

TI 2018-019/I  
Tinbergen Institute Discussion Paper



# Denial and Alarmism in Collective Action Problems

Manuel Foerster<sup>1</sup>  
Joel (J.J.) van der Weele<sup>2</sup>

1: University of Hamburg

2: Universiteit van Amsterdam; Tinbergen Institute, The Netherlands

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

Contact: [discussionpapers@tinbergen.nl](mailto:discussionpapers@tinbergen.nl)

More TI discussion papers can be downloaded at the [Tinbergen Site](#)

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam  
Gustav Mahlerplein 117  
1082 MS Amsterdam  
The Netherlands  
Tel.: +31(0)20 598 4580

Tinbergen Institute Rotterdam  
Burg. Oudlaan 50  
3062 PA Rotterdam  
The Netherlands  
Tel.: +31(0)10 408 8900

# Denial and Alarmism in Collective Action Problems

Manuel Foerster\*      Joël J. van der Weele†

February 28, 2018

## Abstract

We analyze communication about the social returns to investment in a public good. We model two agents who have private information about these returns as well as their own taste for cooperation, or social preferences. Before deciding to contribute or not, each agent submits an unverifiable report about the returns to the other agent. We show that even if the public good benefits both agents, there are incentives to misrepresent information. First, others' willingness to cooperate generates an incentive for "alarmism", the exaggeration of social returns in order to opportunistically induce more investment. Second, if people also want to be *perceived* as cooperators, a "justification motive" arises for low contributors. As a result, equilibrium communication features "denial" about the returns, depressing contributions. We illustrate the model in the context of institutional inertia and the climate change debate.

**JEL classification:** C72, D64, D82, D83, D91.

**Keywords:** cheap talk, cooperation, image concerns, information aggregation, public goods.

---

\*University of Hamburg, email: manuel.foerster@wiso.uni-hamburg.de.

†University of Amsterdam, Tinbergen Institute, email: vdweele@uva.nl.

We would like to thank Roland Bénabou, Anke Gerber, Gero Henseler, Andreas Nicklisch, Jeroen van de Ven, Achim Voss, and seminar participants at the WZB Berlin, the University of Amsterdam and the University of Hamburg for useful comments. We thank Ivar Kolvoort for research assistance. Joël van der Weele gratefully acknowledges financial support from the NWO through VIDI grant 452-17-004.

# 1 Introduction

Many public goods have uncertain social returns to investment. To determine and organize the appropriate level of collective action, a constructive aggregation of all available information is crucial. In practice however, debates about public goods are often fraught and result in gridlock and inaction. One example is institutional change in organizations, where benefits to different stakeholders are often uncertain. Change usually generates resistance, which hampers communication and leads to organizational inertia. Another example is the problem of climate change. While the large majority of scientists agree that man-made climate change is happening, there is uncertainty and disagreement about the efficient level of mitigation of greenhouse gases. The public debate about climate change is highly polarized and politicized, with substantial minorities in most countries disbelieving the scientific consensus.

To better understand information sharing in public good environments, we investigate a stylized model of communication. Two agents each receive a private signal about the return to the public good, derived from personal experience or other private sources. Both agents also have private information about their intrinsic motivation to contribute to the public good. In the communication stage, each agent submits a public report about the social returns of the public good. There is no cost of submitting a biased report, so communication is cheap talk. After observing both reports, agents choose their contribution level, which is observed by the other agent.

In this context, we consider the possibility of truthful communication under different preference structures, derived from the empirical literature on public good contributions. First, we assume that some agents care sufficiently about the public good to contribute when they perceive the return to be high. We show that this introduces an incentive for “alarmism”: the opportunistic exaggeration of the social return on investment in order to induce contributions by others. Alarmism prevents truthful communication, causing agents’ actions to be based solely on their own private signal. Some alarmists will also be “hypocritical”, in that they do not contribute themselves. Hypocrisy is often seen as a sign of bad character, leading to a loss of reputation.

To capture the reputation consequences of dishonest communication in our model, we introduce an “image concern” to be perceived as a cooperative type. This is in line with a large literature on the importance of audience effects in social dilemmas. We show that image concerns create the possibility for truthful communication by deterring hypocritical reports. However, image concerns also introduce a “justification motive”: agents with low intrinsic motivation to contribute will downplay the returns to social investment, in order to justify their selfish actions. This strategy increases the likelihood that both types pool on inaction, thus avoiding a separating equilibrium where low contributors receive a low image. The result is equilibrium communication that is characterized by “denial”, an overrepresentation of negative signals about the social

returns.

From a welfare perspective, the denial equilibrium features inefficiently low contributions compared to fully honest communication, particularly if the costs of contributing to the public good and image concerns are relatively low. However, when the costs of contributing are rather high, denial may actually increase welfare by reducing over-contributions by agents with high intrinsic motivation to contribute. In particular, we demonstrate that image concerns may increase the efficiency of public goods provision by enabling the denial equilibrium.

Our stylized model shows that denial and alarmism pose fundamental problems for decentralized information aggregation in public good environments, *even if* all agents benefit from public good contributions. In Section 6, we will discuss corroborating evidence in the context of institutional inertia and the climate change debate, showing that denial results from a reluctance to change personal behavior and the wish to maintain a reputation as a good employee or citizen. Finally, we will extend our model to impure public goods and discuss evidence in the context of charitable giving in Section 7.

## 2 Literature

This paper combines different strands of literature to investigate communication in public good environments. The first is the empirical literature on heterogeneous motives that drive people to contribute to public goods, revealed in laboratory and field experiments. A minority of agents act as free riders or unconditional altruists, while others are willing to contribute as long as enough people do so (Fischbacher, Gächter, and Fehr, 2001; Kurzban and Houser, 2005; Burlando and Guala, 2004). In addition to these “social preferences”, people care about others’ *perceptions* of their degree of altruism or civic-mindedness. Such audience effects are an important driver of contributions in social dilemma settings, as has been demonstrated in the lab (e.g. Rege and Telle, 2004; Andreoni and Petrie, 2004; Andreoni and Bernheim, 2009; Ariely, Bracha, and Meier, 2009) and in the field (e.g. Harbaugh, 1998; Soetevent, 2005; Lacetera and Macis, 2010; Karlan and McConnell, 2014).

The way we incorporate preference heterogeneity and image concerns in our model is inspired by Ellingsen and Johannesson (2008) and Ali and Bénabou (2016).<sup>1</sup> Private information about the level of “altruism” is conveyed through costly signaling. The main difference from traditional signaling models like Spence (1973) is that the agent cares directly about the beliefs of the observer instead of only her actions. Formally, this turns the model into a psychological game (Geanakoplos, Pearce, and Stacchetti, 1989). This modeling strategy is not inconsistent with a view where a good reputation serves the individual in the longer run by generating profitable interactions.

---

<sup>1</sup>See Bénabou and Tirole (2006), Andreoni and Bernheim (2009) and Grossman and van der Weele (2017) for closely related models.

The second ingredient of our model is the use of pre-play communication in the form of cheap talk (Crawford and Sobel, 1982). There exists a large experimental literature on pre-play communication in dilemma problems. This literature focuses mostly on the communication of intentions to cooperate (Sally, 1995). One exception is Serra-Garcia, Damme, and Potters (2011), who consider an experimental public good game in which one player has private information about the payoffs of the game. They find that if the return to public goods is such that it gives rise to a prisoner’s dilemma, informed players resort to either lies or vague messages to obfuscate this fact.

When it comes to formal theories of pre-play communication, several papers focus on coordination games rather than public good environments that are the focus of our paper. Ellingsen and Östling (2010) investigate how pre-play communication of intentions can improve outcomes in such games under  $k$ -level thinking. Hagenbach and Koessler (2010) and Galeotti, Ghiglino, and Squintani (2013) study cheap talk about the parameters of the game and employ statistical models related to our framework. Hagenbach, Koessler, and Perez-Richet (2014) look at the possibility of truthful communication in a general class of games, where messages are verifiable.

To our knowledge, this is one of the first papers to explore the relation between pre-play communication and the image associated with subsequent actions. Some similar ideas are explored in Bénabou, Falk, and Tirole (2018), where agents with image concerns first search for and then disclose verifiable information about the size of an externality. Agents may withhold positive information to justify their inaction, similar to the “justification motive” that drives denial in our framework. Furthermore, agents have an “influence motive” to increase contributions by others similar to our incentive for hypocrisy/alarmism. The authors also study communication of an informed principal with an image-concerned agent who may contribute to the public good. The principal may either disclose verifiable information or submit a cheap-talk message on her preferred action. Instead, our paper considers simultaneous cheap talk between two image-concerned agents, both of whom may contribute. This combination of cheap talk with subsequent actions also requires us to develop a new refinement criterion, which as far as we know is novel in the literature.

Several studies explore how agents may change the parameters of the game to influence subsequent signaling equilibria. In Henry and Louis-Sidois (2015), agents vote on sanctions for non-compliance in a public good environment. When the vote is secret, some contributors may vote against sanctions, in order to increase the signaling value of their contributions in the subsequent public good problem. Bénabou and Tirole (2011) investigate how an authority can affect compliance by agents with prosocial and image concerns, by managing information about ethical values and compliance in society. In Ali and Bénabou (2016), a principal optimally sets the level of privacy or image concerns. Higher image concerns lead to socially desirable behavior, but also induce conformity that obscures shifts in societal values over time.

Finally, our paper relates to Kuran (1997)’s concept of “preference falsification”, the public misrepresentation of private preferences or opinions. In Kuran’s theory, misrepresentation occurs in order to conform with the majority opinion, whatever that happens to be. In our theory, misrepresentation serves to influence other agents’ actions (alarmism) or to justify one’s own selfish actions (denial). Since our theory formally models the Bayesian formation of beliefs, it offers more specific predictions about the direction of misrepresentation and its welfare consequences.

### 3 Model and notation

In our model there are two agents, indexed  $i = 1, 2$ , who have the option to *contribute* to a public good,  $\hat{a}_i = 1$ , or not,  $\hat{a}_i = 0$ .<sup>2</sup> This contribution, also called (*prosocial*) *action*, has *cost*  $c > 0$  to the agent, but a positive *benefit*  $W$  for both parties. In addition,  $W$  could confer advantages to other agents who are not explicitly modeled and not part of the communication environment.

The value of this externality  $W$  is unclear, but known to be uniformly distributed on  $[0, 1]$ .<sup>3</sup> First, each agent receives an unbiased but noisy *signal*  $s_i \in S = \{0, 1\}$  about  $W$ , where  $s_i = 1$  with probability  $W$  and  $s_i = 0$  with probability  $1 - W$ . Second, each agent  $i$  submits a *report*  $\hat{m}_i \in M = \{0, 1\}$  about her signal  $s_i$  via cheap talk, that is, she can lie. Third, each agent observes the report  $\hat{m}_j$  of the other agent  $j \neq i$  and decides whether to contribute to the public good. Finally, each agent observes the action of the other agent.

Agents’ beliefs are updated based on the standard Beta-binomial model. Agent  $i$ ’s posterior is  $f(W|s_i) = 2W^{s_i}(1 - W)^{1-s_i}$  if she only knows her own private signal and  $f(W|s_i, s_j) = 6/((s_i + s_j)!(2 - s_i - s_j)!)W^{s_i+s_j}(1 - W)^{2-s_i-s_j}$  if she also knows the signal of the other agent. Hence,

$$E[W|s_i] = \frac{1 + s_i}{3} \text{ and } E[W|s_i, s_j] = \frac{1 + s_i + s_j}{4}.$$

The *preferences* of agent  $i$  are given as:

$$u_i(\theta_i, s_i, \hat{m}, \hat{a}) \equiv (1 + \theta_i)(\hat{a}_i + \hat{a}_j)W - \hat{a}_i c + \mu E_i [E_j[\theta_i|\theta_j, s_j, \hat{m}_i, \hat{a}_i] | \theta_i, s_i, \hat{m}_j, \hat{a}_j]. \quad (1)$$

Here, the first term represents the benefits from the public good, which are equal to the sum of the uncertain benefits from each contribution. Additionally, these benefits are increased depending on the agent’s preference parameter  $\theta_i \in \Theta = \{0, 1\}$ , which is private information and takes the value of 1 with prior probability  $\pi \in (0, 1)$  and the value of 0 with prior probability  $1 - \pi$ . We refer to  $\theta_i$  as the *type* of the agent,

<sup>2</sup>To distinguish actual decisions of the agents from strategies, we indicate them by a “hat” symbol.

