

# On the burden of accountability\*

Paul Steffens<sup>†</sup>

March 30, 2009

## Abstract

An accountable committee is prone to two temptations: pretend agreement when members disagree (to deceive) and deliberately implement a bad project (to distort). The temptation to deceive arises when pretending agreement is rewarded with a high enough probability, and does not depend on reputation concerns. Distortion is induced by reputation concerns as the committee accepts distortion when a reputation gain is attained. When a committee is not accountable reputation concerns can induce distortion. It is demonstrated that accountability (weakly) decreases the quality of decisions. The reason is that reputation beliefs are more pronounced when members are evaluated on earlier remarks.

JEL-codes: .

Keywords: .

---

\*We gratefully acknowledge comments and suggestions by

<sup>†</sup>Tinbergen Institute, Erasmus University Rotterdam. E-mail:

# 1 Introduction

Committees taking decisions with broad consequences are under increasing demand to increase the transparency of their procedures, and to improve the comprehensibility of their decisions. While decisions are commonly observed, the preceding phase of information exchange, deliberation and aggregation of opinions typically is not. Obligations are imposed on committees to realize a higher degree of transparency and comprehension. Examples being the requirement of the FOMC to publish its minutes, and the requirement of boards of directors to motivate their M&A decisions (*references*).

These obligations are attempts to improve accountability. In this paper we determine that imposing accountability has a negative effect on the quality of decisions.

In October 2002 the Amsterdam City Council voted in favour of the construction of the North-South metro line after its members made a personal assessment, and after holding Council hearings. Before construction started an expert report (*reference*) containing a risk assessment was presented to the City Council. Although mildly in tone the report was critical on previous risk assessments: the financial repercussions of the worst case scenario would be more severe than previously predicted. The City Council accepted the report's conclusions yet construction of the metro line proceeded as planned. In recent years the metro line became highly controversial as costs have increased exponentially, (largely due to a more complicated construction environment), putting a heavy burden on the city's budget.

We deduce two questions from this ill-example of decision-making: Why did the City Council not react to the expert report? And whether the City Council's reaction would have been different had it not voted in favour of construction before receiving the expert's report?

We study the decision-making of a committee, whose members have reputation concerns. This committee is obliged to report its intention (implementation or cancelling) before gaining access to a final piece of information. Only thereafter the committee has to decide (implement or cancel). The obligation to report one's intention prior to deciding is interpreted as that the committee is accountable. Our paper's main objective is to determine the effects of imposing accountability on the quality of decisions. In order to attain this objective we compare the quality of decisions under

accountability and non-accountability respectively. Our foremost conclusion is that the quality of decisions under accountability is at most of equal and in many instances of inferior quality.

We assume members care to identical degrees about their reputation, they have homogeneous reputation concerns. Reputation is defined as the expected probability with which a member is perceived able. That decision-makers have reputation concerns is widely acknowledged (Prat), members of a committee appear no exception to this (Levy). Especially, since members are often appointed because of their perceived expected ability. Once appointed members can then be keen to demonstrate their appointment was justifiable.

Members are known to be one of two types: smart or dumb. Smart members receive informative signals, whereas dumb members receive uninformative signals. Since members want to be perceived as smart they are inclined to take actions which put their expected ability in a positive light.

The obligation to report one's intention is reasonable when outsiders have an interest in the odds with which the project will be implemented. For example, since outsiders want to anticipate on implementation. The level of agreement can serve as a indication of the odds with which the project will be implemented. When the time that passes between the initial meeting of a committee and the final decision (implementation or not) is long, the more likely it is that outsiders want to anticipate on the likely decision, and hence the more comprehensible the obligation to report.

We provide a brief sketch of the model's timing.

First, members receive a private signal, which can be informative (depends on members' types). Subsequently, members share these signals. In the appendix it is demonstrated that members share truthfully. Conditional on their signals the committee has to report its level of agreement.

Thereafter, the committee receives a new piece of information (the file), which might warn against implementation. A warning is known to be of perfect information, a non warning is less informative.

Then, the committee has to decide on implementation. It either implements or maintains the status quo.

Depending on the decision and the state of the world project payoffs are realized. Moreover, the market forms beliefs on the members' expected abilities. It can condition these beliefs on the reported level of agreement and the decision taken.

Our three main results depend on two inclinations, which are driven by reputation concerns, and by the probability with which a warning can be received.

(1) Committee is to deceive; to report agreement when members disagree (or report disagreement when members agree)

and

(2) committee is to distort, implement a project that is known to be bad.

The temptation to deceive arises when members acknowledge that they disagree. Since informative signals are correlated, members can be inclined to form a united front to call the image of agreement (Visser and Swank, 2008). This temptation is specific for committees.

The temptation to distort arises when the committee reported agreement and received a warning afterwards. Cancelling the project entails admitting the reception of a warning, which demonstrates that the committee received incorrect signals. The temptation to distort arises for individuals and committees alike. Admitting flaws in private assessment is detrimental for one's reputation.

*First*, the committee is tempted to agree in public prior to taking a decision. Agreeing in public resembles the united front result in Visser and Swank (2008); committees form a united front once decision is taken. Our result is less straightforward and requires conditions on reputation concerns and the warning probability. These additional conditions are required since in our model the committee might receive a warning, and has to take a decision after reporting agreement. Thus, decisions can be affected by reporting agreement.

*Secondly*, obliging committees to claim does not improve quality of decisions. In fact for most reputation concerns accountability yields inferior outcomes as more distortion is spurred. That is under non-accountability the decision-making is of superior quality. The intuition is that when claims are made the reputation gap between implementing and canceling becomes more pronounced.

Committees do not claim disagreement to ensure optimal decisions are taken. By claiming disagreement committee can put reputation concerns offside. Yet the inclination to boost reputation by claiming agreement appears to be dominant.

*Thirdly*, increasing the warning probability has two conflicting effects. On the one hand it implies the quality of information ameliorates as committee is warned more often when it is to implement a bad project. On the other hand it implies that more distortion occurs as a higher

warning probability increases the reputation gap, which on its turn spurs distortion. Better information does not imply per se higher quality of decisions. By disentangling these conflicting effects we determine how optimal decision-making depends on the value of the warning probability.

A brief intuition into the two inclinations that arise because of reputation concerns is useful.

A committee is tempted to deceive when it attains either a gain in reputation *or* in project payoffs. By reporting agreement when members disagree, the committee (possibly) can attain a gain in reputation, which might come at the cost of incurring a project loss.

By reporting disagreement when members agree, the committee attains an inferior reputation, yet it knows its reputation will not be affected by the decision it takes. That is, a committee commits to first best decisions as it is insulated from reputation concerns by reporting disagreement.

When reporting agreement results in distortion the committee can commit to first best decisions by reporting disagreement since the committee insulates itself from reputation concerns. In the equilibria we characterize only the first form of deception exists; the committee reports agreement when members disagree.

