behavior (e.g. Scharfstein and Stein 1990). Our results then raise the possibility that there is so much herding precisely *because* there is so much information available about the quality of the manager's choice: a manager who observes a signal that a crash is possible, but not likely, would be reluctant to act on such information for fear of being proven wrong.[2] So the career concerns of the manager lead him to expose investors to excessive risk of catastrophic loss – a distortion which is heightened when the manager's performance is more easily evaluated.

Another situation in which our model could be applied is to a legislator who must decide whether to approve an executive's proposal. Consider, for example, a U.S. Senator asked to vote on whether to authorize the Bush administration to use force against Iraq. Suppose the Senator (or more accurately the Senator's constituents) felt that war was justified if weapons of mass destruction existed, but not justified if they didn't. As, in the lead up to the war, the prior that weapons of mass destruction existed was high, our results indicate that if the Senator had weak information that weapons of mass destruction did not exist, he would have been more likely to oppose the war

[2] Robert Rodriquez, CEO of First Pacific Advisors, a hedge fund which divested its portfolio of subprime mortgages well before the financial crises, argues that, when it comes to risky but widely held assets, managers feel compelled to "be fully invested for fear of underperforming" (Rodriquez 2009).

if it would not have been learned whether they existed.[3] Further, if the costs of an unjustified war were sufficiently high, this opposition would have been in the voters' interest.

As the above examples indicate, in environments with a sufficiently unbalanced prior, learning the consequences of an expert's action only increases the incentive to herd on the prior. As such, efforts to improve transparency, and thereby increase the speed with which the consequences of the expert's action are revealed, though frequently beneficial, can, in some cases, make the principal worse off. Consequently, increasing the frequency of disclosure for financial institutions may have the counter-productive effect of increasing the manager's incentive to herd. Similarly, increasing the effectiveness of the media, which makes it more likely that the consequences of a politician's actions will be learned before the next election, can make politicians more reluctant to go against the conventional wisdom, even when doing so would benefit the public.

While for most agency problems, the more information the principal has, the better the agent's behavior, the literature provides some examples where more information can be harmful.[4] One paper which is especially relevant for our work, Prat (2005), argues that observing the action a career-minded expert takes may lead to socially harmful distortions. That said, in the established results in the literature (e.g. Canes-Wrone et al. 2001, Maskin and Tirole 2004, Prat 2005), it is never harmful to observe the consequences of the expert's action. In particular, Prat (2005, 863) describes "the main contribution of [Prat's] paper is to show that, while transparency on consequences is beneficial, transparency on action can have detrimental effects." When the prior on the state of the world is sufficiently strong, and the costs are sufficiently asymmetric, however, we show this is reversed: the principal is made better off observing the action ("transparency of action") but worse off observing the consequences of that action ("transparency of consequence").

To understand how principal welfare varies with the transparency regime, and how the standard welfare results can be reversed in an asymmetric cost environment, consider the expert's incentives when the prior is sufficiently unbalanced that, even if a low type expert observes a contrary signal, the state is more likely than not to match the prior. As previously noted, when the principal observes the expert's action, it is more difficult for a low type expert to mimic the high type by going against the prior when the principal also observes the consequences of that action. The fear of being proven wrong then makes the low type expert more reticent about acting on a signal contrary to the prior when both the consequences and the action are observed than when only the action is. However, when some low types follow the prior rather than their signal, if the principal sees the expert go against the prior, this reveals positive information about the expert's type. So, when the consequences are observed, also observing the action taken causes the expert to go against the prior

[3] This assumes that the objective of the Senator was to signal competence rather than to signal ideology or toughness. It also assumes that the Senator's vote would not affect the probability of learning the state.

[4] See Prat (2005) for a discussion.

more often. Because transparency of consequence and transparency of action push the expert's behavior in opposite directions, they have an opposite effect on principal welfare.

This means that we can order the likelihood of the low type following his signal across the three transparency regimes: when the principal observes only the action taken but not the consequences, when the principal observes only the consequences but not the action, and when both the action and the consequences are observed. When the prior is sufficiently unbalanced, and the expert knows his type, the expert is most likely to herd on the prior when only the consequences are observed and least likely to when only the action is. Which transparency regime results in the highest welfare depends on the cost structure. When costs are symmetric, so the first-best decision rule involves the expert taking the action most likely to match the state, it is optimal to have transparency over consequence, but not action. However, when the costs are sufficiently asymmetric, so the first-best decision rule involves the expert matching his action to his signal, even if it goes against the prior, it is optimal to have transparency over action, but not consequences. As transparency of consequence is almost always harmful when transparency of action is beneficial, generically, some form of non-transparency increases welfare.

The paper is organized as follows: Section 2 describes the model, Section 3 presents our results, and Section 4 concludes. The proofs of our results are left to the Appendix.

2 Model

A privately informed expert makes a decision on behalf of a principal.[5] We assume that there are two states of the world, $\omega \in \{0, 1\}$, with prior probability that the state is 1, $Pr(\omega = 1) = \pi > 1/2$. There are two possible actions which can be taken, $a \in \{0, 1\}$. The utility to the principal depends on the action and the state,

$$u(a, \omega) = \begin{cases} 0 \text{ if } a = \omega, \\ -1 \text{ if } a = 0, \omega = 1, \\ -\kappa \text{ if } a = 1, \omega = 0. \end{cases} \tag{1}$$

Normalizing the payoff when the action matches the state to be 0 in either state is without loss of generality, as what matters is the difference between the payoff when the action matches the state and when it does not. Notice that we are not assuming that the costs are symmetric: if κ is large, taking action 1 in state 0 is more costly than taking action 0 in state 1. We focus on the case in which $\kappa \geq 1$, so while state $\omega = 0$ is less likely to occur, the cost from not matching the action to the state may be larger in state 0.

The expert observes a private signal, $s \in \{0, 1\}$, of the state of the world. Experts are heterogenous with regard to their ability, t, and can be either the high type, $t = h$, or the low type, $t = l$. We assume the expert is the high type with probability $\alpha \in (0, 1)$. The expert knows his own

[5] We use male pronouns for the expert and female pronouns for the principal.

type but the principal knows only the distribution of expert types.[6] A high type expert observes a signal that is accurate with probability q_h, whereas a low type expert observes a signal that is accurate with probability q_l. That is, $Pr(s = 1|\omega = 1, t = h) = Pr(s = 0|\omega = 0, t = h) = q_h$ and $Pr(s = 1|\omega = 1, t = l) = Pr(s = 0|\omega = 0, t = l) = q_l$. We assume $1/2 < q_l < q_h \leq 1$.

The expert is the only active player in our model. His strategy maps his type and his signal of the state into a probability of taking each action. Formally, the expert's strategy is represented by

$$\sigma(t, s) \in [0, 1],$$

the probability with which the expert takes action 0 for each type, $t \in \{l, h\}$, and signal of the state, $s \in \{0, 1\}$. The principal takes no action but updates her belief about the expert's type based on the information she observes. Finally, we assume that the expert's concern is to appear competent, so he seeks to maximize the probability that the principal places on him being the high type. The key assumptions underlying this payoff structure are that long term contracting between the principal and the expert is not possible[7] and that an expert who is the high type in the current period is likely to be more competent than a low type expert at making decisions in the future. We compare the efficiency – defined as the principal's expected utility in the current period – of different information structures.[8] Following Prat (2005), we describe learning whether or not the expert's action matches the state of the world before assigning his reputation as "transparency on consequence" and observing the action taken by the expert as "transparency on action".