<sup>3</sup>This assumption merely eases the exposition. Our results are qualitatively robust to changes in the distribution of  $W$ , requiring only a continuous and strictly positive density.

and will refer to  $\theta_i = 0$  as a *low type* and  $\theta_i = 1$  as a *high type*. One can interpret  $\theta_i$  as the degree of intrinsic motivation to contribute to the public good, which measures the (psychological) payoff an agent derives from increasing the public good and societal welfare.<sup>4</sup> Thus, each agent  $i$  has two pieces of private information, denoted as a *type-signal pair*  $(\theta_i, s_i) \in \Theta \times S$ . In Section 7, we extend the model to impure public goods and lower intrinsic motivation of high types to contribute.

The second term of (1) represents the cost of contributing. The last term reflects *image concerns*: the expectation of agent  $i$  about what the other agent  $j$  infers about her type, which follows the modeling in Ellingsen and Johannesson (2008) and Ali and Bénabou (2016). We assume that the inference of agent  $j$  about  $i$  can depend on the report  $\hat{m}_i$  and action  $\hat{a}_i$  of the other agent as well as her own signal  $s_j$ . The parameter  $\mu > 0$  measures the importance of these image concerns to the agent. While we assume that people care directly about the inferences of others about their character, one could see this as a proxy for the continuation value in a game in which agents with a good reputation will reap additional benefits from future interactions.

**Timing.** Summarizing the arguments above, the timing is as follows:

1. Nature determines the state  $W \in [0, 1]$  and the types  $\theta_i \in \Theta$  of both agents.
2. Each agent receives a signal  $s_i \in S$  about  $W$ .
3. Each agent submits a report  $\hat{m}_i \in M$ .
4. Each agent observes  $\hat{m}_j$ , and decides whether to contribute,  $\hat{a}_i \in \{0, 1\}$ .
5. Each agent observes the action  $\hat{a}_j$  of the other agent.

**Parameter restrictions.** We make the following assumption on the parameters.

**Assumption 1.**  $3/4 < c - \mu < 3/2$ .

Assumption 1 says that image concerns are neither very large nor very small relative to the cost of contributing. The first inequality rules out equilibria where all agents contribute purely for image reasons. More precisely, low types will never contribute as contributing is strictly dominated for them for any signal structure. We henceforth assume without further statement that low types do not contribute to the public good. The second inequality ensures that high types contribute at least under some circumstances.

---

<sup>4</sup>In principle,  $\theta_i$  could also represent differences in the economic benefit from the public good that are private information. However, it is not clear that people would care about the inferences other people make about such benefits.



**Solution Concept.** The solution concept we employ is perfect Bayesian equilibrium. We restrict our attention to pure strategies. A (*pure*) *strategy*  $(m_i, a_i)$  for agent  $i$  is a pair of mappings

$$m_i : \Theta \times S \rightarrow M \text{ and } a_i : \Theta \times S \times M \rightarrow \{0, 1\}$$

which assign a report to each type-signal pair (first stage) and an action to each type-signal pair and report received from the other agent (second stage), respectively.<sup>5</sup> We denote the set of strategy profiles by  $\mathcal{S}$ . Let  $\tilde{\theta}_i(\theta_j, s_j, \hat{m}_i, \hat{a}_i) \equiv E_j[\theta_i | \theta_j, s_j, \hat{m}_i, \hat{a}_i]$  denote the posterior belief of agent  $j$  about agent  $i$  based on her type  $\theta_j$ , signal  $s_j$  and  $i$ 's report  $\hat{m}_i$  and action  $\hat{a}_i$ . We denote the set of posterior belief systems that are consistent with strategy profile  $(m, a) \in \mathcal{S}$  by  $\Delta^*(m, a)$ .

**Definition 1** (Perfect Bayesian equilibrium). *A strategy profile  $(m^*, a^*) \in \mathcal{S}$  is a “perfect Bayesian equilibrium” of the game if, for some  $\tilde{\theta} \in \Delta^*(m^*, a^*)$ , all agents  $i = 1, 2$  and type-signal pairs  $(\theta_i, s_i) \in \Theta \times S$ ,*

$$m_i^*(\theta_i, s_i) \in \operatorname{argmax}_{m_i} E_i[u_i(\theta_i, s_i, m_i, m_j^*, a^*) \mid \theta_i, s_i, \tilde{\theta}_i], \text{ and}$$

$$a_i^*(\theta_i, s_i, \hat{m}_j) \in \operatorname{argmax}_{a_i} E_i[u_i(\theta_i, s_i, m_i^*, \hat{m}_j, a_i, a_j^*) \mid \theta_i, s_i, \hat{m}_j, \tilde{\theta}_i] \text{ for all } \hat{m}_j \in M.$$

We restrict our attention to symmetric equilibria with positive measure of support in the parameter space. Equilibria with positive measure of support can be considered as robust with respect to changes in the underlying parameters.<sup>6</sup>

**Off-equilibrium beliefs.** As in most (costly) signaling games, some criterion is necessary to rule out equilibria based on implausible off-equilibrium beliefs. In our case, this is more complex than usual as we need to check for possible deviations at each of the two stages of the game. To achieve this, we extend the D1-criterion by Cho and Kreps (1987), see also Banks and Sobel (1987). We proceed backwards and first apply the D1-criterion to the second stage, conditional on the reports that have been submitted in the first stage. That is, we check which type is most likely to have deviated to a certain off-equilibrium action. Given this restriction on beliefs, we go to the first stage and again apply the D1-criterion. We check which type is most likely to have deviated to a strategy that involves the observed message and action (possibly conditional on the report submitted by the other player). This yields a further restriction on beliefs, and we reject an equilibrium candidate if some off-equilibrium deviation is still profitable despite the restriction.

<sup>5</sup>Notice that without loss of generality the strategy of player  $i$  in the second stage does not explicitly depend on the report that she has submitted in the first stage. For each type-signal pair, the report submitted is given by the strategy in the first stage.

<sup>6</sup>Without the restriction to equilibria with positive measure of support, there would exist an equilibrium in which all communication is truthful iff  $\mu = 1/2$ , see Appendix A.2 for details.

Formally, consider any equilibrium  $(m^*, a^*)$  and let  $\tilde{\theta}_i$  denote any posterior belief system for agent  $j$  regarding the type of player  $i$  that is consistent with  $(m^*, a^*)$ . Furthermore, let

$$U_i^*(m_i, a_i | \theta_i, s_i, \tilde{\theta}_i) \equiv E_i[u_i(\theta_i, s_i, m_i, m_j^*, a_i, a_j^*) | \theta_i, s_i, \tilde{\theta}_i]$$

and

$$U_i^*(m_i, a_i | \theta_i, s_i, \hat{m}_j, \tilde{\theta}_i) \equiv E_i[u_i(\theta_i, s_i, m_i, \hat{m}_j, a_i, a_j^*) | \theta_i, s_i, \hat{m}_j, \tilde{\theta}_i]$$

denote agent  $i$ 's expected utility from strategy  $(m_i, a_i)$  before and after receiving the report  $\hat{m}_j$  respectively, conditional on her type  $\theta_i$  and signal  $s_i$  and the belief system  $\tilde{\theta}_i$  of agent  $j$ , and when agent  $j$  behaves according to the equilibrium strategy  $(m_j^*, a_j^*)$ . Next, consider any off-equilibrium choices  $(\hat{m}_i, \hat{a}_i)$  of agent  $i$ . First, we define

$$D^*(\theta_i, s_i, \hat{m}, \hat{a}_i) \equiv \bigcup_{\tilde{\theta} \in \Delta^*(m^*, a^*)} \left\{ \tilde{\theta}_i(\cdot, \hat{m}_i, \hat{a}_i) \mid U_i^*(\hat{m}_i, a_i^* | \theta_i, s_i, \hat{m}_j, \tilde{\theta}_i) \leq U_i^*(\hat{m}_i, \hat{a}_i | \theta_i, s_i, \hat{m}_j, \tilde{\theta}_i) \right\}$$

for each report  $\hat{m}_j \in m_j^*(\Theta \times S)$  of agent  $j \neq i$  and let

$$\Theta^*(\hat{m}, \hat{a}_i) \equiv \left\{ \theta \in \Theta \mid \nexists \theta' \neq \theta, s' : D^*(\theta, s, \hat{m}, \hat{a}_i) \subseteq (\subsetneq) D^*(\theta', s', \hat{m}, \hat{a}_i) \text{ for all (some) } s \right\}$$

denote the set of types that are most likely, that is, for the largest set of beliefs that are consistent with the equilibrium, to take action  $\hat{a}_i$  in the second stage, conditional on the submitted reports  $\hat{m}$ . Second, let  $\Delta(\cdot)$  denote the set of probability distributions over a finite set and define

$$D^{**}(\theta_i, s_i, \hat{m}_i, a_i) \equiv \bigcup_{\tilde{\theta} \in \Delta^*(m^*, a^*)} \left\{ \tilde{\theta}_i(\cdot, \hat{m}_i, a_i) \mid \tilde{\theta}_i(\cdot, \hat{m}_i, \hat{a}_i) \in \Delta(\Theta^*(\hat{m}_i, m_j^*, \hat{a}_i)), \right. \\ \left. U_i^*(m_i^*, a_i^* | \theta_i, s_i, \tilde{\theta}_i) \leq U_i^*(\hat{m}_i, a_i | \theta_i, s_i, \tilde{\theta}_i) \right\}$$

for any  $a_i \in A_i^*(\theta_i, s_i, \hat{a}_i) \equiv \{a_i \mid a_i(\theta_i, s_i, \hat{m}_j) = \hat{a}_i \text{ for some } \hat{m}_j \in m_j^*(\Theta \times S)\}$  and let

$$\Theta^{**}(\hat{m}_i, \hat{a}_i) \equiv \left\{ \theta \in \Theta \mid \nexists \theta' \neq \theta, s', a'_i \in A_i^*(\theta', s', \hat{a}_i) : \right. \\ \left. D^{**}(\theta, s, \hat{m}_i, a_i) \subseteq (\subsetneq) D^{**}(\theta', s', \hat{m}_i, a'_i) \text{ for all (some) } s, a_i \in A_i^*(\theta, s, \hat{a}_i) \right\}$$

denote the set of types that ex-ante are most likely, taking into account the first restriction of beliefs, to deviate to  $(\hat{m}_i, \hat{a}_i)$ . Finally, we reject the equilibrium  $(m^*, a^*)$  if this deviation is still profitable despite the restriction of beliefs to  $\Theta^{**}(\hat{m}_i, \hat{a}_i)$ .

**Definition 2** (Double D1-criterion). *Consider a perfect Bayesian equilibrium  $(m^*, a^*)$ , any agent  $i = 1, 2$  and off-equilibrium choices  $(\hat{m}_i, \hat{a}_i)$ . If for any  $\theta_i \in \Theta^{**}(\hat{m}_i, \hat{a}_i)$ ,*

$s_i \in S$  and  $a_i \in A_i^*(\theta_i, s_i, \hat{a}_i)$ ,

$$U_i^*(m_i^*, a_i^* | \theta_i, s_i, \tilde{\theta}_i) < U_i^*(\hat{m}_i, a_i | \theta_i, s_i, \tilde{\theta}_i)$$

for all  $\tilde{\theta} \in \Delta^*(m^*, a^*)$  such that  $\tilde{\theta}_i(\cdot, \hat{m}_i, \hat{a}_i) \in \Delta(\Theta^{**}(\hat{m}_i, \hat{a}_i))$ , then we say that  $(m^*, a^*)$  violates the “double D1-criterion (DD1-criterion)”. If there are no such off-equilibrium choices, we say that  $(m^*, a^*)$  survives the DD1-criterion and refer to it as a “DD1-equilibrium”.

## 4 Equilibrium analysis

In this section, we provide a full characterization of equilibria and investigate the possibility of truthful communication in equilibrium. Equilibrium strategies will differ across types but not across agents since we consider symmetric equilibria, and our interest is therefore in the types’ equilibrium behavior. We categorize strategies according to four communication patterns that play an important role in our analysis.