A committee is tempted to distort when it reported agreement *and* the file contains a warning. Then a committee faces a trade off: it can attain a reputation gain by implementing yet at the cost of incurring a project loss. A reputation gain can be attained since implementation confirms the committee's initial assessment was likely to be correct, whereas maintenance of the status quo proves that previous assessments were plainly wrong. For strong reputation concerns the committee (at times) implements bad projects when it reported agreement since the reputation gain outweighs the project loss. There is inclination to be consistent., which is in line with Dur (2001) who demonstrates policy makers stick to inefficient policies as admitting flaws in one's judgement signals low ability.

An alternative explanation of the neglectence of new information is offered by the Groupthink hypothesis (Janis, 1981), which suggests that members of a homogeneous group can be inclined not to share and deliberate on new information that goes against the group's previous assessment and opinion.

This paper proceeds as follows. In section 2 the model is explained. Then in section 3 the model is analysed resulting in the characterization of equilibria in Proposition 1. In section 4 we

analyse the model when the committee is not held accountable. Thereafter, in section 5 we compare the outcomes of the accountability and non-accountability models. In section 6 we propose some extensions and elaborate on some discussion points. Section 7 concludes.

## 2 The Model

A two-member committee has to decide on the implementation of a project. Members are distinguished by subscripts,  $i \in \{1, 2\}$ . The committee either maintains the status quo or implements  $X \in \{0, 1\}$ . Maintenance of the status quo generates project payoffs of 0; and implementation of  $p + \mu$ .  $p$  is the members' predisposition, which is assumed to be positive,  $p > 0$ .  $\mu$  reflects that the state of the world can be bad or good,  $\mu \in \{-h, h\}$ . Both states of the world occur with equal probabilities,  $\Pr(\mu = -h) = \Pr(\mu = h) = \frac{1}{2}$ . The project's a priori expected value is positive:  $p + E(\mu) = p > 0$ . Because of the randomness of  $\mu$  project payoffs depend on the realization of the state of the world. When the state of the world is good implementation is optimal,  $p + h > 0$ , whereas when the state of the world is bad maintenance of the status quo is optimal (we assume  $p - h < 0$ ). Members come in two abilities, high and low,  $a_i \in A = \{h, l\}$ . A member does not know his own ability, but he knows the probability,  $\pi$ , with which he is of high ability. Ability determines the informativeness of a member's private assessment. Moreover, we assume members have reputation concerns and want to come across as of high ability.

At  $t = 1$  each member receives a private signal,  $s_i \in S_i = \{s^b, s^g\}$ , about the state of the world ( $b$  stands for bad and  $g$  stands for good). Members of high ability receive informative signals,  $\Pr(\mu = h|s^g, h) = \Pr(\mu = -h|s^b, h) = 1$ , and members of low ability receive uninformative signals,  $\Pr(\mu = h|s^{g,b}, l) = \Pr(\mu = -h|s^{g,b}, l) = \frac{1}{2}$ .

The *ex ante* probabilities of  $\mu$  and  $\pi$  are common knowledge.

At  $t = 2$  the members share their signals truthfully, which as demonstrated in the appendix is incentive compatible. A signal set is formed. Denote the number of good signals by  $|s^g|$ . We assume that for  $|s^g| = 0$  a committee does not arrive in the decision phase. Any committee deciding on reporting and implementation received either  $|s^g| = 1$  or  $|s^g| = 2$ . We refer to  $|s^g| = 1$  as that members disagree, and to  $|s^g| = 2$  as that members agree.

At  $t = 3$  the committee reports its level of agreement,  $r \in R = \{d, a\}$ , with  $r = d$  reflecting that the members disagree, and  $r = a$  that members agree. Reporting is costless and the reported level of agreement does not have to reflect the signal set. The reported level of agreement is observed by the market.

At  $t = 4$  the committee receives a warning or not,  $w$  or  $n$ , depending on the state of the world and the warning probability  $\sigma = \Pr(w|\mu = -h)$ . A warning indicates that the state of the world is bad;  $\Pr(\mu = -h|w) = 1$ . When the state of the world is bad with probability  $\sigma$  a warning is received, and with probability  $1 - \sigma$  no warning is received. When the state of the world is good no warning is received. Consequently, when no warning is received the probability that the state of the world is good increases.

$$\Pr(\mu = h|n, |s^p|) > \Pr(\mu = h||s^p|) \geq \frac{1}{2} \quad \text{for } |s^p| \in \{1, 2\}$$

At  $t = 5$  the committee decides on implementation. It is socially optimal to implement when no warning is received.

At  $t = 6$  the market observes the implementation decision and forms beliefs of the posterior probability that a member is of high ability,  $\Pr(a_i = hg|X, r) = \hat{\pi}_i(r, X)$ . Beliefs are formed conditional on the level of agreement reported, on the implementation decision and on the members' strategies. The state of the world is realized, but not observed by the market, and the members receive their project and reputation payoffs. Members' total payoffs,  $U_i(\cdot)$ , are given by

$$U_i(r, X = 0) = \lambda \hat{\pi}_i(r, X = 0) \tag{1}$$

$$U_i(r, X = 1) = p + \mu + \lambda \hat{\pi}_i(r, X = 1) \tag{2}$$

The parameter  $\lambda$  denotes the weight members attach to their reputation relative to the project payoff, their reputation concerns. We assume the members have identical preferences. Consequently, we can speak of the committee as one entity as they agree on the course of action.

*Solution concept*

Our game is a dynamic game with imperfect information. To solve it, we identify perfect Bayesian equilibria in which (i) the committee’s strategy is optimal and time-consistent given the posterior probabilities, and (ii) posterior probabilities are based on Bayes’ rule.

### 3 Analysis

We start by deriving the conditions under which a first best equilibrium exists. An equilibrium is first best when the committee reports truthfully and takes socially optimal decisions, and thus abstains from deception and distortion. Reporting truthfully requires admitting disagreement, and taking socially optimal decisions requires maintaining the status quo when a warning is received and agreement was pretended.

Consider the beliefs in a first best equilibrium.

Since the committee is believed to report truthfully, pretending agreement indicates that both members received a positive signal, and admitting disagreement indicates that members received conflicting signals.

Suppose agreement is pretended. Maintenance of the status quo indicates a warning is received. Members are considered of low ability since their private assessments were incorrect. In contrast, implementation indicates no warning is received. Consequently, the probability with which a member’s assessment is correct, increases. And so does the expected probability that a member is of high ability.

Suppose disagreement is admitted. At most one member can have made a correct assessment. The posterior probability that a member is of high ability decreases. Moreover, since the states of the world are equally likely a warning (or no warning) does not provide additional information on the probability a member is of high ability. Consequently, updating the probability that a member is of high ability, is only based on admitting disagreement.