Under the first information structure, the principal observes both the action taken by the expert and the state of the world before assigning the expert's reputation. We refer to this as *Full Transparency* (FT). The expert's reputation is given by

$$\lambda(a, \omega) \equiv Pr(t = h|a, \omega), (a, \omega) \in \{0, 1\}^2.$$

We next consider the information structure in which the principal does not observe the consequences of the expert's action. That is, the principal does not observe whether the action taken by

[6] It is important for our results that the expert has private information about his own type. That the expert reveals positive information about his own type by going against the prior (e.g. Levy 2004) will play an important role in our analysis. It is not necessary, however, for the expert to know his own type with certaintly. See Fox and Van Weelden (2010) for a model in which an expert receives a noisy, but informative, signal of his own type. The expert's incentives are fundamentally unchanged when this signal is sufficiently informative.

[7] Clearly, if the principal were able write contracts with the expert which conditioned his payoff on the observed action or consequences, greater transparency can never be harmful to the principal.

[8] This means that we are ignoring the possible benefits to the principal from learning about the expert's type. One setting in which the principal would derive no benefit from learning about the expert's type is a perfectly competitive market without the possibility of long term contracting. As all experts would then be paid according to their expected value, all the gains or losses from learning about the expert's competence would accrue to the expert. Assuming a competitive labor market without the possibility of long term contracting is standard in the career concerns literature pioneered by Holmstrom (1999).

the expert matches the state of the world. The expert's reputation in this setting is

$$\lambda(a) \equiv Pr(t = h|a), a \in \{0, 1\}.$$

We refer to this as *Non-Transparent Consequences* (NC). With *Non-Transparent Consequences*, the expert receives his repuational payoff before the state of the world is learned and the principal receives her utility. For example, a manager may be "judged" by the market (e.g. Brandenburger and Polak 1996), or a politician may come up for re-election (e.g. Canes-Wrone et al. 2001, Maskin and Tirole 2004), before the consequences of their actions are known.

Finally, we consider the case in which the principal does not observe the expert's action but does observe whether the action matches the state of the world. The expert's reputation is then

$$\bar{\lambda}(i) \equiv Pr(t = h|\mathbb{I}(a, \omega) = i), i \in \{0, 1\}.$$

We refer to this as *Non-Transparent Action* (NA).[9]

Under all three information structures, in equilibrium, the principal forms her beliefs via Bayesian updating, and the expert's strategy maximizes his expected reputation given these beliefs. Note that, as the expert's action doesn't affect the probability of learning the state, differences in equilibrium behavior across information structures will not be driven by asymmetric observability.[10]

Before proceeding to the results, we discuss the behavior of the expert that would maximize the principal's welfare. In order to make the problem interesting, we focus on parameters for which delegating to an expert is potentially beneficial. We first assume that $\kappa \leq \frac{\pi}{1-\pi}$. So, if the principal did not have access to an expert, she would go with her prior and take action 1. This also means that the principal's welfare is higher when any expert, regardless of his type, chooses action $a = 1$ upon observing $s = 1$. Second, we assume that $q_h > \pi$, which ensures that it is in the principal's interest for the high type expert to follow his signal even if it goes against the prior.

We now consider the socially optimal action by a low type expert who observed signal $s = 0$. Note that

$$Pr(\omega = 1|t = l, s = 0) = \frac{\pi(1 - q_l)}{\pi(1 - q_l) + q_l(1 - \pi)},$$

which is greater than 1/2 when $\pi > q_l$. As the principal's expected utility is $-\kappa Pr(\omega = 0|t = l, s = 0)$ if action $a = 1$ is taken, and $-Pr(\omega = 1|t = l, s = 0)$ if $a = 0$ is, we have the following remark.

[9] As costs are asymmetric, we assume that the principal observes whether the action matches the state but not her utility; consequently, the principal cannot infer the action taken from the utility she receives. If the principal were to observe her utility, when costs are not symmetric, the principal must be able to infer the action taken either when the action matches the state or when the action and the state do not match. Note also that the information possessed by the principal, and the principal's updating, is exactly the same as when Prat (2003, 2005) considers non-transparent action.

[10] That the principal may be more likely to observe the state after certain actions are taken has been explored in the previous literature. See Canes-Wrone et al. (2001), Milbourn et al. (2001), and Suurmond et el. (2004) for models in which increasing the probability that the state of the world is revealed after one action is taken, holding the probability the state is revealed after the other action is, can create, or increase, distortions in the agent's behavior.

Remark 1 *If $\kappa > \frac{\pi(1-q_l)}{q_l(1-\pi)}$, it is socially optimal for low type experts to select $a = 0$ after observing $s = 0$; if $\kappa < \frac{\pi(1-q_l)}{q_l(1-\pi)}$, it is socially optimal for low type experts to select $a = 1$ after observing $s = 0$.*

Notice that, when $\pi > q_l$, though a low type expert who observes $s = 0$ believes state $\omega = 1$ is more likely, when $\kappa > 1$ it may nevertheless be socially optimal for him to take action $a = 0$. As our main contribution – showing transparency over consequence can be harmful – will require $\pi > q_l$, in our analysis we assume $\pi \in (q_l, q_h)$. The case in which $\pi < q_l$ has been more heavily studied in the literature (e.g. Canes-Wrone et al. 2001), and, for such parameters, only when $\kappa < 1$ can transparency over consequence be harmful.

3 Results

3.1 Preliminaries

We now turn to the analysis of equilibrium behavior. As the expert's payoff does not depend on the action taken, there will be many equilibria of this game. As is standard in the literature (e.g. Levy 2004, Prat 2005), we restrict attention to informative equilibria and ignore "mirror" or "perverse" equilibria in which the high type expert chooses the action opposite to his signal. We say that an equilibrium is *non-pooling* if both actions, $a = 1$ and $a = 0$, are taken by the expert with positive probability along the equilibrium path. Further, to rule out babbling and perverse equilibria, we focus on equilibria in which the expert's strategy is *monotone*.

Definition 1 *A strategy, σ, is monotone if, for any type and signal combination $(t, s) \in \{0, 1\}^2$,*

1. *if $\sigma(t, s) < 1$, then, for all (t', s') such that $Pr[\omega = 0|(t', s')] < Pr[\omega = 0|(t, s)]$, $\sigma(t', s') = 0$.*
2. *if $\sigma(t, s) > 0$, then, for all (t', s') such that $Pr[\omega = 0|(t', s')] > Pr[\omega = 0|(t, s)]$, $\sigma(t', s') = 1$.*

We refer to an equilibrium as monotone if the expert's strategy is monotone. Monotone equilibria are those in which experts are more likely to take action $a = 1$ ($a = 0$) the more likely they believe the state is 1 (respectively 0).

As many non-monotone equilibria depend critically on the expert's indifference over the action, one way to motivate our focus on monotone equilibria is to perturb the expert's payoff so that, in addition to seeking to maximize his reputation, he attaches a small positive weight to the principal's welfare. Then, provided the expert does not suffer a reputational penalty when his action matches the state, his incentive to take a given action is increasing in the probability he assigns to it matching the state. As such, with some weak additional restrictions on the principal's beliefs at certain information sets, any Perfect Bayesian Equilibrium must be non-pooling and monotone.[11] Under all three information structures, there will be a unique equilibrium of this form.

[11] Under *Non-Transparent Consequences*, when the expert places positive weight on the principal's welfare, in any non-pooling equilibrium, the expert's strategy must be monotone. Further, when the expert places a positive weight

Proposition 1 *Suppose $q_l < \pi < q_h$. Under all three information structures, $j \in \{FT, NA, NC\}$, there exists a non-pooling, monotone Perfect Bayesian Equilibrium. This equilibrium involves the high type expert taking action $a = s$, and the low type expert taking action $a = 1$ if he observes $s = 1$, and action $a = 0$ with probability $\sigma_0^j \in [0, 1]$ if he observes $s = 0$. Further, this equilibrium is unique, up to the beliefs at off-path information sets.*

By Proposition 1, the expert's strategy in the non-pooling, monotone equilibrium in each environment can be characterized by the number σ_0^j, the probability with which a low type expert takes action $a = 0$ after observing signal $s = 0$ under information structure $j \in \{FT, NA, NC\}$. We now compare the non-pooling, monotone equilibria across the different information structures.