**Definition 3.** Consider any agent  $i = 1, 2$  and strategy profile  $(m, a) \in \mathcal{S}$ .

- (i) **Honesty.** Type  $\theta_i \in \Theta$  is “honest” or “truthful” if she always submits a report that corresponds to her signal,  $m_i(\theta_i, s_i) = s_i$  for all  $s_i \in S$ .
- (ii) **Alarmism.** Type  $\theta_i \in \Theta$  is an “alarmist” if she submits a high report regardless of her signal,  $m_i(\theta_i, s_i) = 1$  for all  $s_i \in S$ .
- (iii) **Denial.** Type  $\theta_i \in \Theta$  is a “denialist” if she submits a low report regardless of her signal,  $m_i(\theta_i, s_i) = 0$  for all  $s_i \in S$ .
- (iv) **Hypocrisy.** Type  $\theta_i \in \Theta$  is a “hypocrite” if she submits a high report and does not contribute for some signal and report received,  $(m_i(\theta_i, s_i), a_i(\theta_i, s_i, \hat{m}_j)) = (1, 0)$  for some  $s_i \in S$  and  $\hat{m}_j \in M$ .

Note that honesty, alarmism and denial are mutually exclusive. By contrast, honesty and hypocrisy as well as alarmism and hypocrisy may occur together, but are conceptually distinct. In particular, while honesty, alarmism and denial only relate to communication, hypocrisy also relates to actions. Furthermore, we introduce the following definitions.

**Definition 4.** Consider any agent  $i = 1, 2$  and strategy profile  $(m, a) \in \mathcal{S}$ .

- (i) **Information transmission.** We say that there is “information transmission” or “truthful communication” if at least one type  $\theta_i \in \Theta$  is truthful.
- (ii) **Consequential communication.** We say that there is “consequential communication” if receiving a high report increases the likelihood of a contribution,  $E(a_i(\cdot, \hat{m}_j) | \hat{m}_j = 1) > E(a_i(\cdot, \hat{m}_j) | \hat{m}_j = 0)$ .

Note that information transmission is necessary for consequential communication. Consequential communication means that communication affects contributions to the public good. Contributions may not only bring a personal benefit to the high type, but may also bring utility from a higher image assigned by the other agent. In particular, contributions serve as costly signals for being a high type, and in case an agent does not contribute, her report could contain additional information about her type. Hence, while both types may have an incentive to turn alarmist or hypocrite to increase contributions by high types, it is primarily the low type who may have an incentive to turn denialist to avoid the loss in image that contributions by high types entail.

Before we turn to the possibility of truthful communication, we first establish the existence of an equilibrium in which no information transmission takes place.

**Proposition 1.** *There exists at least one DD1-equilibrium without information transmission. In particular, there exists such an equilibrium in which the high type*

(i) *never contributes iff*

$$\frac{3c - 4}{3(1 - \pi)} \geq \mu \text{ and } c - \frac{3}{4} > \mu > c - \frac{3}{2}. \quad (2)$$

(ii) *contributes conditional on a high signal iff*

$$c - \frac{3}{4} > \mu \geq \frac{(3c - 4)(3 - \pi)(3 - 2\pi)}{3(1 - \pi)(9 - 4\pi)}. \quad (3)$$

The proof of this and all other results is presented in Appendix A. Notice that we include the bounds from Assumption 1 in our results whenever binding. Proposition 1 (i) shows that a high type will never contribute in absence of information transmission if image concerns are too low relative to the cost of contributions and the common prior. Part (ii) shows that if image concerns are high enough relative to the cost and the common prior, a high type will contribute conditional on a high signal, independent of any communication. In addition, the minimum image concerns necessary to induce a contribution increase in the prior, because a higher prior increases the “outside image” obtained if not contributing. Notice that since  $(3 - \pi)(3 - 2\pi) < 9 - 4\pi$  for all  $\pi$ , both equilibria may exist at the same time. This potential for multiple equilibria reflects the fact that the high type’s contribution behavior changes the perception of different actions by the other agent. Particularly, when the high type does not contribute in equilibrium, this raises the image associated with not contributing and makes contributing relatively less attractive.

Next, we investigate the possibility of truthful communication. We are interested in situations in which, different from the equilibria derived so far, communication matters for contributions to the public good, that is, strategy profiles with consequential commu-

nication.<sup>7</sup> In a first step, we narrow down the set of potential equilibria. We show that all equilibria with consequential communication share a unique communication strategy: the low type is a denialist and the high type is truthful.

**Proposition 2.** *In any DD1-equilibrium with consequential communication, the low type is a denialist and the high type is truthful.*

The proof of Proposition 2 establishes that if there is consequential communication and the low type is *not* a denialist, then at least one of the types has incentives to deviate. The main intuition is that if image concerns are low, the high type has incentives to turn to hypocrisy/alarmism to raise contributions by the other agent. On the other hand, high image concerns introduce a “justification motive”: the low type has incentives to turn to denial to avoid the loss in image that contributions by the other agent entail. Notably, Proposition 2 rules out truth-telling by both types under consequential communication.

In the following, we restrict our attention to equilibria with consequential communication in which the low type is a denialist and the high type is truthful. We refer to such equilibria as *denialist-equilibria*. The contribution strategy of the high type takes one of two forms in these equilibria: either she contributes conditional on a high signal *or* a high report (from the other agent), or she contributes conditional on a high signal *and* a high report (from the other agent).

We first study the case when image concerns are low. With low (enough) image concerns and consequential communication, it is clear that truth-telling by the high type cannot be an equilibrium.<sup>8</sup> The high type has incentives to turn alarmist as this makes the other player more likely to contribute to the public good. Thus, there does not exist an equilibrium with consequential communication.

**Proposition 3.** *If  $\mu < 1/2$ , there does not exist a DD1-equilibrium with consequential communication.*

The proof first relies on Proposition 2 to reduce the set of potential equilibria to denialist-equilibria. Second, we establish that if  $\mu < 1/2$ , high types have incentives to turn alarmist to increase contributions by the other agent. The expected benefit from the increase in contributions exceeds the expected loss in image.<sup>9</sup>

Next, we turn to our main results on high image concerns. Consider the contribution strategy of the high type in which she contributes conditional on a high signal or a high report. With high (enough) image concerns, the incentives for high types to turn alarmist (and hypocrite) vanish. The expected benefit from the increase in contributions by the other agent is smaller than the expected loss in image. Thus, there exists a denialist-equilibrium.

---

<sup>7</sup>Notice that the equilibria in Proposition 1 may also exist if some type would be truthful instead. We do not discuss these equilibria here as they are outcome-equivalent to the equilibria covered in Proposition 1.

<sup>8</sup>Recall that the low type is a denialist by Proposition 2.

<sup>9</sup>We rely on Theorem 1 and Proposition 4 presented hereafter to establish this argument for reasons of brevity.

**Theorem 1.** *There exists a DD1-denialist-equilibrium such that the high type contributes conditional on a high signal or a high report iff*

$$c - \frac{3}{4} > \mu \geq \max \left\{ c - 1, \frac{(3 - \pi)(3 - 2\pi)}{\pi^2 - 12\pi + 15}, \frac{((3 - 2\pi)(c - 1) - 1)(3 - \pi)(3 - 2\pi)}{3(1 - \pi)(2\pi^2 - 10\pi + 9)} \right\}. \quad (4)$$

This result shows that unlike in the low-image case, truthful communication is possible when image concerns are high and high types contribute to the public good. We establish the existence of a denialist-equilibrium in which the high type contributes conditional on a high signal or a high report, that is, if she has got at least some positive information about the public good.

First, notice that the upper bound on image concerns in (4) stems from Assumption 1. Second, the first part of the lower bound on image concerns,  $\mu \geq c - 1$ , ensures that high types who have submitted a low report contribute if the other agent has submitted a high report (and hence expects a contribution from high types). The second part deters hypocrisy by the high type. As hypocrisy is off-equilibrium, the image assigned to this deviation is determined by the DD1-criterion. While low types are better off by choosing not to contribute for any beliefs assigned to this deviation by the other player, this is not the case for high types: if the belief assigned is low enough, she prefers to contribute. Hence, the DD1-criterion assigns the lowest image to this deviation and thereby deters hypocrisy. The last part deters denial by the high type. Denial would allow to avoid a contribution without appearing hypocritical, but also lower contributions by the other agent. The following example illustrates the results we have obtained so far.

**Example 1.** *Let  $\pi = 1/2$ . There exists a DD1-equilibrium in which*

(i) *there is no information transmission and the high type never contributes iff*

$$2c - \frac{8}{3} \geq \mu \text{ and } c - \frac{3}{4} > \mu > c - \frac{3}{2}.$$

(ii) *there is no information transmission and the high type contributes conditional on a high signal iff*

$$c - \frac{3}{4} > \mu \geq \frac{10(3c - 4)}{21}.$$

(iii) *the low type is a denialist and the high type is truthful (denialist-equilibrium) and contributes conditional on a high signal or a high report iff*

$$c - \frac{3}{4} > \mu \geq \max \left\{ c - 1, \frac{20}{37}, \frac{20(2c - 3)}{27} \right\}.$$

Figure 1 illustrates these equilibria. Notice that we employ the  $\mu - (c - \mu)$  scale to ease the exposition, in which moving south-east (north) represents increasing image

concerns (costs). The equilibria without information transmission partially exist on the same range as discussed above. Furthermore, comparing equilibria with contributions, the denialist-equilibrium requires a minimum level of image concerns, but also exists for larger values of image concerns and costs of contributing to the public good.

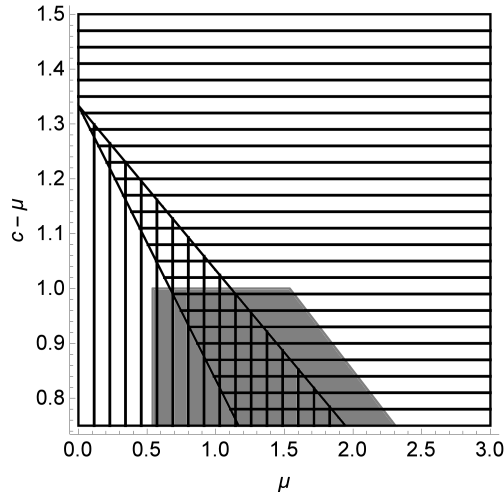


Figure 1: Equilibria for the prior  $\pi = 1/2$ . The equilibrium without information transmission and contributions conditional on the signal (no contributions) is indicated by vertical (horizontal) lines, and the denialist-equilibrium with contributions conditional on a high signal or a high report is indicated in grey.

The insights from Example 1 are qualitatively robust to changes in the prior. However, the parameter ranges of equilibria with contributions essentially decrease in the prior. A higher prior increases the “outside image” that the high type obtains if not contributing and therefore makes not contributing more attractive.

Finally, we consider the contribution strategy of the high type in which she contributes conditional on a high signal and a high report. Notice that under this contribution strategy, hypocrisy only yields a low image if the other agent also has submitted a high report. Moreover, it even yields a high image if the other agent has submitted a low report. This makes truthful communication difficult to sustain in equilibrium. Consequently, it turns out that such a denialist-equilibrium only exists if costs are larger relative to image concerns compared to the former denialist-equilibrium and the share of high types in society is large.

**Proposition 4.** *There exists a DD1-denialist-equilibrium such that the high type contributes conditional on a high signal and a high report iff*

$$B(\pi) \equiv \pi(2\pi^2 - 24\pi + 51) - 27 > 0, \quad (5)$$

$$c - 1 \geq \mu \geq \pi(3 - \pi)(3 - 2\pi) \max \left\{ \frac{1}{B(\pi)}, \frac{2(c - 3)}{3(1 - \pi)(9 - 4\pi)} \right\} \text{ and } \mu > c - \frac{3}{2}. \quad (6)$$

This result establishes the existence of a denialist-equilibrium in which the high type

contributes conditional on a high signal and a high report, that is, if she has got only positive information about the public good.

The upper bound on image concerns in (6) ensures that high types who have turned alarmist have no incentives to contribute if the other agent has submitted a high report. Notice that the other agent expects a contribution in this case and hence not contributing would be off-equilibrium and yield a low image by the DD1-criterion. Compared to the other denialist-equilibrium, this requires higher costs relative to image concerns. The first part of the lower bound on image concerns,  $\mu \geq \pi(3 - \pi)(3 - 2\pi)/B(\pi)$ , and (5) deter the high type from turning alarmist in the first place. In particular, notice that a rather high share of high types in the population—more than 4/5—is necessary for (5) to hold. Otherwise, turning alarmist and not contributing becomes beneficial for high types. The reason is that this yields a high image if the other agent submits a low report, which is likely if there are many low types and her signal is low. The second part deters denial by the high type. Denial would allow to avoid a contribution without appearing hypocritical if the other agent submits a high report, but also lower contributions by the other agent. The last inequality in (6) stems from Assumption 1. Notice that this equilibrium does not exist under the intermediate prior considered in Example 1.