Beliefs satisfy

$$\hat{\pi}(r = a, X = 1) > \pi > \hat{\pi}(r = d) > \hat{\pi}(r = a, X = 0) = 0 \tag{3}$$



By applying backward induction, we first determine which condition has to be satisfied for the committee to maintain the status quo when a warning is received given that agreement was reported. Thereafter, assuming the committee takes socially optimal decisions we determine which condition has to be satisfied for the committee to report disagreement when members disagree. We refer to the difference between two levels of reputation as the reputation gap, and to the product of reputation gap and reputation concerns as the reputation gain.

Suppose the committee pretended agreement and received a warning. The reception of a warning implies that the project value is  $p - h$ . The committee faces a trade off between implementing and maintaining the status quo. While implementing yields a superior reputation, it also implies incurring a project loss. The committee maintains the status quo for

$$p - h + \lambda \hat{\pi}(r = a, X = 1) < 0 \quad \text{or} \quad (4)$$

$$\lambda < \lambda^* = \frac{h - p}{\hat{\pi}(r = a, X = 1)}$$

For weak reputation concerns,  $\lambda < \lambda^*$ , the committee maintains the status quo since the reputation gain compensates insufficiently for the project loss.

Suppose members disagree. Pretending *agreement* yields an expected reputation, which is a weighted average of two beliefs. With probability  $1 - \frac{\sigma}{2}$  no warning is received and the committee implements, and with probability  $\frac{\sigma}{2}$  a warning is received and the status quo is maintained. The expected reputation for reporting agreement is  $(1 - \frac{\sigma}{2}) \hat{\pi}(r = a, X = 1)$ . Admitting *disagreement* yields a reputation of  $\hat{\pi}(r = d)$ .

The committee admits disagreement for

$$\hat{\pi}(r = d) > \left(1 - \frac{\sigma}{2}\right) \hat{\pi}(r = a, X = 1) \quad \text{or} \quad (5)$$

$$\sigma > \sigma^* = 2 \left[ \frac{\hat{\pi}(r = a, X = 1) - \hat{\pi}(r = d)}{\hat{\pi}(r = a, X = 1)} \right]$$

We refer to the warning probability as high for  $\sigma > \sigma^*$ , and as low for  $\sigma < \sigma^*$ . When the warning probability is high the committee admits disagreement as the probability with which reporting agreement is punished is large enough. When the warning probability is low the committee pretends

agreement. Note that the decision to admit disagreement (or not) does not depend on reputation concerns. The commission report the level of agreement yielding the superior expected reputation.

The first best equilibrium exists for weak reputation concerns are weak and a high warning probability. When one or both conditions are not satisfied deviations persist in equilibrium. Either the committee pretends agreement when members disagree, referred to as *deception*, or it implements when a warning is received and agreement was pretended, referred to as *distortion*.

We denote  $\alpha = \Pr(r = a | |s^p| = 1)$  as the deception probability, and  $\beta = \Pr(X = 1 | r = a, w)$  as the distortion probability and their equilibrium values by  $\alpha^*$  and  $\beta^*$ .

We provide a number of thresholds, indifference equations and beliefs to characterize the equilibria. Derivation is postponed to the appendix. Thresholds used are the previously derived  $\sigma^*$  and  $\lambda^*$ , and the unknown thresholds  $\lambda^{**}$ ,  $\lambda^{***}$  and  $\tilde{\lambda}$ . The common denominator of reputation thresholds is that these encompass a trade off between reputation and project payoffs. Since choosing the levels of deception and distortion is at the core of the committee's decision-making we provide the equations based on which the committee derives  $\alpha^*$  and  $\beta^*$ .

In case of "a deception only" equilibrium the value of  $\alpha^*$  is determined by

$$(1 - \alpha^*) \left[ \left(1 - \frac{\sigma}{2}\right) \hat{\pi}(r = a, X = 1, \alpha^*) + \frac{\sigma}{2} \hat{\pi}(r = a, X = 0, \alpha^*) - \hat{\pi}(r = d) \right] = 0 \quad (6)$$

Note that reputation concerns play no role in the determination of  $\alpha^*$ .

In case of "a deception / distortion" equilibrium the values of  $\alpha^*$  and  $\beta^*$  are determined by

$$\begin{aligned} (1 - \alpha^*) (1 - \beta^*) \left[ \lambda \left[ \left(1 - \frac{\sigma}{2}\right) \hat{\pi}(r = a, X = 1, \alpha^*, \beta^*) + \frac{\sigma}{2} \hat{\pi}(r = a, X = 0, \alpha^*) - \hat{\pi}(r = d) \right] + \frac{\sigma}{2} \beta^* (p - h) \right] &= (70) \\ (1 - \alpha^*) (1 - \beta^*) [\lambda [\hat{\pi}(r = a, X = 1, \alpha^*, \beta^*) - \hat{\pi}(r = a, X = 0, \alpha^*)] + p - h] &= 0 \end{aligned}$$

In case of "a distortion only" equilibrium the value of  $\beta^*$  is determined by

$$(1 - \beta^*) [\lambda \hat{\pi}(r = a, X = 1, \beta^*) + p - h] = 0 \quad (8)$$

$$\hat{\pi}(r = a, X = 1, \beta^*) > \pi > \hat{\pi}(r = d) = \hat{\pi}(r = a, X = 0) = 0 \quad \text{for } \beta^* > 0 \quad (9)$$

$$\hat{\pi}(r = a, X = 1, \alpha^*) > \pi > \hat{\pi}(r = d) = \hat{\pi}(r = a, X = 0, \alpha^*) > 0 \quad \text{for } \alpha^* > 0 \quad (10)$$

$$\hat{\pi}(r = a, X = 1, \alpha^*, \beta^*) \geq \pi > \hat{\pi}(r = d) = \hat{\pi}(r = a, X = 0, \alpha^*) > 0 \quad \text{for } \alpha^*, \beta^* > 0 \quad (11)$$

**Proposition 1** (A) *If the warning probability is high,  $\sigma > \sigma^*$ ,*

*i) for weak reputation concerns,  $\lambda \leq \lambda^*$ , the committee neither deceives nor distorts;  $\alpha^* = \beta^* = 0$  given beliefs (3)*

*ii) for intermediate reputation concerns, either  $\lambda^* < \lambda < \tilde{\lambda}$  in case of  $\lambda < \tilde{\lambda}$ , or  $\lambda^* < \lambda < \lambda^{***}$  in case of  $\lambda > \tilde{\lambda}$ , the committee does not deceive, and distorts with a strictly positive probability,  $\beta^* > 0$ , where  $\beta^*$  satisfies (??) given beliefs (9)*

*iii) for strong reputation concerns,  $\lambda > \lambda^{***}$ , the committee deceives and distorts with strictly positive probabilities,  $\alpha^* > 0$ , and  $\beta^* > 0$ , where  $\alpha^*$  and  $\beta^*$  satisfy (7) given beliefs (11).*

(B) *If the warning probability is low,  $\sigma \leq \sigma^*$ , the committee deceives with a strictly positive probability,  $\alpha^* > 0$ , and*

*i) for weak reputation concerns,  $\lambda \leq \lambda^{**}$ , the committee does not distort, and  $\alpha^*$  satisfies*

*(6) given beliefs (10)*

*ii) for strong reputation concerns,  $\lambda > \lambda^{**}$ , the committee distorts with a strictly positive probability,  $\beta^* > 0$ , where  $\alpha^*$  and  $\beta^*$  satisfy (7) given beliefs (11).*

Suppose the warning probability is *high*.