3.2 Results ($q_h = 1$)

We first consider the case in which the high type's signal is perfectly accurate. That is, we assume that $q_h = 1$ and $q_l = q$. This simplifies the algebra as the principal would know with certainty that any "mistake" must have been made by a low type expert. The main result of this section is that the low type expert is most willing to act on his private information when information about the consequences of his action is suppressed. We further show that the low type expert is least likely to act on his private information when the principal cannot observe the action taken.

Proposition 2 *Define $\pi^* \equiv \frac{q}{q+\alpha(1-q)} \in (q, 1)$. Then,*

1. *for $\pi \in [\pi^*, 1)$,*
$$0 = \sigma_0^{NA} = \sigma_0^{FT} < \sigma_0^{NC}.$$
2. *there exists $\pi_* \in (q, \pi^*)$ such that for $\pi \in (\pi_*, \pi^*)$,*
$$0 = \sigma_0^{NA} < \sigma_0^{FT} < \sigma_0^{NC}.$$

To provide an intuition for this result, we first discuss the relationship between *Non-Transparent Consequences* (NC) and *Full Transparency* (FT) when $\pi \geq \pi^*$. Under (NC), if only the high type expert were to ever select $a = 0$, taking action 0 would reveal the expert to be the high type with certainty, giving the low type an incentive to take action $a = 0$ as well. Hence, in equilibrium, the low type expert must take action $a = 0$ with positive probability after observing $s = 0$. Under

on principal welfare, the off-path beliefs in the pooling equilibria would not satisfy criterion D1 of Cho and Kreps (1987). See Fox and Van Weelden (2010) for the details of this approach in a closely related environment. When the consequences are observed, even with a positive weight on the principal's welfare, there can exist perverse equilibria in which the expert takes the action which he believes less likely to match the state and the expert's reputation is higher when his action does not match the state. However, when the expert places positive weight on principal welfare, under the range of parameters we consider, non-pooling, monotone Perfect Bayesian Equilibria are the only equilibria to satisfy D1, in which, for each action, the expert's reputation is no lower if the action matches the state ($\lambda(1, 1) \geq \lambda(1, 0)$, $\lambda(0, 0) \geq \lambda(0, 1)$, and $\bar{\lambda}(1) \geq \bar{\lambda}(0)$).

(FT), as the principal observes the consequences of the expert's action, it is no longer possible for the low type expert to mimic the high type expert by going against the prior: if the low type expert received an incorrect signal this will be revealed. Hence, if, conditional on the low type expert's signal, the probability that the state is 1 is sufficiently high, he would decide to take action $a = 1$ rather than increase the probability of revealing himself to have observed an incorrect signal by taking action $a = 0$.

As a high type expert never observes an incorrect signal, the expert's expected reputation, from either action, is equal to his reputation if proven correct multiplied by the probability his action matches the state. So, there exists a Perfect Bayesian Equilibrium in which all low type experts take action $a = 1$ if, and only if, $Pr(\omega = 1|t = l, s = 0)\lambda(1, 1) \geq Pr(\omega = 0|t = l, s = 0)\lambda(0, 0)$. As the principal would infer that the expert is the high type with certainty if he is proven correct after taking action 0, but not after action 1 ($\lambda(0, 0) = 1 > \lambda(1, 1)$), to support these strategies as an equilibrium, $Pr(\omega = 1|t = l, s = 0)$ must be sufficiently greater than 1/2. So we need

$$\pi \geq \frac{q}{q + \alpha(1 - q)} \equiv \pi^*,$$

which is a stronger condition than $\pi > q$. We can then conclude that $0 = \sigma_0^{FT} < \sigma_0^{NC}$ when $\pi \geq \pi^*$.

Note also that, while the low type expert must choose $a = 0$ with positive probability under *Full Transparency* when $\pi < \pi^*$, by continuity, for a non-degenerate interval (π_*, π^*), he will do so with greater probability under *Non-Transparent Consequences*. As the ranking of the probabilities of the low type expert taking action $a = 0$ in these two information structures is reversed when $\pi \approx q$, we must have that $\pi_* \in (q, \pi^*)$.[12]

Combing the above with Remark 1, when $\pi > \pi_*$ and $\kappa > \frac{\pi(1-q)}{q(1-\pi)}$, as the first-best decision rule involves low type experts taking action $a = 0$ when $s = 0$, but a low type who observes $s = 0$ is more likely to take action 0 with *Non-Transparent Consequences* than *Full Transparency*, transparency of consequence decreases principal welfare. So, learning the consequences of the expert's action is harmful if two things are true: the prior the state is 1 is high enough that a low type expert believes state 1 is sufficiently more likely than 0 even if he observes a signal of 0, and the costs are asymmetric enough that it is in the principal's interest for him to follow his signal anyways.

We now discuss expert behavior under *Non-Transparent Action* (NA) and how it compares to *Full Transparency.* Observing the consequences of the expert's action, but not the action itself, is the form of non-transparency considered in Prat (2005).[13] Under *Non-Transparent Action* the

[12] So, for some $\pi \in (q, \pi_*)$, a low type expert is more likely take action $a = 0$ under (FT) than (NC), even though, regardless of his signal, state $\omega = 1$ is more likely. For these parameters, when costs are symmetric, welfare is higher with (NC) than (FT). As such, the standard result that transparency of consequences can only be beneficial when costs are symmetric (e.g. Prat (2003, 2005), Canes-Wrone et al. 2001) depends critically on the assumption that $q_l \geq \pi$.

[13] In Prat (2005), the expert has no private information about his own ability, though his results continue to hold if the expert does. The unpublished working paper version, Prat (2003), considers experts who have private information

expert's action cannot signal anything about his competence (since the principal doesn't observe it). Consequently, in the non-pooling, monotone equilibrium, as the expert's reputation is enhanced if his action matches the state of the world, he always takes the action most likely to match the state given his private information. When $\pi > q$, this means that the low type takes action $a = 1$ regardless of his signal. So, while the behavior under (FT) and (NA) is identical for $\pi \geq \pi^*$, when $\pi \in (q, \pi^*)$, the low type expert takes action $a = 0$ with a greater probability under *Full Transparency*. Hence, when $\pi \in (q, \pi^*)$ and κ is large, the principal's welfare is higher under *Full Transparency* than *Non-Transparent Action*. So, as in Prat (2005), we find that non-transparency of action increases the incentive for the expert to take the action most likely to match the state of the world. However, when costs are asymmetric, this can decrease welfare.

3.3 Results ($q_h < 1$)

We now consider the equilibrium behavior when the high type's signal is not perfectly accurate, and focus on the case in which the prior is sufficiently unbalanced ($\pi > \pi_*$). While assuming that the high type's signal is perfectly accurate simplifies the algebra, it also means that the principal will believe that the expert is the low type with certainty if she ever learns that the expert's signal did not match the state. Hence, for a wide range of parameters, if the principal observes the consequences of the expert's action, the low type expert will always stick with the prior out of fear of being proven wrong. In contrast, if the high type expert's signal is not perfect, the expert who received an incorrect signal may still be the high type; consequently, if only high type experts were to take a certain action, the principal would infer that they were the high type regardless of the consequences of the action. So, in the information structures in which the action is observed, the low type expert must also be willing to go against the prior with some probability in equilibrium. Under *Non-Transparent Action*, conversely, the expert is still incentived to take the action most likely to match the state, and so all low type experts will follow the prior. Our next result shows that, if the high type expert's signal is sufficiently accurate, but not perfectly accurate, we can rank the probability of a low type expert going against the prior under the three information structures.