## 5 Welfare

Next, we analyze welfare and derive some comparative statics. Given any strategy profile  $(m, a)$ , *ex-ante expected utilitarian welfare* (henceforth *welfare*) is given by

$$2E[E_i[u_i(\theta_i, s_i, m, a) \mid \theta_i, s_i, \tilde{\theta}_i]],$$

where  $\tilde{\theta}_i$  denotes any belief system consistent with  $(m, a)$ . Similar to Bénabou and Tirole (2006), the welfare ranking is not affected if we only consider costs and benefits of the public good, that is, disregard image.

**Remark 1.** *The ex-ante expected total image in society is independent of the agents' strategies and given by  $2\pi$ .*

We refer to Lemma 1 in Appendix A.6 for details. Furthermore, let

$$\mathcal{S}^* = \left\{ (m, a) \in \mathcal{S} \mid \begin{array}{l} m_i(0, s) = m_i(1, s) \text{ for all } s \in S \text{ and } a_i(0, \cdot) \equiv 0 \text{ for } i = 1, 2, \text{ or} \\ m_i(1, s) = s \text{ and } m_i(0, s) = 0 \text{ for all } s \in S, a_i(0, \cdot) \equiv 0, \\ a_i(1, 0, 0) = 0 \text{ and } a_i(1, 1, 1) = 1 \text{ for } i = 1, 2 \end{array} \right\}$$

denote the restriction to strategy profiles in which both types employ the same communication strategy and the low type does not contribute, as well as the potential denialist-equilibria. This set contains all strategy profiles that we have identified as



potential equilibria in Section 4 as well as all strategy profiles with truthful communication.<sup>10</sup> Truthful communication serves as a benchmark and may be interpreted as a situation where information is public and contributions are made or enforced by a social planner.

In the following, we restrict our attention to the strategy profiles in  $\mathcal{S}^*$  and refer to the strategy profile which yields the highest welfare among the strategy profiles in  $\mathcal{S}^*$  as *welfare-maximizing*. Comparing welfare yields the following result.

**Proposition 5.** *The welfare-maximizing strategy profile is such that*

- (i) *if  $c \leq 3/2$ , both types are truthful and the high type contributes conditional on a high signal or a high report.*
- (ii) *if  $3/2 \leq c \leq (3 + 5\pi)/(2 + 2\pi)$ , the low type is a denialist and the high type is truthful and contributes conditional on a high signal or a high report.*
- (iii) *if  $(3 + 5\pi)/(2 + 2\pi) \leq c \leq 9/4$ , both types are truthful and the high type contributes conditional on a high signal and a high report.*
- (iv) *if  $9/4 \leq c \leq 3$ , the low type is a denialist and the high type is truthful and contributes conditional on a high signal and a high report.*
- (v) *if  $c \geq 3$ , there is no information transmission (both types are truthful) and the high type never contributes.*

Notice that we omit the bounds from Assumption 1 here, as Proposition 5 also holds outside these bounds. This result shows that if costs of contributing to the public good are low, then full honesty and contributions by high types conditional on positive information about the public good would be optimal for society. Hence, denial by low types results in inefficiently low contributions in this case.

Interestingly, however, denial by low types may also increase welfare relative to honesty if costs of contributing to the public good are intermediate. The high type contributes partially because of the gain in image that comes with a contribution, at the expense of the other agent's image. Denial may therefore increase welfare by mitigating the adverse consequences of image concerns, namely over-contributions by high types when the other agent is a low type. This is not merely an academic possibility, and we discuss applications related to climate change and energy conservation in Section 6. Note that contributing solely conditional on a high signal, in absence of any information transmission, never yields the highest welfare, i.e., some information transmission always improves the outcome given high types contribute under some circumstances.

Since Proposition 5 does not restrict image concerns, actual equilibria may overlap with these optimal regions. The following example illustrates our results and relates

---

<sup>10</sup>Notice that the set contains in total six potential equilibrium strategy profiles and three strategy profiles with truthful communication. We refer to the proof of Proposition 5 for details.

them to the denialist-equilibrium with contributions conditional on a high signal or a high report (Theorem 1).

**Example 2.** Let  $\pi = 1/2$ . Figure 2 illustrates the welfare-maximizing strategy profiles derived in Proposition 5. Recall that moving south-east (north) in the  $\mu - (c - \mu)$  scale represents increasing image concerns (costs). The strategy profile in which the low type is a denialist and the high type is truthful and contributes conditional on a high signal or a high report is welfare-maximizing iff  $3/2 \leq c \leq 11/6$ . Combined with the conditions derived in Example 1, this strategy profile is a DD1-equilibrium and welfare-maximizing iff

$$c - \frac{3}{4} > \mu \geq \max \left\{ c - 1, \frac{20}{37} \right\} \text{ and } \frac{11}{6} \geq c \geq \frac{3}{2}. \quad (7)$$

Given sufficiently high image concerns, the denialist-equilibrium with contributions conditional on a high signal or a high report is welfare-maximizing if costs of contributing to the public good are intermediate. Otherwise, efficiency would require truthful communication (except for rather high costs). In particular, if the costs of contributing to the public good and image concerns are rather low, the denialist-equilibrium features inefficiently low contributions, see Figure 2.

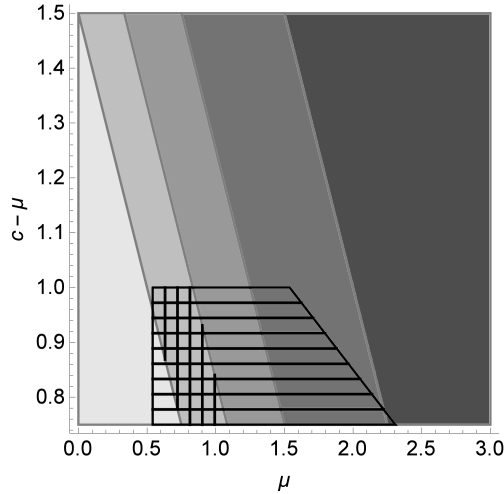


Figure 2: Welfare-maximizing strategy profiles for the prior  $\pi = 1/2$ . The strategy profiles from Proposition 5 (i)–(v) are indicated in grey (from light (i) to dark (v)). The denialist-equilibrium with contributions conditional on a high signal or a high report is indicated by horizontal lines. The parameter range on which this equilibrium is welfare-maximizing (7) is indicated by vertical lines.

Next, we derive comparative statics on how image concerns affect welfare in equilibrium. As welfare of a particular equilibrium is increasing in image concerns, we only consider costs and benefits of the public good. It follows from Remark 1 that *net-of-image-welfare* of a strategy profile  $(m, a)$  is given by  $2E[E_i[u_i(\theta_i, s_i, m, a) \mid \theta_i, s_i, \tilde{\theta}_i]] - 2\pi\mu$ ,

where  $\tilde{\theta}_i$  denotes any belief system consistent with  $(m, a)$ . Combining our results on equilibria and welfare, we show that when costs of contributing to the public good are intermediate, sufficiently high image concerns increase net-of-image-welfare in equilibrium.

**Corollary 1.** *Suppose that  $27/20 < c < 39/20$  and that agents coordinate on the welfare-maximizing equilibrium. If  $\mu < 1/2$ , then there exists  $\mu' > \mu$  such that  $\mu'$  yields higher net-of-image-welfare in equilibrium than  $\mu$ .*

Corollary 1 establishes that if costs of contributing to the public good are intermediate, then image concerns may increase net-of-image-welfare in equilibrium through a change in the set of equilibria. Using Proposition 5, we first show that in the considered cost-range, the denialist-equilibrium with contributions by the high type conditional on a high signal or a high report yields the highest welfare among the equilibrium candidates. Next, recall from Proposition 3 that this equilibrium does not exist if image concerns are low. We establish the claim by showing that we can choose high enough image concerns for which the equilibrium does exist and hence net-of-image-welfare in equilibrium is higher. This shows that image concerns may increase the efficiency of public goods provision, *even though* it results in denialist communication.

## 6 Evidence

Most collective action problems feature uncertainty about the return on investment. Many important examples concern institutional and behavioral change. In this section, we highlight three applications of our model to institutional inertia, climate change denial, and energy conservation. The model predicts that communication about the necessity of change will be contentious, and that denial will arise naturally even if everybody would benefit from change.

The (mostly qualitative) evidence on these topics is highly suggestive that the mechanisms in our model are at work. Nevertheless, they don't fully identify these mechanisms, as it is difficult to identify private information and preferences in field studies, and control for all aspects of communication. Generating more direct evidence may necessitate additional controls, for instance in the laboratory.

### 6.1 Institutional inertia

It is known since the days of Machiavelli that organizations are difficult to reform, even if changes are potentially beneficial for everyone. This phenomenon is referred to as *institutional inertia*, the tendency of (large) organizations to respond slowly to changes in the environment. These include technology, laws and regulation, public opinion as well as available information. A well-known example is General Motors, which has once been the world leader in automobile production efficiency, but lost this position

to Japanese manufacturers during the 1970s. Although senior management understood the nature of the problem by the end of the decade, they were not able to close the productivity gap (Rumelt, 1995).

Many studies conclude that change initiatives in organizations inevitably generate resistance (Agocs, 1997; Armenakis and Bedeian, 1999). In particular, change requires people to “[take] risks, particularly those associated with self esteem—loss of face, appearing incompetent, seemingly unable or unwilling to learn, etc.” (Jick, 2008, p.407). Thus, change may require costly actions that threaten the image of those who wish to keep the reputation of being a good employee.

Such pressure may generate denial about the benefits of change. Agocs (1997) mainly discusses two forms of denial: denial of the legitimacy of the case for change, and refusal to recognize responsibility to address the change issue. He justifies the focus on these early stages of the change process as “reflecting the fact that change agents most frequently deal with denial, and many proposals never get beyond that point” (p.920). In this context, denial is understood as downplaying the credibility of the change message, or the credibility of the messenger, or both.

This parallels the mechanisms in our model: downplaying the credibility of the change message can be interpreted as denial in our terminology, and refusal to recognize responsibility to address the change issue can be interpreted as inaction or not contributing to the public good. Combined with the observation that people wish to preserve a “good image” outlined above, this describes the behavior of low types in our model. Hence, our stylized model is consistent with key findings on institutional inertia.

## 6.2 Climate change denial

One of the most prominent and important public good debates in the beginning of the 21st century concerns climate change. While there is little scientific doubt about the occurrence of climate change and the role of human produced greenhouse gases like CO<sub>2</sub>, there is still substantial uncertainty about the future costs of climate change, and the effectiveness and cost of abatement technologies.<sup>11</sup> Many studies document denial about the importance or impact of climate change among substantial minorities in Western countries (Hobson and Niemeyer, 2013). Only about half of Americans believe that human activity is responsible for warming the planet, a fraction that has remained rather stable over time (Pew Research Center, 2016).

Evidence from interviews and interactive focus groups show that the motives behind such denial correspond to those in the model and involve the cost of behavioral

---

<sup>11</sup>Since 2007, the Intergovernmental Panel on Climate Change (IPCC) reports that the scientific evidence for climate change is “unequivocal”, and says men-made activity has resulted in warming with “very high confidence” (IPCC, 2007). At the same time, modeling the appropriate response to climate change has yielded mostly speculative results. Pindyck (2013) reviews “integrated assessment models” of the relation between economic policy and the climate, and concludes that such analyses “create a perception of knowledge and precision, but that perception is illusory and misleading” (p.860).

change and the wish to preserve the image of being a good citizen.<sup>12</sup> Particularly, Stoll-Kleemann, O’Riordan, and Jaeger (2001) organize focus groups in which randomly selected individuals in Switzerland discussed possible consequences of climate change. They find that even though people expressed awareness and anxiety about the occurrence of climate change, they frequently downplay the problem by emphasizing uncertainty around its impact, questioning the possibility for meaningful individual actions, or by making optimistic projections about the impact of new technologies. Stoll-Kleemann, O’Riordan, and Jaeger (2001) highlight the importance of image concerns. They argue that denial helps to deflect personal responsibility, and serves to “justify why they should not act either individually or through collective institutions” (p.107).