For *weak* reputation concerns,  $\lambda < \lambda^*$ , the committee neither deceives nor distorts.

The committee does not deceive, admits disagreement, since this yields a superior reputation. When the warning probability is high the probability that pretending agreement results in a reputation loss is high, making admitting disagreement the more attractive option. The committee does not distort since the project loss is insufficiently compensated by the reputation gain.

For *intermediate* reputation concerns,  $\lambda^* < \lambda < \lambda^{***}$  or  $\lambda^* < \lambda < \tilde{\lambda}$  depending on whether  $\lambda > \tilde{\lambda}$  is satisfied the committee distorts, but not deceives. The committee does not deceive when

the payoffs of admitting disagreement are superior. Project payoffs are superior when disagreement is admitted as distortion is precluded. The reputation when admitting disagreement can be higher or lower than the expected reputation of pretending agreement. When the reputation of admitting disagreement is superior, reporting disagreement is strictly superior and the committee does not deceive. When the reputation of pretending agreement is superior the reporting decision entails a trade off. Admitting disagreement yields superior project and inferior reputation payoffs. Only for not too strong reputation concerns the committee is willing to admit disagreement. Note that the temptation to deceive requires that the committee distorts with some positive probability.

The committee distorts when the reputation gain exceeds the project loss. Distortion occurs when agreement is pretended and a warning is received. The level of distortion  $\beta^*$  increases in reputation concerns.

For *strong* reputation concerns,  $\lambda > \lambda^{***}$ , the committee deceives and distorts. The committee deceives since the expected payoffs of pretending agreement are superior. This requires that the expected reputation of pretending agreement strongly exceeds the reputation of admitting disagreement, which on its turn requires that the committee distorts with a sufficiently high probability. For the committee to consider pretending agreement it is a prerequisite that a superior expected reputation is attained. Moreover, the weight attached to reputation payoffs has to be sufficiently high for total expected payoffs of pretending agreement to be superior, since pretending agreement is inferior projectwise.

Since deception only arises when the committee distorts with a sufficiently high probability it is superfluous for the committee to prefer implementing when a warning is received and agreement has been pretended.

The deception and distortion probabilities  $\alpha^*$  and  $\beta^*$  increase in reputation concerns.

Suppose the warning probability is *low*,  $\sigma < \sigma^*$ .

For *weak* reputation concerns,  $\lambda \leq \lambda^{**}$ , the committee deceives, and does not distort. The committee deceives since the expected reputation of reporting agreement is superior. Note that the temptation to report agreement does not depend on reputation concerns; it is a calculative contemplation. The reporting decision is relatively simple as the committee cannot be induced to

distort. The deception probability  $\alpha^*$  decreases in the warning probability, and does not depend on reputation concerns.

The committee does not distort since the project loss is insufficiently compensated by the reputation gain.

For *strong* reputation concerns,  $\lambda > \lambda^{**}$ , the committee deceives and distorts. The committee's decision to deceive depends on a comparison of expected payoffs for reporting agreement and payoffs of reporting disagreement. Since the committee engages in distortion reporting agreement results in distortion and is inferior projectwise. Deception requires that the expected reputation for reporting agreement strongly exceeds the reputation for reporting disagreement.

The committee distorts when the reputation gain more than compensates for the project loss.

Note the similarities between the equilibria when the committee engages in both deception and distortion. The levels of deception and distortion are determined by the same equations. And for *extremely strong* reputation concerns the committee exhausts its deception and distortion opportunities. It can so since there remains a reputation gap between implementing and maintaining the status quo when agreement was pretended, and between pretending agreement and admitting disagreement.

## 4 A non-accountable committee

Consider a non-accountable committee, which is not obliged to report its level of agreement. We even assume that it cannot communicate with the market in any meaningful way. Then beliefs can only be conditioned on the implementation decision. We start by deriving the conditions under which a first best equilibrium exists.

In a first best equilibrium the status quo is maintained when a warning is received, and the project is implemented when no warning is received. If such an equilibrium exists beliefs satisfy

$$\hat{\pi}(X = 1) > \pi > \hat{\pi}(X = 0) > 0 \tag{12}$$

The beliefs for status quo and implementation are weighted average of the probabilities the status quo results out of two positive signals and a warning respectively conflicting signals and a warning times the beliefs attached to the different information sets

When the project is implemented in any case no warning is received, and either two positive or conflicting signals.

When the status quo is maintained in any case a warning is received, and either two positive or conflicting signals.

When conflicting signals are received the occurrence of a warning (or no warning) is uninformative, whereas when two positive signals and a warning are received members are uncovered as of low ability, since their private assessment was incorrect.

When two positive signals and no warning are received the probability that a member is of high ability increases, since the no warning increases the probability that their private signals are informative. As a consequence, a superior posterior belief is attached to implementation.

Suppose a warning is received. The committee faces the familiar trade off between project and reputation payoffs. By implementing a superior reputation can be attained at the cost of a project loss. The committee does not distort for

$$p - h + \lambda \hat{\pi}(X = 1) < \lambda \hat{\pi}(X = 0) \quad \text{or} \quad (13)$$

$$\lambda < \underline{\lambda} = \frac{h - p}{\hat{\pi}(X = 1) - \hat{\pi}(X = 0)}$$

For weak reputation concerns,  $\lambda < \underline{\lambda}$ , the committee maintains the status quo when a warning is received since the project loss is insufficiently compensated by the reputation gain.

For stronger reputation concerns,  $\lambda > \underline{\lambda}$ , the committee implements with a positive probability when a warning is received, denoted  $\gamma = \Pr(X = 1|w)$ . The equilibrium value of this distortion probability is denoted  $\gamma^*$ .

Consider an equilibrium in which the committee distorts with  $\gamma^* > 0$ . Beliefs based on  $\gamma^*$  satisfy

$$\hat{\pi}(X = 1, \gamma^*) \geq \pi > \hat{\pi}(X = 0) > 0 \quad (14)$$

The belief for implementation decreases in the distortion probability, since the higher  $\gamma^*$  the lower the posterior probability that a member is of high ability when the project is implemented since the probability increases that implementation was granted when a warning was received.