Proposition 3 *For all $\pi > \pi_*$, there exists a $\bar{q}_h(q_l, \pi) < 1$ such that, for all $q_h \in (\bar{q}_h(q_l, \pi), 1)$,*

$$0 = \sigma_0^{NA} < \sigma_0^{FT} < \sigma_0^{NC}.$$

Therefore, the low type expert is most likely to act on his signal of the state when the consequences of his action are unobserved and least likely when only the consequences are observed. If $\kappa \neq \frac{\pi(1-q_l)}{q_l(1-\pi)}$, the principal will not be indifferent over the low type expert's decision after observing $s = 0$. As such an expert randomizes with a non-degenerate probability under *Full Transparency*, some form of non-transparency will lead to a strict increase in the principal's welfare.

about their own competence.

Corollary 1 *Suppose $\pi > \pi_*$ and $q_h \in (\bar{q}_h(q_l, \pi), 1)$. Then,*

1. *if $\kappa > \frac{\pi(1-q_l)}{q_l(1-\pi)}$, the principal's welfare is highest with* Non-Transparent Consequences*.*
2. *if $\kappa < \frac{\pi(1-q_l)}{q_l(1-\pi)}$, the principal's welfare is highest with* Non-Transparent Action*.*
3. Full Transparency *is only welfare maximizing if $\kappa = \frac{\pi(1-q_l)}{q_l(1-\pi)}$, in which case all three information structures generate the same payoff to the principal.*

So, generically, when the prior is sufficiently unbalanced and the high type expert's signal is sufficiently accurate, some form of non-transparency strictly increases the principal's welfare. Which type of non-transparency is beneficial depends on the cost structure, and, when one form of non-transparency is beneficial the other is harmful. Thus, *Full Transparency* is always the second most preferred information structure under the range of parameters we consider. So, if some relevant parameters (π, q_h, q_l, κ) are unknown, it may be the optimal information structure ex-ante. Further, if the principal benefits from learning about the expert's type – which she does not in this model – this could provide additional benefits to *Full Transparency.* Finally, if the principal could contract with the expert based on her observations, greater transparency can only be beneficial.

3.4 Overconfident Experts

While we have assumed that all players are fully rational, with the expert knowing his own type, it has been well documented that individuals are often overconfident in their own ability.[14] In our setting, one possible form of overconfidence is for some fraction, μ, of the low type experts to believe themselves to be the high type. Interestingly, if the fraction of overconfident low type experts is not too large, our results don't change very much at all. If σ_0^j, the probability of randomization with fully rational experts under information structure $j \in \{FT, NA, NC\}$, is positive, introducing a positive fraction of overconfident low types, $\mu \leq \sigma_0^j$, won't affect the probability that each ability type takes each action. As an overconfident low type expert believes himself to be the high type, he will behave like a high type expert, and always take action $a = s$ in any non-pooling, monotone equilibrium. Then, in order for the rational low types to be indifferent over actions when $s = 0$, they must randomize with a probability such that the fraction of low types, including those who are overconfident, who take action $a = 0$ after observing $s = 0$ is σ_0^j – the same fraction as when all experts are fully rational. Only when $\mu > \sigma_0^j$ would the probability of each action change.

When $j \in \{FT, NA\}$, if $\mu > \sigma_0^j$, this increases the fraction of low type experts taking action $a = 0$: as overconfident low types behave like high types, in equilibrium, all high types and overconfident low types take action $a = s$, and all rational low types always take action $a = 1$. So, when κ is large,

[14] See, for example, Thaler (2000) for an overview of results on cognitive bias in economic decision making. See also Odeon (1998), among others, for a model in which traders are overconfident in the quality of their information.

expert overconfidence can increase welfare. When $j = NC$, however, if $\mu > \sigma_0^{NC}$, there no longer exists a non-pooling, monotone equilibrium. If the fraction of low types who believe themselves to be the high type is large, the reputational benefit the expert would receive from convincing the principal that he believes himself to be the high type would not be enough to offset the reputational penalty from revealing that he observed a signal contrary to the prior. Hence, all experts have a strict incentive to convince the principal that they observed signal $s = 1$, and it is not possible to support a non-pooling, monotone equilibrium with *Non-Transparent Consequences*.

Our results are then robust to deviations from full rationality, provided that the fraction of overconfident experts is not too large ($\mu \in [0, \sigma_0^{NC})$). When $\mu \in [0 = \sigma_0^{NA}, \sigma_0^{FT})$, the ranking of the randomization probabilities in Proposition 3, and hence the welfare comparisons in Corollary 1, continue to hold. Further, when $\mu \in [\sigma_0^{FT}, \sigma_0^{NC})$, while expert behavior is identical under *Full Transparency* and *Non-Transparent Action*, low type experts are still most likely to go against the prior with *Non-Transparent Consequences*. So, when $\mu < \sigma_0^{NC}$, *Non-Transparent Consequences* still provides the highest welfare when κ is large.

4 Conclusions

We have considered the behavior of a career-minded expert under three different information structures: when both the expert's action and the consequences of that action are transparent, when only the action is transparent, and when only the consequences are. Our analysis has focused on the case in which there is a strong prior on the state of world, but where the costs of different mistakes may be asymmetric. In contrast to the previous literature (e.g. Canes-Wrone et al. 2001, Maskin and Tirole 2004, Prat 2005), we have shown that observing the consequences of the expert's action can often decrease the principal's welfare.

In our model, if the prior is sufficiently unbalanced, non-transparency of consequences increases the incentive to take the action contrary to the prior. Therefore, when a career-minded expert is too reticent relative to the first-best in going against the prior under full transparency, non-transparecy of consequences will be beneficial. Non-transparency of action has the opposite effect, decreasing the incentive to take the action contrary to the prior. So, when a career-minded expert takes the contrarian action too often relative to the first-best under full transparency, non-transparency of action is beneficial. As equilibrium behavior with full transparency will generically fall short of the first-best, some form of non-transparency will be beneficial; which form is beneficial will depend on the specific cost structure.

As there are many situations in which costs are not symmetric, our results stand as an important caveat to the established results in the literature on the welfare implications of different forms of transparency. In the career concerns literature, the expert's objective, and hence equilibrium behavior, does not depend on the distribution of costs. The first-best decision rule, in contrast, will be closely tied to the cost structure. As such, statements about the welfare implications of different transparency regimes will be sensitive to the costs of different types of mistakes in the environment considered.

5 Acknowledgements

We thank Dirk Bergemann, Anne Guan, Ekaterina Malinova, Massimo Morelli, Gilles Serra, Craig Volden, two anonymous referees, and the editor for helpful comments. Van Weelden thanks Larry Samuelson and the Max Weber Programme for financial support when completing this work.

6 References

Brandenburger, Adam and Ben Polak (1996). “When Managers Cover their Posteriors: Making the Decisions the Market Wants to See.” *RAND Journal of Economics*, 24, 3, 523-541.

Canes-Wrone, Brandice, Michael Herron, and Kenneth Shotts (2001). “Leadership and Pandering: A Theory of Executive Policymaking.” *American Journal of Political Science*, 45, 3, 532-550.

Cho, In-Koo, and David M. Kreps (1987). “Signaling Games and Stable Equilibria.” *Quarterly Journal of Economics* 102, 2, 179-221.

Fearon, James (1999). “Electoral Accountability and the Control of Politicians: Selecting Good Candidates versus Sanctioning Poor Performance.” In Democracy, Accountability, and Representation, eds. A. Przeworski, S.C. Stokes and B. Manin. Cambridge University Press.

Fox, Justin and Richard Van Weelden (2010). “Partisanship and the Effectiveness of Oversight.” *Journal of Public Economics*, 94, 9-10, 684-697.

Holmstrom, Bengt (1999). “Managerial Incentive Problems: A Dynamic Perspective.” *Review of Economic Studies*, 66, 1, 169-182 (Originally appeared in Essays in Honor of Lars Wahlbeck in 1982).

Levy, Gilat (2004). “Anti-herding and Strategic Consultation.” *European Economic Review*, 48, 3, 503-525.

Maskin, Eric and Jean Tirole (2004). “The Politician and the Judge.” *American Economic Review*, 94, 4, 1034-1054.