Similarly, Norgaard (2006a,b) finds denial in her interviews with Norwegian subjects. She cites the “fear of being a bad person” as one of the underlying motivations of this denial. In order to deflect personal responsibility, people downplay the knowledge they have about the effects of climate change:

“The notion that people are not acting against global warming because they do not know reinforces a sense of their innocence in the face of these activities.” (2006b, p.366)

Stoll-Kleemann, O’Riordan, and Jaeger (2001) also link these denials explicitly to the cost of contributions:

“The most powerful zone for denial was the perceived unwillingness to abandon what appeared as personal comfort and lifestyle-selected consumption and behaviour in the name of climate change mitigation.” (2001, p.113)

These findings are thus in line with the idea that denial is motivated by the wish to justify inaction and affects communication about climate change, which is the central result of our model.

### **6.3 Conspicuous conservation**

The literature on energy conservation, and in particular “conspicuous conservation” (Griskevicius, Tybur, and Van den Bergh, 2010), also provides evidence for the mechanisms and welfare implications of our model. A key role for image concerns in our model is that they deter hypocrisy. To avoid discrepancy between their actions and their words, image concerns lead people either to contribute (high types) or engage in denial (low types). Sexton and Sexton (2014) show that signaling “green” motivations indeed leads to an increased willingness to pay of up to \$4000 for hybrid or electric cars.

---

<sup>12</sup>Denial may be aided by the efforts of the fossil fuel industry to supply misinformation (Oreskes and Conway, 2010, 2011). These actors are not included in our model, as we focus on agents who benefit from the public good when others provide it, in this case cleaner air and less fossil fuel consumption. We show that the demand for denial does not depend on the presence of actors with such clear interests in the status quo.

Hards (2013) interviews people about their energy conservation practices, and finds that “eco-bling” can be status enhancing. Particularly,

“[b]eing seen to perform certain energy practices [...] can *protect people from being seen as hypocritical*, especially if they feel part of an ethical-environmental community. For example, Paul (56 years old) said that his wood-burning stove gave him a sense of “satisfaction”, “self-righteousness” or “smugness” because it showed he was not contributing to the environmental problems about which he campaigns.” (2013, p.443, emphasis ours)

Conspicuous conservation also illustrates some of the counterintuitive welfare implications of the model. In particular, when image concerns and the cost of contributions are relatively high, the model shows that denial may improve welfare by reducing over-contributions by high types. Conspicuous conservation may indeed lead to excessive investment in visible conservation measures, and crowd out less “glamorous” but more efficient efforts like weatherizing homes (Wilk and Wilhite, 1985).

## 7 Extension: impure public goods and charitable giving

Our model can easily be extended to impure public goods and charitable giving. To allow for an impure public good, the preferences of agent  $i$  are now given as:

$$u_i(\theta_i, s_i, \hat{m}, \hat{a}) \equiv (1 + \theta_i)(\hat{a}_i + \beta \hat{a}_j)W - \hat{a}_i c + \mu E_i [E_j[\theta_i | \theta_j, s_j, \hat{m}_i, \hat{a}_i] | \theta_i, s_i, \hat{m}_j, \hat{a}_j].$$

The novel parameter  $\beta \in [0, 1]$  allows for imperfect spillovers from contributions by the other agent. The public good is called *pure* if  $\beta = 1$  and *impure* otherwise. Additionally, we allow the high type’s degree of intrinsic motivation to contribute to the public good to take values  $\bar{\theta} \in (0, 1]$ , such that  $\theta_i \in \Theta = \{0, \bar{\theta}\}$ . For instance, if we interpret  $\bar{\theta}$  as the degree to which high types internalize the effect of a contribution on the other agent, then  $\bar{\theta} \leq \beta$  seems reasonable. On the other hand, values  $\bar{\theta} > \beta$  may be interpreted as “warm glow” from the effect of a contribution on a third party (Andreoni, 1989). We discuss this possibility in detail in Section 7.1 in the context of charitable giving. Notice that Assumption 1 now reads  $3/4 < c - \mu < 3(1 + \bar{\theta})/4$ . We recover our baseline model if  $\beta = 1$  and  $\bar{\theta} = 1$ .

We restrict our attention to our main result on the denialist-equilibrium with contributions by the high type conditional on a high signal or a high report (Theorem 1).

**Proposition 6.** *There exists a DD1-denialist-equilibrium such that the high type con-*

tributes conditional on a high signal or a high report iff

$$c - \frac{3}{4} > \mu \geq \max \left\{ c - (1 + \bar{\theta})/2, \frac{(1 + \bar{\theta})\beta(3 - \pi)(3 - 2\pi)}{2(\pi^2 - 12\pi + 15)}, \frac{\beta(3 - \pi)(3 - 2\pi)}{4(\pi^2 - 6\pi + 6)}, \right. \\ \left. \frac{(2(3 - 2\pi)c - (1 + \bar{\theta})(4 - \pi(3 - \beta)))(3 - \pi)(3 - 2\pi)}{6(1 - \pi)(2\pi^2 - 10\pi + 9)} \right\}. \quad (8)$$

To avoid repetition, we mainly comment on the effect of the high types' degree of intrinsic motivation  $\bar{\theta}$  and the level of spillovers  $\beta$ . Recall that the upper bound on image concerns in (8) stems from Assumption 1. The first part of the lower bound on image concerns,  $\mu \geq c - (1 + \bar{\theta})/2$ , ensures that high types contribute if the other agent has submitted a high report. This bound is decreasing in  $\bar{\theta}$ , which implies that contributions become more attractive for high types if  $\bar{\theta}$  increases. In particular, notice that  $\bar{\theta} > 1/2$  is necessary for the equilibrium to exist. The second part deters hypocrisy by the high type. This bound is increasing in  $\bar{\theta}$  and  $\beta$ , which implies that truthful communication becomes more difficult to sustain if  $\bar{\theta}$  or  $\beta$  increases. The third part is new compared to Theorem 1 and deters hypocrisy by the low type. This bound is increasing in  $\beta$ , which implies that denial becomes more difficult to sustain if  $\beta$  increases. This bound does not appear in (4) as it is implied by the second lower bound if  $\bar{\theta} = 1$ . If the high type's degree of intrinsic motivation is large enough, then she is more likely than the low type to turn hypocrite. The last part deters denial by the high type. This bound is decreasing in  $\bar{\theta}$  and  $\beta$ , which implies that truthful communication becomes easier to sustain if  $\bar{\theta}$  or  $\beta$  increases.

In summary, the case for hypocrisy is weaker, while the case for denial is stronger with an impure public good. The intuition is straightforward: if spillovers are low one loses (gains) less from discouraging (encouraging) contributions by others.

## 7.1 Charitable giving

Our extended framework easily accommodates the case of charitable giving to a third entity. In this case there are essentially no spillovers on the communication partner, i.e.,  $\beta = 0$ . Some types may be intrinsically motivated to give, implying that  $\bar{\theta} > \beta$ . This may be interpreted as the high type experiencing "warm glow" from the effect of her own contribution on a third party like a charitable organization.

For this case, we consider the denialist-equilibrium with contributions conditional on a high signal or a high report. First, observe that the second and the third part of the lower bound on image concerns in (8) vanish. Without spillovers from contributions by the other agent, there is no reason to induce more contributions through hypocrisy/alarmism for any of the types. Interestingly, this implies that the equilibrium already exists for very low image concerns, i.e., the presence of spillovers is crucial to our result on low image concerns (Proposition 3). Second, the last part of the lower

bound is decreasing in  $\beta$ , so denial is more attractive for high types than in the baseline model. Without spillovers, denial does not reduce expected benefits from contributions by the other agent any more. This raises the attractiveness of denial for the high type, and implies that the equilibrium no longer exists for intermediate image concerns and costs of contributing. Figure 3 illustrates the equilibrium in the baseline model,  $\beta = 1$ , and without spillovers,  $\beta = 0$ .

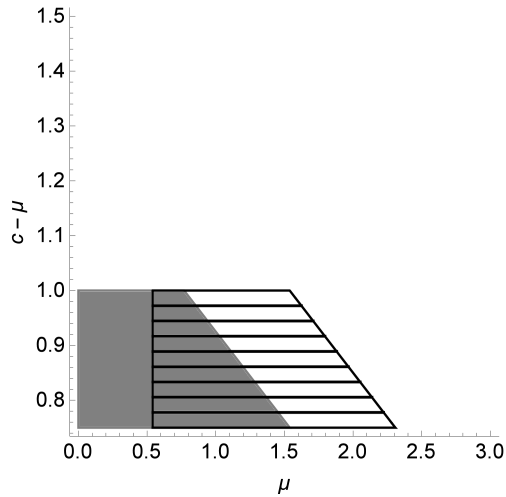


Figure 3: Denialist-equilibrium with contributions conditional on a high signal or a high report for the prior  $\pi = 1/2$  and  $\bar{\theta} = 1$ . The high-spillover case,  $\beta = 1$ , is indicated by horizontal lines, and the no-spillover case,  $\beta = 0$ , is indicated in grey.

In the empirical literature, there is ample evidence that people look for social excuses not to give, providing support for the “justification motive”. For instance, people take the other exit at the supermarket to avoid solicitors for charity (Andreoni, Rao, and Trachtman, 2017), and are less often at home when a solicitation is announced in advance (Dellavigna, List, and Malmendier, 2012). In laboratory studies, subjects choose to opt out of giving situations at a personal cost to themselves (Dana, Cain, and Dawes, 2006), and give less if they can credibly deflect blame for their selfish actions (Dana, Weber, and Kuang, 2007; Andreoni and Bernheim, 2009).

The efficiency or impact of giving plays an important role in giving, and also in excuses not to give. If given the choice, a substantial proportion of subjects in the laboratory chooses not to know information about the impact, even if they would use the information when they are forced to observe it (Dana, Weber, and Kuang, 2007; Grossman and van der Weele, 2017). Exley (2016a) shows that when the impact of a charitable donation is risky, subjects weight this risk more heavily when the cost of the donation is theirs, rather than a third party’s. Exley (2016b) shows that when presented with information about the effectiveness of a charity, subjects will weight low effectiveness strongly as a reason not to give, but again, only when they personally bear the cost of giving.

There is some suggestive evidence about the impact of such excuse-seeking for com-



munication about effectiveness. Andreoni and Sanchez (2014) show that subjects in the lab falsify belief statements to the experimenter about the trustworthiness of others, in order to excuse their own selfish behavior.

## 8 Discussion and conclusion

We investigate communication about the return to public goods in the presence of social preferences and image concerns. If intrinsic motivation to contribute to the public good is sufficiently high, communication could in principle be very helpful for generating contributions. However, our results show that such social preferences generate incentives for “alarmism” and “hypocrisy”: opportunistic exaggeration of the importance of the public good, in order to induce cooperation. Image concerns deter such hypocrisy and make partially truthful communication possible. However, they also generate a “justification motive” that leads to the opposite communication bias: people who are not inclined to contribute to the public good engage in “denial” in order to justify their selfish actions.

These results reveal the inherent difficulties of information aggregation about public goods, *even if* all agents benefit from the public good. Furthermore, if contributions are cheap, denial leads to inefficiently low public goods provision through the suppression of alarming signals. In the context of climate change, Stoll-Kleemann, O’Riordan, and Jaeger (2001) conclude that there is

“both a coherence and a rationality to dissonance and denial that will not make it easy for democracies to gain early consent for tough climate change mitigation measures.” (2001, p.115).

Interestingly, denial may be better than full transparency in occasions where status concerns lead to over-contribution. Section 6 showed examples in the context of energy consumption, where “conspicuous conservation” and “eco-bling” may crowd out more efficient contributions. In general, since partial denial is a feature of any equilibrium with meaningful communication in our model, image concerns may increase the efficiency of public goods provision by enabling such equilibrium.

Our model suggests that there are no straightforward policy solutions to biased communication. Taking away incentives for misrepresentation requires incentives to make public good investments individually rational, for instance through Pigouvian taxation. However, this raises a chicken-and-egg problem, as it assumes the existence of both the information and the political will to implement such policies, exactly the elements that are missing in the context we have considered. The model does suggest that since aggregation of private signals is likely to be flawed, the government may consider it its task to spearhead the production and dissemination of information about the social returns to public goods like climate change mitigation.