The belief for maintenance of the status quo reflects that the committee received a warning and either two positive or conflicting signals. This belief does not depend on  $\gamma^*$ . Irrespective of the distortion probability it holds that at most one member can have made a correct signal.

The distortion probability  $\gamma^*$  satisfies

$$(1 - \gamma^*) [\lambda (\hat{\pi}(X = 1, \gamma^*) - \hat{\pi}(X = 0)) + p - h] = 0 \quad (15)$$

The stronger reputation concerns the more the committee distorts and hence the higher  $\gamma^*$ . For any value of  $\gamma^*$  a member is believed to be of high ability with a higher probability when the project is implemented. For  $\gamma^* = 1$  the committee always implements, and hence no updating on the members' type can occur,  $\hat{\pi}(X = 1, \gamma^* = 1) = \pi$ .

Since there remains a reputation gap,  $\hat{\pi}(X = 1, \gamma^* = 1) - \hat{\pi}(X = 0) > 0$ , the committee can exhaust ( $\gamma^* = 1$ ) its distortion probability. It does so for extremely strong reputation concerns

$$\begin{aligned} p - h + \lambda\pi &> \lambda\hat{\pi}(X = 0) && \text{or} \\ \lambda &> \bar{\lambda} = \frac{h - p}{\pi - \hat{\pi}(X = 0)} \end{aligned} \quad (16)$$

$$\hat{\pi}(X = 1, \gamma^* = 1) = \pi > \hat{\pi}(X = 0) > 0$$

**Proposition 2** *An equilibrium exists in which the committee*

- (i) *does not distort for weak reputation concerns,  $\lambda < \underline{\lambda}$ , given beliefs (12)*
- (ii) *distorts with a strictly positive probability,  $\gamma^* \in (0, 1)$ , for strong, yet not extremely strong, reputation concerns,  $\bar{\lambda} > \lambda > \underline{\lambda}$ , where  $\gamma^*$  satisfies (15) given beliefs (14).*
- (ii) *distorts with the exhaustive probability,  $\gamma^* = 1$ , for extremely strong reputation concerns,  $\lambda > \bar{\lambda}$*

The committee has only one instrument to boost its reputation which is by implementing. Implementation is in higher esteem than maintenance of the status quo since it indicates private signals were likely to be correct.

Implementation is superior project and reputationwise when no warning is received. Hence it is obvious that implementing is always optimal.

On the contrary, when a warning is received there is the trade off between project and reputation payoffs. For weak reputation concerns the committee does not give in to this temptation as the project loss is insufficiently compensated for by the reputation gain. For strong reputation concerns the committee gives in to the temptation, and implements with a positive probability when a warning is received. The stronger reputation concerns the higher the distortion probability, and for extremely strong reputation concerns the committee exhausts its distortion options.

## **5 Comparison of decision quality: accountability versus non-accountability**

In terms of decision quality it appears that the advantage of accountability is that disagreement can be admitted prior to deciding, which allows for commitment to socially optimal decisions. The drawback of accountability is that the reputation gain (of implementing when a warning is received) increases once the committee reported agreement. The larger the reputation gain the more distortion is spurred.

In this section, we establish a notion of decision quality, and thereupon determine under which protocol the highest quality of decisions is attained. Naturally, any sensible comparison of decision quality requires that we consider equal values of the warning probability and the reputation concerns.

The committee makes two types of mistakes: it can accidentally implement a bad project when no warning was received; and it can deliberately implement a bad project when a warning is received.

The accidental mistake type occurs with some probability when no warning is received, and occurs with equal probabilities under both protocols, since the committee implements when no warning is received irrespective of whether it reported (dis)agreement or not.



The deliberate mistake type occurs when the committee implements when a warning is received. Our comparison of decision quality focuses on this type of mistakes. Since the project loss does not differ between protocols, it is sufficient to compare the probabilities with which a project is implemented when a warning is received.

When comparing the quality of decisions under accountability and non-accountability we make use of the distinction between the high and low warning probability (introduced in the section on accountability).

*First*, suppose the warning probability is low.

Under accountability the committee does not distort for  $\lambda < \lambda^*$ , and under non-accountability the committee does not distort for  $\lambda < \bar{\lambda}$ . And since it holds that  $\bar{\lambda} > \lambda^*$ , we can conclude (1) that distortion occurs for weaker reputation concerns under accountability, and (2) that there exists a domain of reputation concerns,  $\lambda \in (\bar{\lambda}, \lambda^*)$ , for which distortion occurs only under accountability.

For reputation concerns,  $\lambda > \bar{\lambda}$ , distortion occurs under both accountability and non-accountability, whereby satisfy respectively

$$\begin{aligned} \hat{\pi}(r = 2, X = 1, \alpha^*, \beta^*) - \hat{\pi}(r = 2, X = 0, \alpha^*) &\geq \frac{h - p}{\lambda} \\ \hat{\pi}(X = 1, \gamma^*) - \hat{\pi}(X = 0) &\geq \frac{h - p}{\lambda} \end{aligned} \tag{17}$$

The RHSs of both equalities are identical. When both equalities hold the LHSs have to be identical too. For  $\alpha^* < 1$  this requires that  $\beta^*$  exceeds  $\gamma^*$ , whereas for  $\alpha^* = 1$  it requires that  $\beta^*$  equals  $\gamma^*$ .

The quality of decisions under accountability is inferior when the distortion and deception probabilities satisfy

$$\frac{\beta^*}{\gamma^*} \geq \frac{\Pr(|s^p| = 1) + \Pr(|s^p| = 2)}{\Pr(|s^p| = 2) + \alpha^* \Pr(|s^p| = 1)} \tag{18}$$

The distortion probability under accountability has to exceed the distortion accountability under non-accountability with a certain degree, which decreases in the deception probability.

*Secondly*, suppose the warning probability is high.

Under accountability the committee does not distort for

$$\lambda < \lambda^{**} = \frac{h-p}{\widehat{\pi}(r=2, X=1)} \quad (19)$$

and under non-accountability the committee does not distort for

$$\lambda < \bar{\lambda} = \frac{h-p}{\widehat{\pi}(X=1) - \widehat{\pi}(X=0)} \quad (20)$$

Note the denominator of  $\lambda^{**}$  exceeds the denominator of  $\bar{\lambda}$ , and thus  $\bar{\lambda} > \lambda^{**}$ . That is distortion is induced for weaker reputation concerns under accountability, and there exists a domain of reputation concerns,  $\lambda \in (\lambda^{**}, \bar{\lambda})$  for which distortion occurs under accountability, and not under non-accountability.

For stronger reputation concerns,  $\lambda > \bar{\lambda}$ , the committee distorts under both accountability and non-accountability.