Milbourn, Todd, Richard Shockley and Anjon Thakor (2001). “Managerial Career Concerns and Investments in Information.” *RAND Journal of Economics*, 32, 2, 334-351.

Odean, Terrence (1998). “Volume, Volatility, Price, and Profit when All Traders are Above Average.” *Journal of Finance*, 53, 6, 1887-1934.

Prat, Andrea (2003). “The Wrong Kind of Transparency.” Center for Economic Policy Research. CEPR Discussion Paper No. 3859.

Prat, Andrea (2005). “The Wrong Kind of Transparency.” *American Economic Review*, 95, 3, 862-877.

Rodriquez, Robert L. (2009). “Reflections and Outrage.” Speech to the Morningstar conference on May 29, 2009.

Scharfstein, David and Jeremy Stein (1990). “Herd Behavior and Investment.” *American Economic Review*, 80, 3, 465-479.

Suurmond, Guido, Otto Swank and Bauke Visser (2004). “On the Bad Reputation of Reputational Concerns.” *Journal of Public Economics*, 88, 11-12, 2817-2838.

Thaler, Richard (2000). “From Homo Economicus to Homo Sapiens.” *Journal of Economic Perspectives*, 14, 1, 133-141.

7 Appendix

In this section we provide the proofs of Propositions 1-3. We begin with the proof of Proposition 1, the existence and uniqueness of a non-pooling, monotone equilibrium. As this result encapsulates statements about three different information structures we prove each one as a separate Lemma.

Lemma 1 *Suppose $\pi > 1/2$ and $j = NC$. Then there exists a unique non-pooling, monotone Perfect Bayesian Equilibrium. In this equilibrium the high type expert always takes action $a = s$. The low type expert takes action $a = 1$ after observing $s = 1$ and randomizes with a non-degenerate probability after observing $s = 0$.*

Proof. We begin by noting that in any non-pooling, monotone Perfect Bayesian Equilibrium the high type expert must always choose $a = s$. This follows immediately from the following argument: in a non-pooling equilibrium both actions must be chosen with positive probability; by monotonicity, if any expert ever chooses a given action, $a' \in \{0, 1\}$, then the high type expert who observed signal $s = a'$ must always choose that action; if the high type were to randomize after observing $s = a'$, all other experts would have to choose action $1 - a'$ with probability 1; if only the high type expert were to choose action a', the principal would assign reputation $\lambda = 1$ after observing action a', so the expert has a strict incentive to take action a'.

Given that the high type expert will always take action $a = s$, we can then restrict our analysis to the low type expert's behavior given that the high type always follows his signal of the state. We can represent any expert strategy by the double (σ_1, σ_0), the probability of the low type expert choosing 0 after observing signal $s = 1$ and $s = 0$ respectively.

Our next task is to show that we must have $\sigma_1 = 0$. We do this by contradiction. Suppose there exists a non-pooling, monotone equilibrium with $\sigma_1 > 0$. Then, by monotonicity, in this equilibrium, $\sigma_0 = 1$. Now we can calculate the reputations of the expert after taking each action for each $\sigma_1 \in (0, 1]$. Note that by Bayes's rule,

$$\lambda(1|\sigma_1, \sigma_0 = 1) = \frac{Pr(h,1)}{Pr(h,1) + Pr(l,1)} = \frac{\alpha[\pi q_h + (1-\pi)(1-q_h)]}{\alpha[\pi q_h + (1-\pi)(1-q_h)] + (1-\alpha)(1-\sigma_1)[\pi q_l + (1-\pi)(1-q_l)]},$$

$$\lambda(0|\sigma_1, \sigma_0 = 1) = \frac{Pr(h,0)}{Pr(h,0) + Pr(l,0)} = \frac{\alpha[(1-\pi) q_h + \pi(1-q_h)]}{1 - \alpha[\pi q_h + (1-\pi)(1-q_h)] - (1-\alpha)(1-\sigma_1)[\pi q_l + (1-\pi)(1-q_l)]}.$$

Note that $\lambda(0|\sigma_1, \sigma_0 = 1)$ is decreasing in σ_1, $\lambda(1|\sigma_1, \sigma_0 = 1)$ is increasing in σ_1, and, when $\sigma_1 = 0$,

$$\lambda(0|\sigma_1 = 0, \sigma_0 = 1) = \frac{\alpha[(1-\pi) q_h + \pi(1-q_h)]}{[(1-\pi) q_h + \pi(1-q_h)] + (1-\alpha)(q_h - q_l)(2\pi - 1)} < \alpha,$$

$$\lambda(1|\sigma_1 = 0, \sigma_0 = 1) = \frac{\alpha[\pi q_h + (1-\pi)(1-q_h)]}{[\pi q_h + (1-\pi)(1-q_h)] - (1-\alpha)(q_h - q_l)(2\pi - 1)} > \alpha,$$

as $\pi > 1/2$. So we have

$$\lambda(0|\sigma_1, \sigma_0 = 1) < \lambda(1|\sigma_1, \sigma_0 = 1),$$

whenever $\sigma_1 > 0$. Hence, the expert's reputation would be strictly higher from taking action $a = 1$ than action $a = 0$, which means that the above strategy cannot be part of an equilibrium. Therefore, we must have $\sigma_1 = 0$ in any non-pooling, monotone PBE.

We now show that there exists a unique equilibrium with $\sigma_1 = 1$ and $\sigma_0 \in (0, 1)$ and all high type experts taking action $a = s$. First, define $\lambda(a|\sigma_0)$ to be the reputation of the expert after each action is chosen, for each $\sigma_0 \in [0, 1]$. Then, given π, q_h, and q_l, for any probability of randomization, σ_0, the reputational difference between actions $a = 0$ and $a = 1$ is

$$\phi(\sigma_0) \equiv \lambda(0|\sigma_0) - \lambda(1|\sigma_0). \tag{2}$$

Notice that we have an equilibrium iff $\phi(\sigma_0) = 0$, and it has been established that $\phi(\sigma_0 = 1) < 0$.

By Bayes's rule,

$$\lambda(0|\sigma_0) = \frac{Pr(h,0)}{Pr(h,0) + Pr(l,0)} = \frac{\alpha[(1-\pi) q_h + \pi(1-q_h)]}{\alpha[(1-\pi) q_h + \pi(1-q_h)] + (1-\alpha)[(1-\pi) q_l + \pi(1-q_l)]\sigma_0}.$$

Notice then that $\lambda(0|\sigma_0)$ is clearly decreasing in σ_0 with $\lambda(0|0) = 1 > \alpha$. Similarly,

$$\lambda(1|\sigma_0) = \frac{Pr(h,1)}{Pr(h,1) + Pr(l,1)} = \frac{Pr(h,1)}{1 - [Pr(h,0) + Pr(l,0)]}.$$

As the denominator is decreasing in σ_0, $\lambda(1|\sigma_0)$ is increasing in σ_0. In addition, as $\lambda(0|0) = 1 > \alpha$,

$$\lambda(1|0) = \frac{\alpha - Pr(a=0|\sigma_0=0)\lambda(0|0)}{1 - Pr(a=0|\sigma_0=0)} = \alpha - \frac{Pr(a=0|\sigma_0=0)(\lambda(0|0)-\alpha)}{1 - Pr(a=0|\sigma_0=0)} < \alpha.$$

Therefore, $\phi(\sigma_0) = \lambda(0|\sigma_0) - \lambda(1|\sigma_0)$ is a decreasing function of σ_0 with $\phi(0) > 0$ and $\phi(1) < 0$, and so there exists a unique solution, σ_0^{NC}, to $\phi(\sigma_0) = 0$. Further, $\sigma_0^{NC} \in (0,1)$.