Future research should delve further into the role of public science communication. Reducing denial may be achieved by changing the moral frames surrounding public good contributions, using what Kahan (2015) calls “disentanglement strategies”. Another avenue is to extend our framework to larger groups of people. On the empirical side, experimental investigations of strategic misrepresentation of information in public good games and its underlying motives would deepen our understanding of the nature of denial and alarmism.

## References

- AGOCs, C. (1997): “Institutionalized resistance to organizational change: Denial, inaction and repression,” *Journal of Business Ethics*, 16(9), 917–931.
- ALI, S. N., AND R. BÉNABOU (2016): “Image Versus Information,” *NBER Working Paper*, 22203.
- ANDREONI, J. (1989): “Giving with impure altruism: Applications to charity and Ricardian equivalence,” *Journal of Political Economy*, 97(6), 1447–1458.
- ANDREONI, J., AND D. B. BERNHEIM (2009): “Social Image and the 50-50 Norm: A Theoretical and Experimental Analysis of Audience Effects,” *Econometrica*, 77(5), 1607–1636.
- ANDREONI, J., AND R. PETRIE (2004): “Public goods experiments without confidentiality: a glimpse into fund-raising,” *Journal of Public Economics*, 88, 1605–1623.
- ANDREONI, J., J. M. RAO, AND H. TRACHTMAN (2017): “Avoiding the Ask: A Field Experiment on Altruism, Empathy, and Charitable Giving,” *Journal of Political Economy*, 125(3), 625–653.
- ANDREONI, J., AND A. SANCHEZ (2014): “Do Beliefs Justify Actions or Do Actions Justify Beliefs? An Experiment on Stated Beliefs, Revealed Beliefs, and Social-Image Manipulation,” *National Bureau of Economic Research Working Paper Series*, No. 20649(November), 1–38.
- ARIELY, D., A. BRACHA, AND S. MEIER (2009): “Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially,” *American Economic Review*, 99(1), 544–555.
- ARMENAKIS, A. A., AND A. G. BEDEIAN (1999): “Organizational change: A review of theory and research in the 1990s,” *Journal of management*, 25(3), 293–315.
- BANKS, J. S., AND J. SOBEL (1987): “Equilibrium Selection in Signaling Games,” *Econometrica*, 55(3), 647–661.
- BÉNABOU, R., A. FALK, AND J. TIROLE (2018): “Narratives, Imperatives and Moral Reasoning,” Unpublished manuscript.
- BÉNABOU, R., AND J. TIROLE (2006): “Incentives and prosocial behavior,” *American Economic Review*, 96(5), 1652–1678.
- (2011): “Laws and Norms,” *NBER Working Paper No. 17579*.

- BURLANDO, R. M., AND F. GUALA (2004): “Heterogeneous Agents in Public Goods Experiments,” *Experimental Economics*, 8, 35–54.
- CHO, I.-K., AND D. M. KREPS (1987): “Signaling games and stable equilibria,” *The Quarterly Journal of Economics*, 102(2), 179–221.
- CRAWFORD, V. P., AND J. SOBEL (1982): “Strategic Information Transmission,” *Econometrica*, 50(6), 1431–1451.
- DANA, J., D. M. CAIN, AND R. M. DAWES (2006): “What you don’t know wont hurt me: Costly (but quiet) exit in dictator games,” *Organizational Behavior and Human Decision Processes*, 100(2), 193–201.
- DANA, J., R. A. WEBER, AND J. X. KUANG (2007): “Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness,” *Economic Theory*, 33(1), 67–80.
- DELLAVIGNA, S., J. A. LIST, AND U. MALMENDIER (2012): “Testing for altruism and social pressure in charitable giving,” *Quarterly Journal of Economics*, 127(1), 1–56.
- ELLINGSEN, T., AND M. JOHANNESSON (2008): “Pride and prejudice: The human side of incentive theory,” *The American Economic Review*, 98(3), 990–1008.
- ELLINGSEN, T., AND R. ÖSTLING (2010): “When Does Communication Improve Coordination?,” *American Economic Review*, 100(September), 1695–1724.
- EXLEY, C. L. (2016a): “Excusing selfishness in charitable giving: The role of risk,” *Review of Economic Studies*, 83(2), 587–628.
- (2016b): “Using charity performance metrics as an excuse not to give,” *Mimeo, Harvard University*.
- FISCHBACHER, U., S. GÄCHTER, AND E. FEHR (2001): “Are People Conditionally Cooperative,” *Economics Letters*, 71(3), 397–404.
- GALEOTTI, A., C. GHIGLINO, AND F. SQUINTANI (2013): “Strategic information transmission networks,” *Journal of Economic Theory*, 148(5), 1751–1769.
- GEANAKOPOLOS, J., D. PEARCE, AND E. STACCHETTI (1989): “Psychological games and sequential rationality,” *Games and Economic Behavior*, 1(1), 60–79.
- GRISKEVICIUS, V., J. M. TYBUR, AND B. VAN DEN BERGH (2010): “Going Green to Be Seen: Status, Reputation, and Conspicuous Conservation,” *Journal of Personality and Social Psychology*, 98(3), 392–404.
- GROSSMAN, Z., AND J. VAN DER WEELE (2017): “Self-Image and Willful Ignorance in Social Decisions,” *Journal of the European Economic Association*, 15(1), 173–217.
- HAGENBACH, J., AND F. KOESSLER (2010): “Strategic Communication Networks,” *Review of Economic Studies*, 77, 1072–1099.
- HAGENBACH, J., F. KOESSLER, AND E. PEREZ-RICHET (2014): “Certifiable Pre-Play Communication: Full Disclosure,” *Econometrica*, 82(3), 1–46.

- HARBAUGH, W. T. (1998): “What Do Donations Buy? A Model of Philanthropy Based on Prestige and Warm Glow,” *Journal of Public Economics*, 67, 269–284.
- HARDS, S. K. (2013): “Status, stigma and energy practices in the home,” *Local Environment*, 18(4), 438–454.
- HENRY, E., AND C. LOUIS-SIDOIS (2015): “Voting and contributing when the group is watching,” *Mimeo, Sciences Po*.
- HOBSON, K., AND S. NIEMEYER (2013): ““What sceptics believe”: The effects of information and deliberation on climate change scepticism.,” *Public Understanding of Science*, 22(4), 396–412.
- IPCC (2007): “Climate Change 2007 Synthesis Report,” Discussion paper.
- JICK, T. D. (2008): “The recipients of change,” in *Organization change: A comprehensive reader*, ed. by W. W. Burke, D. G. Lake, and J. W. Paine, vol. 155. John Wiley & Sons.
- KAHAN, D. M. (2015): “What is the “science of science communication”?,” *JCOM: Journal of Science Communication*, 14(3), 1–12.
- KARLAN, D., AND M. A. MCCONNELL (2014): “Hey look at me: The effect of giving circles on giving,” *Journal of Economic Behavior and Organization*, 106, 402–412.
- KURAN, T. (1997): *Private truths, public lies: The social consequences of preference falsification*. Harvard University Press, Cambridge, MA.
- KURZBAN, R., AND D. HOUSER (2005): “Experiments investigating cooperative types in humans: A complement to evolutionary theory and simulations,” *Proceedings of the National Academy of Sciences*, 102(5), 1803–1807.
- LACETERA, N., AND M. MACIS (2010): “Social image concerns and prosocial behavior: Field evidence from a nonlinear incentive scheme,” *Journal of Economic Behavior and Organization*, 76(2), 225–237.
- NORGAARD, K. M. (2006a): “People to Protect Themselves a Little Bit: Emotions, Denial, and Social Movement Nonparticipation,” *Sociological Inquiry*, 76(3), 372–396.
- (2006b): “We Don’t Really Want to Know. Environmental Justice and Socially Organized Denial of Global Warming in Norway,” *Organization and Environment*, 19(3), 347–370.
- ORESKE, N., AND E. M. CONWAY (2010): “Defeating the merchants of doubt,” *Nature*, 465(7299), 686–687.
- (2011): *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming*. Bloomsbury Press, London.
- PEW RESEARCH CENTER (2016): “The Politics of Climate,” .
- PINDYCK, R. S. (2013): “Climate Change Policy: What Do the Models Tell Us?,” *Journal of Economic Literature*, 51(3), 1–23.

- REGE, M., AND K. TELLE (2004): “The impact of social approval and framing on cooperation in public good situations,” *Journal of Public Economics*, 88, 1625–1644.
- RUMELT, R. P. (1995): “Inertia and Transformation,” in *Resources in an evolutionary perspective: A synthesis of evolutionary and resource-based approaches to strategy*, ed. by C. A. Montgomery. Kluwer Academic Publishers, Norwell, MA.
- SALLY, D. (1995): “Conversation and Cooperation in Social Dilemmas: A Meta-Analysis of Experiments from 1958 to 1992,” *Rationality and Society*, 7, 58–92.
- SERRA-GARCIA, M., E. V. DAMME, AND J. POTTERS (2011): “Hiding an inconvenient truth: Lies and vagueness,” *Games and Economic Behavior*, 73(1), 244–261.
- SEXTON, S. E., AND A. L. SEXTON (2014): “Conspicuous conservation: The Prius halo and willingness to pay for environmental bona fides,” *Journal of Environmental Economics and Management*, 67(3), 303–317.
- SOETEVENT, A. R. (2005): “Anonymity in giving in a natural context a field experiment in 30 churches,” *Journal of Public Economics*, 89, 2301–2323.
- SPENCE, M. (1973): “Job Market Signaling,” *The Quarterly Journal of Economics*, 87(3), 355.
- STOLL-KLEEMANN, S., T. O’RIORDAN, AND C. C. JAEGER (2001): “The psychology of denial concerning climate mitigation measures: evidence from Swiss focus groups,” *Global Environmental Change*, 11(2), 107–117.
- WILK, R. R., AND H. L. WILHITE (1985): “Why Don’t People Weatherize Their Homes? An Ethnographic Solution,” *Energy*, 10(5), 621–629.

## A Appendix: proofs

### A.1 Proof of Proposition 1

Suppose without loss of generality that both types always submit  $\hat{m}_i = 1$ . Hence, reports are uninformative and the high type’s contribution decision is based only on her private signal.

- (i) Suppose that high types never contribute and let  $\tilde{\theta}_i$  denote any consistent belief system. Consider the second stage and suppose that  $\hat{m}_i = 1$ . Notice that not contributing yields an image of  $\pi$ , while contributing is off-equilibrium. Denote the image in this case by  $\hat{\theta}_i$  and let  $\tilde{\theta}_i$  denote any consistent belief system such that  $\tilde{\theta}_i(\theta_j, s_j, 1, 1) \equiv \hat{\theta}_i$ . The high type with  $s_i = 1$  has no incentives to deviate to  $\hat{a}_i = 1$  iff

$$\begin{aligned}
 U_i^*(1, \hat{a}_i = 0 | \theta_i, s_i, \tilde{\theta}_i) &\geq U_i^*(1, \hat{a}_i = 1 | \theta_i, s_i, \tilde{\theta}_i) \Leftrightarrow \mu\pi \geq 2E[W|s_i] - c + \mu\hat{\theta}_i \\
 &\Leftrightarrow \hat{\theta}_i \leq \pi + \frac{c - 4/3}{\mu} \equiv \theta^*(1, 1).
 \end{aligned}$$

As low types never contribute by Assumption 1, a deviation to  $\hat{a}_i = 1$  is assigned to the high type by the DD1-criterion (and hence yields a high image) whenever it is weakly

beneficial for some belief  $\hat{\theta}_i$  of the other agent. Hence, the high type with  $s_i = 1$  has no incentives to deviate to  $\hat{a}_i = 1$  iff

$$\theta^*(1, 1) \geq 1 \Leftrightarrow \mu \leq \frac{3c - 4}{3(1 - \pi)}, \quad (9)$$

which is the desired upper bound on  $\mu$ .<sup>13</sup> Furthermore, (9) implies that also the high type with  $s_i = 0$  has no incentives to deviate to  $\hat{a}_i = 1$  as her expected benefit from contributing is lower.