The distortion probability  $\alpha^*$  under accountability is determined by

$$(1 - \alpha^*) \left[ \widehat{\pi}(r=2, X=1, \alpha^*, \beta^*) - \widehat{\pi}(r=2, X=0, \alpha^*) - \frac{h-p}{\lambda} \right] = 0$$

and the distortion probability  $\gamma^*$  under non-accountability by

$$(1 - \gamma^*) \left[ \widehat{\pi}(X=1, \gamma^*) - \widehat{\pi}(X=0) - \frac{h-p}{\lambda} \right] = 0$$

The decision quality under accountability is inferior when the  $\frac{\beta^*}{\gamma^*}$ -ratio satisfies

$$\frac{\beta^*}{\gamma^*} \geq \frac{\Pr(|s^p|=1) + \Pr(|s^p|=2)}{\Pr(|s^p|=2) + \alpha^* \Pr(|s^p|=1)} \quad (21)$$

Note that when committee does not deceive ( $\alpha^* = 0$ ) the  $\frac{\beta^*}{\gamma^*}$ -ratio simplifies to

$$\frac{\beta^*}{\gamma^*} \geq \frac{\Pr(|s^p|=1) + \Pr(|s^p|=2)}{\Pr(|s^p|=2)}$$

**Proposition 3** *Comparison of the quality of decisions under the accountability and the non-accountability protocol yields that the non-accountability results in an least as good decisions.*

(1) *Accountability and non-accountability yield an identical quality of decisions when reputation concerns do not induce distortion*

(2) *Non-accountability yields a superior quality of decisions when reputation concerns do not induce distortion under non-accountability whereas they do induce distortion under accountability*

(3) *Non-accountability yields a superior quality of decisions when reputation concerns induce positive yet not exhausted levels of distortion under both accountability and non-accountability.*

(4) *Non-accountability and accountability yield an identical quality of decisions when the committee always distorts and deceives.*

## 6 Extensions

We consider two extensions that offer interesting avenues for future research: productive effort of the public, and heterogeneous reputation concerns.

*First*, consider the public can by exerting productive effort contribute to the project, and consider this effort level can be conditioned on the reported agreement. Both the state of the world and the public's effort determine the project's value in case of implementation. Depending on the circumstances the public is only willing to exert effort when the odds are favourable that the implemented project is good. Since the committee does not bear the cost of effort while it does benefit from exerted effort, it can be tempted to encourage effort exertion by reporting agreement.

In addition to the (possible) reputation advantage of reporting agreement, it now also has the advantage of more effort to be exerted. The model becomes a version of a cheap talk game in the spirit of Sobel, whereby the degree of information transmission depends on differences in preferences.

Secondly, consider heterogeneous reputation concerns. Then signal sharing can be distorted as the members anticipate they are to disagree on the reporting and implementation decision. Members share views which not necessarily reflect signals. Moreover the members no longer agree on which level of agreement to report and whether to implement or not. Denote the member with weakest reputation concerns as the dove and the member with strongest reputation concerns as the hawk. The hawk is more inclined to attain a higher reputation than the dove. By appointing two doves (or two hawks) the sharing of information can be facilitated and decisions are conditioned on optimal information. When members differ strongly in reputation concerns the sharing stage is

strongly distorted, and the voting rule plays a vital role. When the members hold homogeneous reputation concerns they share signals truthfully, and agree on the reporting and implementation decision (making the voting rule obsolete). Consider views are shared sequentially. This allows the last member to condition his view. Members are tempted to manipulate their views. The larger differences in reputation concerns the less information is transmitted.

Heterogeneous reputation concerns can explain why political committees appear prone to leak information. The dove de facto decides which level of agreement to report and whether or not to implement, as he can leak disagreement and the reception of a warning. The more decisive the dove in formal decision-making, the stronger the hawk's incentives to manipulate, and the less information is transmitted. By shifting decision power to the hawk more information is transmitted at the cost of higher levels of deception and distortion. For example, under unanimity the dove is formally decisive, but the hawk has a role in the actual decision right. The hawk provides the dove with his information and tries to direct decisions towards more deception and distortion. This discussion relates to the literature on actual and formal decision rights.

## 7 Conclusions

An accountable committee is prone to two temptations: to deceive: pretend agreement when members disagree, and to distort: deliberately implement a bad project after pretending agreement. Both temptations are driven by reputation concerns. The committee not necessarily gives in to temptations as doing so can imply inferior project payoffs. The general rule is that the committee gives in to temptations for strong enough reputation concerns; the exception being that the committee pretends agreement for any level of reputation concerns when this yields a superior reputation and no change in project payoffs.

The committee is prone to distort as a superior belief is held for implementation. A committee is prone to distortion conditional on having pretended agreement and receiving a warning. By pretending agreement the committee puts its reputation at stake, since it has to confirm the members'

prior assessments were correct. This confirmation requires implementation which is painful only when a warning is received. Then it is certain that a project loss is incurred. Admitting private assessments are incorrect decreases posterior beliefs of the probability that a member is of high ability when the committee pretended agreement.

The committee pretends agreement when pretending agreement yields a (strongly) superior expected reputation. Pretending agreement yields a superior expected reputation for a low probability that pretending agreement results in a reputation loss.

The expected reputation of pretending agreement is superior for two distinct circumstances. Either this occurs when the expected reputation of pretending agreement is superior when the committee does not engage in distortion. Then the low probability with which a reputation loss is incurred, makes the committee willing to take the risk of pretending agreement. Or the expected reputation of pretending agreement is superior since a high level of distortion is induced. The higher the probability with which the committee distorts the lower the probability that pretending agreement results in a reputation loss. Deception is costly when the committee engages in distortion. Then pretending agreement entails that the committee implements a bad project with a positive probability when a warning is received. The committee anticipates on this distortionary loss. Talking is not always cheap. Possibly not understanding this feature plays a major role in the arguments of the advocates of accountability.

When the committee is not accountable it communicates by its decision only. Since a superior belief is held for implementation the committee can be induced to distort. The decision to distort entails a trade off between reputation and project payoffs.

A comparison of the level of distortion (or quality of decisions) demonstrates that under accountability a (weakly) higher level of distortion is induced; and likely a higher level of distortion. A broader range of reputation concerns exists for which no distortion is induced under non-accountability. For stronger reputation concerns distortion occurs under both protocols. A comparison of the level of distortion indicates that under accountability a (weakly) superior level of distortion is induced, or in other words a (weakly) lower quality of decisions.

By pretending agreement the committee puts its reputation at stake as the reputation gap becomes more pronounced. This makes the temptation to give in to distortion stronger, increasing

the level of distortion or inducing distortion. The argument that being able to admit disagreement and thereby protect itself from reputation concerns does not apply, since the more valuable the escape option is the stronger the incentive not to use the escape option because of reputation concerns.

## 8 Appendix

Suppose the warning probability is high

*First*, recall that the high warning probability was one condition required for the first best equilibrium. When reputation concerns are weak,  $\lambda \leq \lambda^* = \frac{h-p}{\hat{\pi}(r=a, X=1)}$ , the committee neither distorts nor deceives.