Hence, we conclude there exists a unique non-pooling, monotone PBE. In this equilibrium, the high type follows his signal, and the low type takes action $a = 1$ after observing signal $s = 1$ and action $a = 0$ with probability $\sigma_0^{NC} \in (0,1)$ after observing signal $s = 0$. ■

Lemma 2 *Suppose $\pi \in (q_l, q_h)$ and $j = FT$. Then there exists a non-pooling, monotone Perfect Bayesian Equilibrium which is unique up to the beliefs at off-path information sets. In this equilibrium the high type expert always takes action $a = s$, and the low type expert takes action $a = 1$ when $s = 1$ and action $a = 0$ with probability $\sigma_0 \in [0,1]$ when $s = 0$.*

Proof. We begin by noting that in any non-pooling, monotone Perfect Bayesian Equilibrium the high type expert must always choose $a = s$. To see this, first note that, in a non-pooling equilibrium, both actions must be chosen with positive probability; by monotonicity, if any expert ever takes action $a' \in \{0,1\}$, the high type expert who observed signal $s = a'$ must always take action a'; if the high type were to randomize after signal $s = a'$, all other experts would have to take action $1 - a'$ with probability 1. Further, if only the high type expert ever takes action a', the principal must assign reputation $\lambda = 1$ after observing (a', ω) if $\omega = a'$, and, if $q_h < 1$, also if $\omega \neq a'$. Hence the expected reputation of a high type expert who observed signal $s = a'$ would be $\lambda = 1$ from taking action a', so it would not be an equilibrium to randomize.

Given that any non-pooling, monotone PBE must involve the high type always taking action $a = s$, the expert's strategy can then be represented by the double (σ_1, σ_0) as in the *Non-Transparent Consequences* case. Recall also that monotonicity implies that either $\sigma_1 = 0$ or $\sigma_0 = 1$. We now show that we must have $\sigma_1 = 0$.

To see that we must have $\sigma_1 = 0$, suppose $\sigma_1 > 0$. Then we know $\sigma_0 = 1$ and so can calculate the reputations for the expert after each combination (a, ω):

$$\lambda(1,1|\sigma_1, \sigma_0 = 1) = \frac{\alpha q_h}{\alpha q_h + (1-\alpha)(1-\sigma_1)q_l},$$

$$\lambda(1,0|\sigma_1, \sigma_0 = 1) = \frac{\alpha(1-q_h)}{\alpha(1-q_h) + (1-\alpha)(1-\sigma_1)(1-q_l)},$$

$$\lambda(0,1|\sigma_1, \sigma_0 = 1) = \frac{\alpha(1-q_h)}{\alpha(1-q_h) + (1-\alpha)[(1-q_l) + \sigma_1 q_l]},$$

$$\lambda(0,0|\sigma_1, \sigma_0 = 1) = \frac{\alpha q_h}{\alpha q_h + (1-\alpha)[q_l + \sigma_1(1-q_l)]}.$$

Note that when $q_h = 1$ and $\sigma_1 = 1$, $(a, \omega) = (1, 0)$ is off the equilibrium path. As we are showing an equilibrium cannot exist because the expert would have an incentive to deviate to action $a = 1$, which is least attractive if the belief after $(1, 0)$ is 0, we can, without loss of generality, set $\lambda(1, 0) = 0$.

From the above equations we can note the following properties for any $\sigma_1 > 0$: $\lambda(1, 1) > \lambda(1, 0)$ and $\lambda(0, 0) > \lambda(0, 1)$ so being proven correct is beneficial. In addition we have $\lambda(1, 1) > \lambda(0, 0)$ and $\lambda(1, 0) \geq \lambda(0, 1)$. These inequalities, combined with the observation that $Pr(\omega = 1|t = l, s = 1) > \pi > 1/2$, allow us to conclude that

$$E[\lambda(1, \omega|\sigma_1, \sigma_0 = 1)|t = l, s = 1] > E[\lambda(0, \omega|\sigma_1, \sigma_0 = 1)|t = l, s = 1].$$

As the low type exert would have a strict incentive to take action $a = 1$ rather than $a = 0$ after observing $s = 1$, we cannot have a non-pooling, monotone PBE in which $\sigma_1 > 0$.

We now turn to showing that there exists a unique PBE in which the high type expert always takes action $a = s$ and the low type expert's strategy is represented by $\sigma_1 = 0$ and $\sigma_0 \in [0, 1]$. We begin by calculating, via Bayes's rule, the updated reputation for each action-state pair,

$$\lambda(1, 1|\sigma_0) = \frac{\alpha q_h}{\alpha q_h + (1 - \alpha)[q_l + (1 - \sigma_0)(1 - q_l)]},$$

$$\lambda(1, 0|\sigma_0) = \frac{\alpha(1 - q_h)}{\alpha(1 - q_h) + (1 - \alpha)[(1 - q_l) + (1 - \sigma_0)q_l]},$$

$$\lambda(0, 1|\sigma_0) = \frac{\alpha(1 - q_h)}{\alpha(1 - q_h) + (1 - \alpha)\sigma_0(1 - q_l)},$$

$$\lambda(0, 0|\sigma_0) = \frac{\alpha q_h}{\alpha q_h + (1 - \alpha)\sigma_0 q_l}.$$

Note that when $q_h = 1$ and $\sigma_0 = 0$ the information set $(a, \omega) = (0, 1)$ is off the equilibrium path. As the statement of this lemma says nothing about the uniqueness of off-path beliefs, and because $\lambda(0, 1) = 0$ is the belief most conducive to supporting an equilibrium with $\sigma_0 = 0$, there is no loss of generality in setting the beliefs at this information set $\lambda(0, 1) = 0$.

We now prove that, if the low type expert is optimizing after observing $s = 0$, it will follow that all other experts are optimizing. To see this, note that, as in the previous case, being proven correct is beneficial: $\lambda(0, 0) > \lambda(0, 1)$ and $\lambda(1, 1) > \lambda(1, 0)$ for all σ_0. Hence, if the low type expert is willing to randomize with probability $\sigma_0 \in (0, 1)$ after observing $s = 0$, all experts who observed $s = 1$ have a strict incentive to choose $a = 1$ and a high type expert who observed $s = 0$ has a strict incentive to choose $a = 0$. If $\sigma_0 = 1$ then the low type must have a (weak) incentive to take action $a = 0$ after observing $s = 0$, so the high type has a strict incentive to choose $a = 0$ after observing $s = 0$; further, recall that we established in the first part of the proof that, when $\sigma_0 = 1$, the low type, and therefore also the high type, who observed $s = 1$ will have a strict incentive to choose $a = 1$. Finally, if $\sigma_0 = 0$, the high type who observed $s = 0$ would have an expected reputation of 1 from taking action $a = 0$ and so would have a strict incentive to do so; further, as the low type who observed $s = 0$ has a (weak) incentive to choose $a = 1$ all experts who observe $s = 1$ must have a strict incentive to take action $a = 1$. Hence, if the low type expert is optimizing under the specified strategy, all other experts will have a strict incentive to follow the specified strategies for all $\sigma_0 \in [0, 1]$. So all that remains to show is that there exists a unique $\sigma_0 \in [0, 1]$ such that it is optimal for the low type expert upon observing $s = 0$ to choose $a = 0$ with probability σ_0.

Notice that $\lambda(1, 1)$ is increasing, $\lambda(1, 0)$ is weakly increasing, $\lambda(0, 1)$ is weakly decreasing and $\lambda(0, 0)$ is decreasing in σ_0. Therefore,

$$\psi(\sigma_0) \equiv E[\lambda(0, \omega|\sigma_0) - \lambda(1, \omega|\sigma_0)|t = l, s = 0], \tag{3}$$

is decreasing in σ_0 for all q_h, q_l and π. Hence we can conclude that:

- if $\psi(0) \leq 0$ then we have an equilibrium if and only if $\sigma_0 = 0$.
- if $\psi(0) > 0$ and $\psi(1) < 0$ then there exists a unique $\sigma^* \in (0, 1)$ such that we have an equilibrium if and only if $\sigma_0 = \sigma^*$.
- if $\psi(1) \geq 0$ then we have an equilibrium if and only if $\sigma_0 = 1$.