If  $\hat{m}_i = 0$ , then any action is off-equilibrium. A high type with  $s_i = 1$  prefers to contribute when this yields a high image and not contributing a low image if

$$2E[W|s_i] - c + \mu \geq 0 \Leftrightarrow \mu \geq \frac{3c - 4}{3}. \quad (10)$$

We proceed by case distinction. First, suppose that (10) holds. By Assumption 1, low types prefer not to contribute even if contributing would yield a high image and not contributing a low image. Hence, applying the DD1-criterion yields, for any  $\hat{m}_j$ ,  $D^*(0, s_i, 0, 1, \hat{m}_j) = \emptyset$  for all  $s_i$  and  $D^*(1, 1, 0, 1, \hat{m}_j) \neq \emptyset$ . Therefore,  $\Theta^*(0, 1, \hat{m}_j) = \{1\}$ , i.e., the DD1-criterion assigns a high image to action  $\hat{a}_i = 1$  when  $\hat{m}_i = 0$ . Similarly,  $D^*(0, s_i, 0, 0, \hat{m}_j) = [0, 1]$  for all  $s_i$ ,  $D^*(1, 1, 0, 1, \hat{m}_j) \subsetneq [0, 1]$  and therefore  $\Theta^*(0, 0, \hat{m}_j) = \{0\}$ , i.e., the DD1-criterion assigns a low image to action  $\hat{a}_i = 0$  when  $\hat{m}_i = 0$ . Consider the first stage. There are two possible deviations, either submitting a low report and contributing or submitting a low report and not contributing. The former deviation yields a high image and the latter a low image. Hence, a deviation to  $(0, 0)$  is dominated by  $(1, 0)$  as it yields a lower image, while a deviation to  $(0, 1)$  yields the same payoff as  $(1, 1)$  and is therefore also not profitable. Second, suppose that (10) does not hold. Then, contributing is strictly worse than not contributing for the high type with any signal and for any beliefs assigned by the other player. Hence, applying the DD1-criterion yields, for any  $\hat{m}_j$ ,  $\Theta^*(0, 1, \hat{m}_j) = \Theta^*(0, 0, \hat{m}_j) = \{0, 1\}$ , i.e., the DD1-criterion assigns an intermediate image of  $\pi$  to any action when  $\hat{m}_i = 0$ . Consider the first stage. A deviation to  $(0, 0)$  yields the same payoff as  $(1, 0)$  and is therefore not profitable, while a deviation to  $(0, 1)$  yields at most the same image as  $(1, 1)$  and is therefore also not profitable, which finishes the first part.

- (ii) Suppose that high types contribute if they have got a high signal and let  $\tilde{\theta}_i$  denote any consistent belief system. Consider the second stage and suppose that  $\hat{m}_i = 1$ . The high type with  $s_i = 1$  has no incentives to deviate to  $\hat{a}_i = 0$  iff

$$\begin{aligned} & U_i^*(1, \hat{a}_i = 1 | \theta_i, s_i, \tilde{\theta}_i) \geq U_i^*(1, \hat{a}_i = 0 | \theta_i, s_i, \tilde{\theta}_i) \\ \Leftrightarrow & 2E[W|s_i] - c + \mu \geq \mu(Pr(s_j = 0 | s_i)E[\tilde{\theta}_i(\theta_j, 0, 1, 0)] + Pr(s_j = 1 | s_i)E[\tilde{\theta}_i(\theta_j, 1, 1, 0)]). \end{aligned} \quad (11)$$

Notice that

$$E[\tilde{\theta}_i(\theta_j, 0, 1, 0)] = \frac{2\pi}{3 - \pi} \text{ and } E[\tilde{\theta}_i(\theta_j, 1, 1, 0)] = \frac{\pi}{3 - 2\pi}. \quad (12)$$

<sup>13</sup>Notice that if (9) holds with equality, then the DD1-criterion assigns this deviation to the high type as it is weakly beneficial for her in case it yields a high image. However, not contributing remains a best response.

Substituting (12) into (11) yields

$$\begin{aligned} \frac{4}{3} - c + \mu &\geq \frac{2\pi\mu}{3} \left( \frac{1}{3-2\pi} + \frac{1}{3-\pi} \right) \Leftrightarrow \frac{4}{3} - c \geq -\mu \frac{(1-\pi)(9-4\pi)}{(3-2\pi)(3-\pi)} \\ &\Leftrightarrow \mu \geq \frac{(3c-4)(3-\pi)(3-2\pi)}{3(1-\pi)(9-4\pi)}, \end{aligned} \quad (13)$$

which is the desired lower bound on  $\mu$ . The high type with  $s_i = 0$  has no incentives to deviate to  $\hat{a}_i = 1$  iff

$$\begin{aligned} U_i^*(1, \hat{a}_i = 0 | \theta_i, s_i, \tilde{\theta}_i) &\geq U_i^*(0, \hat{a}_i = 1 | \theta_i, s_i, \tilde{\theta}_i) \\ \Leftrightarrow \mu(Pr(s_j = 0 | s_i)E[\tilde{\theta}_i(\theta_j, 0, 1, 0)] + Pr(s_j = 1 | s_i)E[\tilde{\theta}_i(\theta_j, 1, 1, 0)]) &\geq 2E[W | s_i] - c + \mu, \end{aligned}$$

which always holds as  $2E[W | s_i] - c + \mu = 2/3 - c + \mu < 0$  by Assumption 1.

If  $\hat{m}_i = 0$ , then any action is off-equilibrium. By Assumption 1, low types prefer not to contribute even if contributing would yield a high image and not contributing a low image. A high type with  $s_i = 1$ , on the other hand, would prefer to contribute in this case if (13) holds. Hence, as in part (i), applying the DD1-criterion yields, for any  $\hat{m}_j$ ,  $\Theta^*(0, 1, \hat{m}_j) = \{1\}$  and  $\Theta^*(0, 0, \hat{m}_j) = \{0\}$ , i.e., the DD1-criterion assigns a high (low) image to action  $\hat{a}_i = 1$  ( $\hat{a}_i = 0$ ) when  $\hat{m}_i = 0$ . Consider the first stage. Again, a deviation to  $(0, 0)$  is dominated by  $(1, 0)$  as it yields a lower image, while a deviation to  $(0, 1)$  yields the same payoff as  $(1, 1)$  and is therefore also not profitable, which finishes the second part.

Finally, if (13) does not hold, then (9) does hold since  $(3-\pi)(3-2\pi)/(9-4\pi) < 1$  for all  $\pi \in (0, 1)$ , which establishes existence of an equilibrium without information transmission. Accounting for the bounds from Assumption 1 finishes the proof.

## A.2 Proof of Proposition 2

First, notice that there are in total ten equilibrium candidates with consequential communication: all strategy profiles with truthful communication (both types are truthful; one type is truthful, while the other type over-/underreports) and in which high reports increase the likelihood of a contribution by the high type (contribution conditional on a high signal or/and a high report). We proceed by case distinction with respect to the types' communication strategy and show that all strategy profiles except those in which the low type is a denialist and the high type is truthful cannot be part of a DD1-equilibrium.

- (i) Suppose that both types are truthful and the high type contributes conditional on a high signal or/and a high report.

First, suppose that the high type contributes conditional on a high signal and a high report. Take a low type with  $s_i = 1$  and consider a deviation to  $(m'_i, a'_i) = (0, 0)$ . Agent  $i$  does not have incentives to do this deviation iff

$$\begin{aligned} U_i^*(m_i^*, a_i^* | \theta_i, s_i, \tilde{\theta}_i) &\geq U_i^*(m'_i, a'_i | \theta_i, s_i, \tilde{\theta}_i) \\ \Leftrightarrow Pr(s_j = 0 | s_i)\mu\pi + Pr(s_j = 1 | s_i)\pi E[W | s_i, s_j = 1] &\geq \mu\pi \\ \Leftrightarrow \mu &\leq 3/4. \end{aligned} \quad (14)$$

Next, take a high type with  $s_i = 0$  and consider a deviation to  $(m'_i, a'_i) = (1, 0)$ . Agent  $i$  does not have incentives to do this deviation iff

$$\begin{aligned}
& U_i^*(m_i^*, a_i^* | \theta_i, s_i, \tilde{\theta}_i) \geq U_i^*(m'_i, a'_i | \theta_i, s_i, \tilde{\theta}_i) \\
& \Leftrightarrow \mu\pi \geq \Pr(s_j = 0 | s_i) \mu\pi + \Pr(s_j = 1 | s_i) \pi 2E[W | s_i, s_j = 1] \\
& \Leftrightarrow \mu \geq 1.
\end{aligned} \tag{15}$$

As (14) and (15) cannot both hold, such an equilibrium does not exist.

Second, suppose that the high type contributes conditional on a high signal or a high report. Consider the first stage and let  $\tilde{\theta}_i$  denote any consistent belief system. Take a low type with  $s_i = 1$ . As contributing is never profitable for the low type, we only need to check a deviation to  $(m'_i, a'_i) = (0, 0)$ . Agent  $i$  does not have incentives to do this deviation iff

$$\begin{aligned}
& U_i^*(m_i^*, a_i^* | \theta_i, s_i, \tilde{\theta}_i) \geq U_i^*(m'_i, a'_i | \theta_i, s_i, \tilde{\theta}_i) \\
& \Leftrightarrow \Pr(s_j = 0 | s_i = 1) \pi E[W | s_i = 1, s_j = 0] + \Pr(s_j = 1 | s_i = 1) \pi E[W | s_i = 1, s_j = 1] \\
& \quad \geq \Pr(s_j = 0 | s_i = 1) \mu\pi + \Pr(s_j = 1 | s_i = 1) \pi E[W | s_i = 1, s_j = 1] \\
& \Leftrightarrow \mu \leq 1/2.
\end{aligned} \tag{16}$$

Next, take a high type with  $s_i = 0$ . As contributing after receiving a high report is optimal and always contributing never optimal, we only need to check a deviation to  $(m'_i, a'_i) = (1, \mathbf{1}_{\{\hat{m}_j=1\}})$ . Agent  $i$  does not have incentives to do this deviation iff

$$\begin{aligned}
& U_i^*(m_i^*, a_i^* | \theta_i, s_i, \tilde{\theta}_i) \geq U_i^*(m'_i, a'_i | \theta_i, s_i, \tilde{\theta}_i) \\
& \Leftrightarrow \Pr(s_j = 0 | s_i = 0) \mu\pi + \Pr(s_j = 1 | s_i = 0) [(1 + \pi) 2E[W | s_i = 0, s_j = 1] - c + \mu] \\
& \quad \geq \Pr(s_j = 0 | s_i = 0) \pi 2E[W | s_i = 0, s_j = 0] \\
& \quad + \Pr(s_j = 1 | s_i = 0) [(1 + \pi) 2E[W | s_i = 0, s_j = 1] - c + \mu] \\
& \Leftrightarrow \mu \geq 1/2.
\end{aligned} \tag{17}$$

Combining (16) and (17) yields  $\mu = 1/2$ . We ignore this equilibrium as we concentrate on equilibria with positive measure of support, which finishes the first part.

- (ii) Suppose that only the low type is truthful and the high type contributes conditional on a high signal or/and a high report.