*Secondly*, consider an equilibrium, in which the committee distorts yet not deceives. Then beliefs satisfy

$$\hat{\pi}(r = a, X = 1, \beta^*) > \pi > \hat{\pi}(r = d) > \hat{\pi}(r = a, X = 0) = 0$$

Beliefs encompass truthful reporting of (dis)agreement. Therefore the belief for  $r = d$  does not alter. Since the committee distorts with  $\beta^* > 0$  the belief for  $r = a$  and  $X = 1$  deflates in  $\beta^*$ . The belief for  $r = a$  and  $X = 0$  does not alter as this indicates both members are of low ability as their signals were incorrect.

Committee distorts for

$$\begin{aligned} p - h + \lambda \hat{\pi}(r = a, X = 1) &> \lambda \hat{\pi}(r = a, X = 0) = 0 \quad \text{or} \\ \lambda &> \lambda^* = \frac{h - p}{\hat{\pi}(r = a, X = 1)} \end{aligned}$$

For strong reputation concerns,  $\lambda > \lambda^*$  the committee distorts since the reputation gain outweighs the project loss. The distortion probability  $\beta^*$  satisfies

$$(1 - \beta^*) [p - h + \lambda \hat{\pi}(r = a, X = 1, \beta^*)] = 0$$

The committee truthfully reports disagreement (anticipating  $\beta^*$ ) when the payoffs for reporting

disagreement exceed payoffs for reporting agreement.

$$\lambda \hat{\pi}(r = d) \geq \lambda \left(1 - \frac{\sigma}{2} (1 - \beta^*)\right) \hat{\pi}(r = a, X = 1, \beta^*) + \frac{\sigma}{2} \beta^* (p - h)$$

Note  $\frac{\sigma}{2} \beta^* (p - h) < 0$ . When the reputation for reporting disagreement exceeds the reputation for reporting agreement

$$\hat{\pi}(r = d) > \left(1 - \frac{\sigma}{2} (1 - \beta^*)\right) \hat{\pi}(r = a, X = 1, \beta^*)$$

the committee is better off by reporting disagreement.

The reputation for reporting disagreement is equal to the expected reputation for reporting agreement when the distortion probability  $\tilde{\beta} \in (0, 1)$  satisfies

$$\hat{\pi}(r = d) = \left(1 - \frac{\sigma}{2} (1 - \tilde{\beta})\right) \hat{\pi}(r = a, X = 1, \tilde{\beta})$$

A solution  $\tilde{\beta} \in (0, 1)$  exists since

$$\begin{aligned} \hat{\pi}(r = d) &> \left(1 - \frac{\sigma}{2}\right) \hat{\pi}(r = a, X = 1) \quad \text{and} \\ \hat{\pi}(r = d) &< \hat{\pi}(r = a, X = 1, \beta = 1) \end{aligned}$$

The level of distortion  $\tilde{\beta}$  is induced for reputation concerns  $\tilde{\lambda} = \frac{h-p}{\hat{\pi}(r=a, X=1, \tilde{\beta})}$ .

For  $\lambda \leq \tilde{\lambda}$  the reputation for reporting disagreement (agreement) is superior; vice versa for  $\lambda > \tilde{\lambda}$ . Recall that reporting disagreement is superior projectwise as no distortion follows after reporting disagreement whereas reporting agreement results in distortion.

For  $\lambda \leq \tilde{\lambda}$  the committee truthfully reports disagreements as this yields superior project and reputation payoffs.

For  $\lambda > \tilde{\lambda}$  the committee might be tempted to deceive since reporting agreement yields superior reputation payoffs. For the committee to deceive the increase in reputation payoffs has to exceed the decrease in project payoffs. And since the increase in reputation payoffs depends on reputation concerns it holds that for sufficiently strong reputation concerns the committee is to deceive.

Assume  $\lambda > \tilde{\lambda}$ . Then the committee truthfully reports disagreement for

$$\begin{aligned} \lambda \hat{\pi}(r = d) &\geq \lambda \left(1 - \frac{\sigma}{2}(1 - \beta^*)\right) \hat{\pi}(r = a, X = 1, \beta^*) + \frac{\sigma}{2} \beta^* (p - h) \quad \text{or} \\ \lambda < \lambda^{**} &= \frac{\frac{\sigma}{2} \beta^* (h - p)}{\left(1 - \frac{\sigma}{2}(1 - \beta^*)\right) \hat{\pi}(r = a, X = 1, \beta^*) - \hat{\pi}(r = d)} \end{aligned}$$

*Thirdly*, we consider an equilibrium in which the committee deceives and distorts. In the characterization of the distortion, no deception equilibrium it was demonstrated that reputation concerns have to be stronger to induce distortion and deception than merely distortion. One condition on reputation concerns is sufficient to establish the proposed equilibrium. Deception has to be induced.

In the proposed equilibrium the beliefs are functions of both  $\alpha^*$  and  $\beta^*$ , and satisfy

$$\hat{\pi}(r = a, X = 1, \alpha^*, \beta^*) \geq \pi > \hat{\pi}(r = d) > \hat{\pi}(r = a, X = 0, \alpha^*) > 0$$

Admitting disagreement  $r = d$  indicates that members disagree; the belief for pretending agreement and implementing ( $r = a$  and  $X = 1$ ) decreases in  $\alpha^*$  and  $\beta^*$ , whereas the belief for pretending agreement and maintaining the status quo ( $r = a$  and  $X = 0$ ) increases in  $\alpha^*$ .

The committee deceives when the expected payoffs of reporting agreement exceed the payoffs of reporting disagreement.

$$\lambda \left(1 - \frac{\sigma}{2}(1 - \beta^*)\right) \hat{\pi}(r = a, X = 1, \beta^*) + \frac{\sigma}{2} \beta^* (p - h) \geq \lambda \hat{\pi}(r = d)$$

Since  $\frac{\sigma}{2} \beta^* (p - h) < 0$  reporting disagreement yields superior project payoffs. For the committee to report agreement reputation payoffs for reporting agreement have to (strongly) exceed reputation payoffs for reporting disagreement. First of all, this requires that reporting agreement yields a superior expected reputation. Assume  $\lambda > \tilde{\lambda}$ .

A stronger condition on reputation concerns is required to ensure reporting agreement as the project loss has to be compensated for.

$$\lambda > \lambda^{**} = \frac{\frac{\sigma}{2} \beta^* (h - p)}{\left(1 - \frac{\sigma}{2}(1 - \beta^*)\right) \hat{\pi}(r = a, X = 1, \beta^*) - \hat{\pi}(r = d)}$$



Committee reports agreement with a positive probability when members disagree for reputation concerns satisfying,  $\lambda > \lambda^{**}$ .