Therefore, there exists a non-pooling, monotone PBE. In this equilibrium, the high type always takes action $a = s$, and the low type takes action $a = 1$ whenever he observes signal $s = 1$ and action $a = 0$ with some probability $\sigma_0^{FT} \in [0, 1]$ after observing $s = 0$. This equilibrium is unique up to the belief at off-path information sets (if such information sets exist). ■

Lemma 3 *Suppose $\pi \in (q_l, q_h)$ and $j = NA$. Then there exists a unique non-pooling, monotone Perfect Bayesian Equilibrium. In this equilibrium the high type expert always chooses $a = s$ and the low type expert always chooses $a = 1$ regardless of his signal of the state.*

Proof. We first show that in any non-pooling, monotone PBE, $\bar{\lambda}(1) > \bar{\lambda}(0)$. To see this, first note that both actions, $a = 0$ and $a = 1$, must be taken with positive probability. Now, by monotonicity, at least one action $a' \in \{0, 1\}$ must never be taken when the expert observes signal $s = 1 - a'$. Hence the expert who observes $s = a'$ must choose action a' with probability σ_h if they are the high type, and σ_l if they are the low type. Further, by monotonicity, in a non-pooling equilibrium, $\sigma_h > \sigma_l$. Define $\pi' = Pr(\omega = a') \in \{\pi, 1 - \pi\}$. Now we can calculate $\bar{\lambda}(1)$ from the equation,

$$\bar{\lambda}(1) = \frac{Pr(t = h, a = \omega)}{Pr(a = \omega)} = \frac{Pr(t = h, a = \omega = a') + Pr(t = h, a = \omega = 1 - a')}{Pr(a = \omega = a') + Pr(a = \omega = 1 - a')}.$$

Note that, the denominator, $Pr(a = \omega)$, is

$$\pi'[\alpha q_h \sigma_h + (1 - \alpha) q_l \sigma_l] + (1 - \pi')[\alpha q_h + (1 - \alpha) q_l + \alpha(1 - q_h)(1 - \sigma_h) + (1 - \alpha)(1 - q_l)(1 - \sigma_l)].$$

Note further that, because $q_h > \pi$, $\pi' q_h \sigma + (1-\pi')[q_h + (1-\sigma)(1-q_h)]$ is increasing in σ, and, that $\pi' q\sigma + (1-\pi')[q + (1-\sigma)(1-q)]$ is (weakly) increasing in q. Hence, as $\sigma_h > \sigma_l$ and $q_h > q_l$,

$$\pi' q_l \sigma_l + (1-\pi')[q_l + (1-\sigma_l)(1-q_l)] < \pi' q_h \sigma_h + (1-\pi')[q_h + (1-\sigma_h)(1-q_h)],$$

and so $Pr(a = \omega)$ is less than $\pi' q_h \sigma_h + (1-\pi')(q_h + (1-\sigma_h)(1-q_h))$. Therefore,

$$Pr(t = h, a = \omega) = \alpha[\pi' q_h \sigma_h + (1-\pi')(q_h + (1-\sigma_h)(1-q_h))] > \alpha Pr(a = \omega),$$

and so, $\bar{\lambda}(1) > \alpha$. Now, as $\alpha = Pr(a = \omega)\bar{\lambda}(1) + (1 - Pr(a = \omega))\bar{\lambda}(0)$, we can conclude that $\bar{\lambda}(1) > \bar{\lambda}(0)$ in any non-pooling, monotone equilibrium.

Further, as $\pi \in (q_l, q_h)$, we have that,

$$P(\omega = 0|s,t) = \begin{cases} \frac{q_h(1-\pi)}{q_h(1-\pi)+\pi(1-q_h)} > \frac{1}{2} \text{ if } s = 0, t = h, \\ \frac{q_l(1-\pi)}{q_l(1-\pi)+\pi(1-q_l)} < \frac{1}{2} \text{ if } s = 0, t = l, \\ < 1 - \pi < \frac{1}{2} \text{ if } s = 1, t \in \{l, h\}. \end{cases}$$

Since, in any non-pooling, monotone equilibrium, the expert's reputation is strictly higher when his action matches the state, the expert is optimizing if and only if he always takes the action most likely to match the state of world. We can conclude that, there is a unique non-pooling, monotone Perfect Bayesian Equilibrium, and it involves the high type expert takeing action $a = s$ and the low type expert taking action $a = 1$ regardless of his signal. ■

Proof of Proposition 1. This result is immediate combining Lemmas 1-3. ■

Proof of Proposition 2

Proof of Part 1: Suppose $\pi \geq \pi^*$. It is immediate from Lemma 1 that $\sigma_0^{NC} > 0$, and, because $\pi^* > q$, from Lemma 3 that $\sigma_0^{NA} = 0$. So we need only prove that $\sigma_0^{FT} = 0$ in equilibrium under *Full Transparency.* Further, it is sufficient to show this is an equilibrium, since, by Proposition 1, there is unique behavior which constitutes a non-pooling, monotone equiibrium.

In order to verify that it is an equilibrium for all high type experts to choose $a = s$, and all low types to take $a = 1$ regardless of their signal, we first calculate the reputation $\lambda(a, \omega)$ for any action-state pair given these strategies. Note that the pair $(0, 1)$ is off-path given expert behavior; to make the incentive constraints for the low type as easy to satisfy as possible take $\lambda(0, 1) = 0$, though other beliefs may also support this as an equilibrium. Next we note that, as only the high type expert ever chooses $a = 0$, and, as $q_h = 1$, only the low type ever chooses $a = 1$ when $\omega = 0$, the principal must assign beliefs $\lambda(1, 0) = 0$ and $\lambda(0, 0) = 1$. Note also that, by substituting $q_h = 1$ and $\sigma_0 = 0$ into the equation for $\lambda(1, 1|\sigma_0)$ from Lemma 2, we get $\lambda(1, 1) = \alpha$.

In order to have an equilibrium with $\sigma_0 = 0$, the expected reputation from choosing $a = 1$, must be at least as high as the expected reputation from choosing $a = 0$ for a low type who observes signal $s = 0$:

$E(\lambda(1,\omega)|t = l, s = 0) = \alpha Pr(\omega = 1|l, 0) \geq 1 - Pr(\omega = 1|l, 0) = Pr(\omega = 0|l, 0) = E(\lambda(0,\omega)|t = l, s = 0).$

This is satisfied if and only if $Pr(\omega = 1|t = l, s = 0) \geq \frac{1}{1+\alpha}$. As $Pr(\omega = 1|t = l, s = 0) = \frac{\pi(1-q)}{\pi(1-q)+q(1-\pi)}$, this holds if and only if $\pi \geq \pi^*$.

As the incentive constraints for all other expert's are immediate, we can then conclude that, in the unique non-pooling, monotone PBE, the low type expert always takes action $a = 1$ after observing signal $s = 0$ when $\pi \geq \pi^*$ under *Full Transparency.*

Proof of Part 2: We first note that, by Lemma 3, $\sigma_0^{NA} = 0$ on the interval (π_*, π^*) for all $\pi_* \geq q$. We now consider the equations that determine σ_0^{FT} for each prior π. We found in Part 1 that the unique non-pooling, monotone equilibrium involves $\sigma_0^{FT} = 0$ if and only if $\pi \geq \pi^*$. We now consider $\pi < \pi^*$, where we must then have $\sigma_0^{FT} > 0$. Note that when $\sigma_0^{FT} > 0$ there are no off-path information sets, so all beliefs can be derived by Bayes's rule. Substituting $q_h = 1$ and $q_l = q$ into the updated reputations we calculated in the proof of Lemma 2 we have,

$$\lambda(1,1|\sigma_0) = \frac{\alpha}{\alpha + (1-\alpha)[q + (1-q)(1-\sigma_0)]},$$
$$\lambda(1,0|\sigma_0) = \lambda(0,1|\sigma_0) = 0,$$
$$\lambda(0,0|\sigma_0) = \frac{\alpha}{\alpha + (1-\alpha)q\sigma_0}.$$

Now recall the definition of ψ from equation (3), $\psi(\sigma_0;\pi) = E[\lambda(0,\omega|\sigma_0) - \lambda(1,\omega|\sigma_0)|t = l, s = 0]$, and recall that we have an equilibrium with $j = FT$ when $\psi(\sigma_0;\pi) = 0$. Recall also that $\psi(\sigma_0;\pi)$ is decreasing in σ_0. We write ψ as a function of π to make explicit that it depends on π through $Pr(\omega = 0|t = l, s = 0)$. Since $Pr(\omega = 0|t = l, s = 0)$ is continuously decreasing in π, so too is $\psi(\sigma_0;\pi)$. Further, when $\pi = q$, $Pr(\omega = 0|t = l, s = 0) = 1/2$ and so $\psi(1;\pi = q) = 0$ and $\sigma_0^{FT}(q) = 1$. Hence, for all $\pi \in (q, \pi^*)$, $\psi(1;\pi) < 0$, and so there exists a unique solution, $\sigma_0^{FT}(\pi) \in (0,1)$, to $\psi(\sigma_0;\pi) = 0$. Now, as ψ is continuously differentiable and decreasing in σ_0, by the implicit function theorem, $\sigma_0^{FT}(\pi)$ is a continuous function of π on (q, π^*).

Now recall, from equation (2), the definition of $\phi(\sigma_0;\pi) = \lambda(0|\sigma_0,\pi) - \lambda(1|\sigma_0,\pi)$, and recall that, for each π, we have an equilibrium with $j = NC$ when $\phi(\sigma_0;\pi) = 0$. Further, substituting $q_h = 1$ and $q_l = q$ into the updated reputations derived in Lemma 1,

$$\lambda(0|\sigma_0,\pi) = \frac{(1-\pi)\alpha}{(1-\pi)\alpha + (1-\alpha)((1-\pi)q + \pi(1-q))\sigma_0},$$
$$\lambda(1|\sigma_0,\pi) = \frac{\pi\alpha}{1 - [(1-\pi)\alpha + (1-\alpha)((1-\pi)q + \pi(1-q))\sigma_0]}.$$

So we can see immediately that $\phi(\sigma_0;\pi)$ is a continuous in π and σ_0 and decreasing in σ_0. In addition, since $\sigma_0^{FT}(\pi^*) = 0$ and $\sigma_0^{FT}(q) = 1$, whereas, by Lemma 1, $\sigma_0^{NC}(\pi) \in (0,1)$, we have that

$$\phi(\sigma_0^{FT}(\pi^*);\pi^*) > 0 > \phi(\sigma_0^{FT}(q);q).$$

We now define,

$$\pi_* = inf\{\pi \in (q, \pi^*] : \forall \pi' > \pi, \phi(\sigma_0^{FT}(\pi'); \pi') > 0\}.$$

To conclude, as $\phi(\sigma_0^{FT}(\pi^*); \pi^*) > 0$ and $\phi(\sigma_0^{FT}(q); q) < 0$, by the continuity of ϕ, we have $\pi_* \in (q, \pi^*)$. Consequently, for all $\pi \in (\pi_*, \pi^*)$, $\phi(\sigma_0^{FT}(\pi); \pi) > 0$ and so $\sigma_0^{NC}(\pi) > \sigma_0^{FT}(\pi)$. ■

Proof of Proposition 3. Suppose $\pi > \pi_*$. Note first that, by Lemma 3, $\sigma_0^{NA} = 0$ when $\pi \in (q_l, q_h)$. So, as $\pi_* > q_l$, it is sufficient to show that, for all $\pi > \pi_*$, there exists a $\bar{q}_h(q_l, \pi) \geq \pi$ such that $0 < \sigma_0^{FT} < \sigma_0^{NC}$ for all $q_h \in (\bar{q}_h(q_l, \pi), 1)$. First note that it is immediate that $\sigma_0^{FT} > 0$. If $\sigma_0^{FT} = 0$, only high type experts would ever take action $a = 0$ and the principal would then infer that any expert who took action 0 was the high type with certainty, regardless of the consequences, and all experts would then have a strict incentive to take action $a = 0$.

We now establish the second inequality, that $\sigma_0^{FT} < \sigma_0^{NC}$ when $\pi > \pi_*$ and q_h is sufficiently high. Note that, by Lemma 1, $\sigma_0^{NC} \in (0, 1)$ for all paramters. Now recall the definition of $\phi(\sigma_0; q_h) = \lambda(0|\sigma_0, q_h) - \lambda(1|\sigma_0, q_h)$ from equation (2), and allow q_h to vary while holding π and q_l fixed. We first note that $\phi(\sigma_0; q_h)$ is continuously differentiable and strictly decreasing in both its arguments, q_h and σ_0. To see this, recall that

$$\lambda(0|\sigma_0, q_h) = \frac{\alpha[\pi + (1 - 2\pi)q_h]}{\alpha[\pi + (1 - 2\pi)q_h] + (1 - \alpha)((1 - \pi)q_l + \pi(1 - q_l))\sigma_0}.$$

So we can see immediately that, as $\pi > 1/2$ and $(1 - \alpha)((1 - \pi)q_l + \pi(1 - q_l))\sigma_0 > 0$ when $\sigma_0 > 0$, that $\lambda(0|\sigma_0, q_h)$ is strictly decreasing in σ_0 and strictly decreasing in q_h when $\sigma_0 > 0$. Further, recall that

$$\begin{aligned}\lambda(1|\sigma_0, q_h) &= \frac{Pr(t = h, a = 1)}{1 - [Pr(t = h, a = 0) + Pr(t = l, a = 0)]} \\ &= \frac{\alpha[(2\pi - 1)q_h + (1 - \pi)]}{\alpha[(2\pi - 1)q_h + (1 - \pi)] + (1 - \alpha)[1 - ((1 - \pi)q_l + \pi(1 - q_l))\sigma_0]}.\end{aligned}$$

So, as $\pi > 1/2$ and $(1 - \alpha)(1 - ((1 - \pi)q_l + \pi(1 - q_l))\sigma_0) > 0$, we see that $\lambda(1|\sigma_0, q_h)$ is strictly increasing in q_h and σ_0. Hence, we can conclude that $\phi(\sigma_0; q_h) = \lambda(0|\sigma_0, q_h) - \lambda(1|\sigma_0, q_h)$ is strictly decreasing in q_h and σ_0.

Therefore, by the implicit function theorem, we can implicitly define the continuous function $\sigma_0^{NC}(q_h)$ as the solution to $\phi(\sigma_0; q_h) = 0$ for each q_h. Now, recall the definition of $\psi(\sigma_0; q_h)$ from equation (3) and recall also that ψ is continuous and, by Proposition 2, that $\psi(\sigma_0^{NC}(1); q_h = 1) < 0$. We now define, for each π and q_l,

$$\bar{q}_h(q_l, \pi) = inf\{q_h \in (\pi, 1] : \forall q_h' > q_h, \psi(\sigma_0^{NC}(q_h'); q_h') < 0\}.$$

Note that, since $\psi(\sigma_0^{NC}(1); q_h = 1) < 0$, it follows from continuity that $\bar{q}_h(q_l, \pi) < 1$. From the definition of $\bar{q}_h$, for all $q_h \in (\bar{q}_h, 1)$, $\psi(\sigma_0^{NC}(q_h); q_h) < 0$, so, as ψ is decreasing in σ_0 and $\psi(\sigma_0^{FT}(q_h); q_h) = 0$, we have that $\sigma_0^{FT}(q_h) < \sigma_0^{NC}(q_h)$. Hence, for all $q_h \in (\bar{q}_h, 1)$, we can conclude that $0 = \sigma_0^{NA} < \sigma_0^{FT} < \sigma_0^{NC}$. ■