Then deception and distortion probabilities satisfy

$$(1 - \alpha^*)(1 - \beta^*) \left[ \lambda \left[ \left(1 - \frac{\sigma}{2}(1 - \beta^*)\right) \hat{\pi}(r = a, X = 1, \alpha^*, \beta^*) + \frac{\sigma}{2}(1 - \beta^*) \hat{\pi}(r = a, X = 0, \alpha^*) - \hat{\pi}(r = d) \right] + \frac{\sigma}{2}\beta^* \right. \\ \left. (1 - \alpha^*)(1 - \beta^*) \lambda [\hat{\pi}(r = a, X = 1, \alpha^*, \beta^*) - \hat{\pi}(r = a, X = 0, \alpha^*)] \right]$$

Suppose the warning probability is low.

Consider an equilibrium in which the committee deceives, but not distorts. Then beliefs satisfy

$$\hat{\pi}(r = a, X = 1, \alpha^*) > \pi > \hat{\pi}(r = d) > \hat{\pi}(r = a, X = 0, \alpha^*) > 0$$

We consider the committee's commitment to maintain the status quo when a warning is received, and thereafter its inclination to pretend agreement when members disagree.

*First*, given equilibrium beliefs the committee maintains the status quo when a warning is received and agreement was pretended for

$$p - h + \lambda \hat{\pi}(r = a, X = 1, \alpha^*) \leq \lambda \hat{\pi}(r = a, X = 0, \alpha^*) \quad \text{or} \\ \lambda \leq \lambda^{**} = \frac{h - p}{\hat{\pi}(r = a, X = 1, \alpha^*) - \hat{\pi}(r = a, X = 0, \alpha^*)}$$

For weak reputation concerns,  $\lambda \leq \lambda^{**}$ , the committee does not distort since the reputation gain insufficiently compensates for the project loss. Assume  $\lambda \leq \lambda^{**}$ .

*Secondly*, the committee deceives with a positive probability for

$$\sigma < \sigma^* = 2 \left[ \frac{\hat{\pi}(r = a, X = 1) - \hat{\pi}(r = d)}{\hat{\pi}(r = a, X = 1)} \right]$$

When the warning probability is sufficiently small the committee is willing to deceive with a positive probability as only with a small probability pretending agreement results in a reputation loss.

The deception probability  $\alpha^*$  is determined by the subsequent indifference equation, which incorporates that for equilibrium beliefs based on  $\alpha^*$  the committee is indifferent between admitting disagreement and pretending agreement when members disagree. Or the committee strictly prefers pretending agreement for beliefs based on  $\alpha^* = 1$ .

$$(1 - \alpha^*) \left[ \left(1 - \frac{\sigma}{2}\right) \hat{\pi}(r = a, X = 1, \alpha^*) + \frac{\sigma}{2} \hat{\pi}(r = a, X = 0, \alpha^*) - \hat{\pi}(r = d) \right] = 0$$

The deception probability  $\alpha^*$  decreases in the warning probability  $\sigma$ . The lower the warning probability the more attractive it is to pretend agreement when members disagree.

When the warning probability is extremely low the committee pretends agreement with  $\alpha^* = 1$ , since

$$\begin{aligned} \left(1 - \frac{\sigma}{2}\right) \hat{\pi}(r = a, X = 1, \alpha^* = 1) + \frac{\sigma}{2} \hat{\pi}(r = a, X = 0, \alpha^* = 1) &> \hat{\pi}(d = 1) \quad \text{or} \\ \sigma < \sigma^{**} = 2 \left[ \frac{\hat{\pi}(r = a, X = 1, \alpha^* = 1) - \hat{\pi}(d = 1)}{\hat{\pi}(r = a, X = 1, \alpha^* = 1) - \hat{\pi}(r = a, X = 0, \alpha^* = 1)} \right] \end{aligned}$$

Consider an equilibrium in which the committee deceives and distorts with positive probabilities. Beliefs are functions of  $\alpha^*$  and  $\beta^*$ , and satisfy

$$\hat{\pi}(r = a, X = 1, \alpha^*, \beta^*) \geq \pi > \hat{\pi}(r = d) > \hat{\pi}(r = a, X = 0, \alpha^*) > 0$$

The beliefs satisfy the order that is familiar by now. The belief for  $r = a$  and  $X = 1$  deflates in  $\alpha^*$  and  $\beta^*$  as the event becomes less informative on a member's probability to have made a correct assessment. When the committee deceives and distorts with probabilities  $\alpha^* = \beta^* = 1$  the committee always ends up in the event  $r = a$  and  $X = 1$ , and hence this event does not allow for updating:  $\hat{\pi}(r = a, X = 1, \alpha^* = \beta^* = 1) = \pi$ . For all other values of  $\alpha^*$  and  $\beta^*$  implementation after pretending agreement entails a reputation gain in comparison with the prior. The belief for  $r = a$  and  $X = 0$  increases in  $\alpha^*$  as the more the committee pretends agreement when members

disagree the higher the probability the members disagree when agreement was pretended. The belief for  $r = d$  is a constant factor, and reflects that the committee admitted disagreement which is only done when members disagree. Moreover, then the decision provides no further information on members.

The committee has to prefer implementing when a warning is received, and it pretended agreement. It does so when implementing yields superior payoffs

$$p - h + \lambda \hat{\pi}(r = a, X = 1, \alpha^*) \geq \lambda \hat{\pi}(r = a, X = 0, \alpha^*) \quad \text{or}$$

$$\lambda > \lambda^{**} = \frac{h - p}{\hat{\pi}(r = a, X = 1, \alpha^*) - \hat{\pi}(r = a, X = 0, \alpha^*)}$$

For strong reputation concerns,  $\lambda > \lambda^{**}$ , the committee distorts with a positive probability as given beliefs based on  $\alpha^*$  and  $\beta = 0$  the reputation gain exceeds the project loss.

Recall that the committee already deceives when it does not distort. Then when the committee does distort the incentive to deceive becomes even stronger.

The deception and distortion probabilities are determined by

$$(1 - \alpha^*)(1 - \beta^*) \left[ \lambda \left[ \left(1 - \frac{\sigma}{2}(1 - \beta^*)\right) \hat{\pi}(r = a, X = 1, \alpha^*, \beta^*) + \frac{\sigma}{2}(1 - \beta^*) \hat{\pi}(r = a, X = 0, \alpha^*) - \hat{\pi}(r = d) \right] + \frac{\sigma}{2} \beta^* \right. \\ \left. (1 - \alpha^*)(1 - \beta^*) \left[ \hat{\pi}(r = a, X = 1, \alpha^*, \beta^*) - \hat{\pi}(r = a, X = 0, \alpha^*) \right] \right]$$

Note that a high warning probability,  $\sigma > \sigma^*$ , only possibly exists for  $\sigma^* < 1$ . This holds when the reputation ratio is moderate and satisfies,  $\frac{\hat{\pi}(r=a, X=1)}{\hat{\pi}(r=d)} < 2$ . We assume that the reputation ratio is moderate.