First, suppose that the high type underreports and contributes conditional on a high signal or a high report. Consider the second stage. If  $\hat{m}_j = 1$  and  $\hat{m}_i = 0$ , then not contributing yields a low image and contributing a high image. Furthermore, notice that  $\hat{m}_j = 1$  implies that  $j$  is a low type with  $s_j = 1$ . Hence, contributing independent of the signal  $s_i$  is optimal for the high type iff

$$c - \mu \leq 2 \min_{s_i} E[W | s_i, \hat{m}_j = 1] = 2E[W | s_i = 0, s_j = 1] = 1 \Leftrightarrow \mu \geq c - 1. \tag{18}$$

If  $\hat{m}_i = 1$ , then not contributing yields a low image, while contributing is off-equilibrium. Notice that contributing may be beneficial if the other player assigns a sufficiently high image to this deviation. Denote this image by  $\hat{\theta}_i$  and let  $\tilde{\theta}_i$  denote any consistent belief



system such that  $\tilde{\theta}_i(\theta_j, s_j, 1, 1) \equiv \hat{\theta}_i$ . Suppose  $\hat{m}_j = 0$  and take a high type with  $s_i = 1$ . Agent  $i$  does not have incentives to take action  $\hat{a}_i = 1$  iff

$$\begin{aligned} & U_i^*(\hat{m}_i, \hat{a}_i = 0 | \theta_i, s_i, \hat{m}_j, \tilde{\theta}_i) \geq U_i^*(\hat{m}_i, \hat{a}_i = 1 | \theta_i, s_i, \hat{m}_j, \tilde{\theta}_i) \\ \Leftrightarrow & 0 \geq \Pr(s_j = 0 | s_i)(2E[W | s_i, s_j = 0] - c + \mu\hat{\theta}_i) + \Pr(s_j = 1 | s_i)\pi(2E[W | s_i, s_j = 1] - c + \mu\hat{\theta}_i) \\ \Leftrightarrow & 0 \geq 1 + 3\pi + (1 + 2\pi)(-c + \mu\hat{\theta}_i) \\ \Leftrightarrow & \hat{\theta}_i \leq \frac{(1 + 2\pi)c - (1 + 3\pi)}{(1 + 2\pi)\mu} \equiv \theta^*(1, 1). \end{aligned}$$

Notice that (18) implies that

$$\theta^*(1, 1) \leq 1 - \frac{\pi}{(1 + 2\pi)\mu} < 1,$$

i.e., if the assigned image is larger than  $\theta^*(1, 1) < 1$ , then the high type with  $s_i = 1$  prefers to take action  $\hat{a}_i = 1$ . Next, consider a low type with any  $s_i$ . Agent  $i$  does not have incentives to take action  $\hat{a}_i = 1$  iff

$$U_i^*(\hat{m}_i, \hat{a}_i = 0 | \theta_i, s_i, \hat{m}_j, \tilde{\theta}_i) \geq U_i^*(\hat{m}_i, \hat{a}_i = 1 | \theta_i, s_i, \hat{m}_j, \tilde{\theta}_i) \Leftrightarrow \hat{\theta}_i \leq \theta^*(0, s_i),$$

where  $\theta^*(0, s_i) > 1$  since by Assumption 1 low types never contribute. Now, we need to apply the DD1-criterion to determine the type that is most likely to take action  $\hat{a}_i = 1$  when  $\hat{m}_j = 0$  and  $\hat{m}_i = 1$ . Notice that  $\theta^*(\theta_i, s_i) > 1$  implies  $D^*(\theta_i, s_i, 1, 1, 0) = \emptyset$ , which therefore holds for low types, while  $D^*(1, 1, 1, 1, 0) \neq \emptyset$  as  $\theta^*(1, 1) < 1$ . Therefore,  $\Theta^*(1, 1, 0) = \{1\}$ , i.e., the criterion assigns a high image,  $\hat{\theta}_i = 1$ , to action  $\hat{a}_i = 1$  when  $\hat{m}_j = 0$  and  $\hat{m}_i = 1$ . Similarly, suppose  $\hat{m}_j = 1$  and take a high type with  $s_i = 1$ . Notice that  $\hat{m}_j = 1$  implies that  $j$  is a low type with  $s_j = 1$ . Agent  $i$  does not have incentives to take action  $\hat{a}_i = 1$  iff

$$\begin{aligned} & U_i^*(\hat{m}_i, \hat{a}_i = 0 | \theta_i, s_i, \hat{m}_j, \tilde{\theta}_i) \geq U_i^*(\hat{m}_i, \hat{a}_i = 1 | \theta_i, s_i, \hat{m}_j, \tilde{\theta}_i) \\ \Leftrightarrow & 0 \geq 2E[W | s_i = 1, s_j = 1] - c + \mu\hat{\theta}_i \\ \Leftrightarrow & 0 \geq 3/2 - c + \mu\hat{\theta}_i \\ \Leftrightarrow & \hat{\theta}_i \leq \frac{c - 3/2}{\mu} \equiv \theta^{**}(1, 1). \end{aligned}$$

Notice that (18) implies  $\theta^{**}(1, 1) < 1$ . Applying the same reasoning as above yields  $\Theta^*(1, 1, 1) = \{1\}$ , i.e., the criterion assigns a high image,  $\hat{\theta}_i = 1$ , to action  $\hat{a}_i = 1$  when  $\hat{m}_j = 1$  and  $\hat{m}_i = 1$ . Therefore, let  $\theta_i$  denote any consistent belief system such that  $\tilde{\theta}_i(\theta_j, s_j, 1, 1) \equiv 1$  and consider the first stage. Take a high type with  $s_i = 1$  and consider a deviation to  $(m'_i, a'_i) = (1, 1)$ . Agent  $i$  does not have incentives to do this deviation iff

$$\begin{aligned} & U_i^*(m_i^*, a_i^* | \theta_i, s_i, \tilde{\theta}_i) \geq U_i^*(m'_i, a'_i | \theta_i, s_i, \tilde{\theta}_i) \\ \Leftrightarrow & \Pr(s_j = 0 | s_i)2E[W | s_i, s_j = 0] + \Pr(s_j = 1 | s_i)(1 + \pi)2E[W | s_i, s_j = 1] - c + \mu \\ & \geq \Pr(s_j = 0 | s_i)(1 + \pi)2E[W | s_i, s_j = 0] + \Pr(s_j = 1 | s_i)(1 + \pi)2E[W | s_i, s_j = 1] - c + \mu \\ \Leftrightarrow & 0 \geq \pi, \end{aligned}$$

which implies that such an equilibrium does not exist.

Second, suppose that the high type overreports and contributes conditional on a high signal

or a high report. Then, submitting a low report and not contributing yields a low image. Take a low type with  $s_i = 0$  and consider a deviation to  $(m'_i, a'_i) = (1, 0)$ . Agent  $i$  does not have incentives to do this deviation iff

$$\begin{aligned} & U_i^*(m_i^*, a_i^* | \theta_i, s_i, \tilde{\theta}_i) \geq U_i^*(m'_i, a'_i | \theta_i, s_i, \tilde{\theta}_i) \\ \Leftrightarrow & Pr(s_j = 1 | s_i) \pi E[W | s_i, s_j = 1] \geq Pr(s_j = 0 | s_i) [\pi E[W | s_i, s_j = 0] \\ & + (1 - \pi) \mu E[\tilde{\theta}_i(\theta_j, 0, 1, 0)]] + Pr(s_j = 1 | s_i) \pi E[W | s_i, s_j = 1] \\ \Leftrightarrow & 0 \geq \pi/4 + (1 - \pi) \mu E[\tilde{\theta}_i(\theta_j, 0, 1, 0)], \end{aligned}$$

which does not hold as the right-hand-side is strictly positive. Hence, such an equilibrium does not exist.

Third, suppose that the high type underreports and contributes conditional on a high signal and a high report. Consider the second stage. If  $\hat{m}_j = 1$  and  $\hat{m}_i = 1$ , then not contributing yields a low image, while contributing is off-equilibrium. Denote the image in this case by  $\hat{\theta}_i$  and let  $\tilde{\theta}_i$  denote any consistent belief system such that  $\tilde{\theta}_i(0, 1, 1, 1) = \hat{\theta}_i$ . Also notice that  $\hat{m}_j = 1$  implies that  $j$  is a low type with  $s_j = 1$ . Take a high type with  $s_i = 1$ . Agent  $i$  does not have incentives to take action  $\hat{a}_i = 1$  iff

$$\begin{aligned} U_i^*(\hat{m}_i, \hat{a}_i = 0 | \theta_i, s_i, \hat{m}_j, \tilde{\theta}_i) \geq U_i^*(\hat{m}_i, \hat{a}_i = 1 | \theta_i, s_i, \hat{m}_j, \tilde{\theta}_i) & \Leftrightarrow 0 \geq 2E[W | s_i, s_j = 1] - c + \mu \hat{\theta}_i \\ & \Leftrightarrow 0 \geq 3/2 - c + \mu \hat{\theta}_i \\ & \Leftrightarrow \hat{\theta}_i \leq \frac{c - 3/2}{\mu} \equiv \theta^{***}(1, 1). \end{aligned}$$

Notice that Assumption 1 implies  $\theta^{***}(1, 1) < 1$ . Applying the same reasoning as in the first case yields  $\Theta^*(1, 1, 1) = \{1\}$ , i.e., the criterion assigns a high image,  $\hat{\theta}_i = 1$ , to action  $\hat{a}_i = 1$  when  $\hat{m}_j = 1$  and  $\hat{m}_i = 1$ . Therefore, let  $\tilde{\theta}_i$  denote any consistent belief system such that  $\tilde{\theta}_i(0, 1, 1, 1) = 1$  and consider the first stage. Take a low type with  $s_i = 1$  and consider a deviation to  $(m'_i, a'_i) = (0, 0)$ . Agent  $i$  does not have incentives to do this deviation iff

$$\begin{aligned} & U_i^*(m_i^*, a_i^* | \theta_i, s_i, \tilde{\theta}_i) \geq U_i^*(m'_i, a'_i | \theta_i, s_i, \tilde{\theta}_i) \\ \Leftrightarrow & Pr(s_j = 1 | s_i) \pi E[W | s_i, s_j = 1] \geq Pr(s_j = 0 | s_i) \mu E[\tilde{\theta}_i(\theta_j, 0, 0, 0)] \\ & + Pr(s_j = 1 | s_i) [\pi \mu E[\tilde{\theta}_i(\theta_j, 1, 0, 0)] + (1 - \pi) \mu \pi] \\ \Leftrightarrow & 3\pi/2 \geq \mu (E[\tilde{\theta}_i(\theta_j, 0, 0, 0)] + 2\pi E[\tilde{\theta}_i(\theta_j, 1, 0, 0)] + 2\pi(1 - \pi)). \end{aligned} \quad (19)$$

Next, notice that

$$E[\tilde{\theta}_i(\theta_j, 0, 0, 0)] = \frac{3\pi}{2 + \pi} \text{ and } E[\tilde{\theta}_i(\theta_j, 1, 0, 0)] = \frac{3\pi}{1 + 2\pi}. \quad (20)$$

Substituting (20) into (19) yields

$$\begin{aligned} 3\pi/2 \geq \mu \left( \frac{3\pi}{2 + \pi} + \frac{6\pi^2}{1 + 2\pi} + 2\pi(1 - \pi) \right) & \Leftrightarrow 3/2 \geq \mu \frac{7 + 24\pi - 4\pi^3}{(2 + \pi)(1 + 2\pi)} \\ & \Leftrightarrow \mu \leq \frac{3(2 + \pi)(1 + 2\pi)}{2(7 + 24\pi - 4\pi^3)}. \end{aligned} \quad (21)$$

Next, take a high type with  $s_i = 1$  and consider a deviation to  $(m'_i, a'_i) = (1, \mathbf{1}_{\{\hat{m}_j=1\}})$ .

Recall that this deviation yields a high image if  $\hat{m}_j = 1$  by the DD1-criterion. Agent  $i$  does not have incentives to do this deviation iff

$$\begin{aligned}
& U_i^*(m_i^*, a_i^* | \theta_i, s_i, \tilde{\theta}_i) \geq U_i^*(m_i', a_i' | \theta_i, s_i, \tilde{\theta}_i) \\
& \Leftrightarrow Pr(s_j = 0 | s_i) \mu E[\tilde{\theta}_i(\theta_j, 0, 0, 0)] + Pr(s_j = 1 | s_i) [\pi \mu E[\tilde{\theta}_i(\theta_j, 1, 0, 0)] \\
& \quad + (1 - \pi)(2E[W | s_i, s_j = 1] + \mu - c)] \geq Pr(s_j = 1 | s_i) [2E[W | s_i, s_j = 1] + (1 - \pi)(\mu - c)] \\
& \Leftrightarrow \mu(E[\tilde{\theta}_i(\theta_j, 0, 0, 0)] + 2\pi E[\tilde{\theta}_i(\theta_j, 1, 0, 0)]) \geq 3\pi. \tag{22}
\end{aligned}$$

Substituting (20) into (22) yields

$$\mu \left( \frac{3\pi}{2 + \pi} + \frac{6\pi^2}{1 + 2\pi} \right) \geq 3\pi \Leftrightarrow \mu \geq \frac{(2 + \pi)(1 + 2\pi)}{1 + 6\pi + 2\pi^2}. \tag{23}$$

As

$$\frac{1}{1 + 6\pi + 2\pi^2} > \frac{3}{2(7 + 24\pi - 4\pi^3)},$$

(21) and (23) cannot both hold, which implies that such an equilibrium does not exist.

Finally, suppose that the high type overreports and contributes conditional on a high signal and a high report. Then, submitting a low report and not contributing yields a low image. Take a low type with  $s_i = 0$  and consider a deviation to  $(m_i', a_i') = (1, 0)$ . Agent  $i$  does not have incentives to do this deviation iff

$$\begin{aligned}
& U_i^*(m_i^*, a_i^* | \theta_i, s_i, \tilde{\theta}_i) \geq U_i^*(m_i', a_i' | \theta_i, s_i, \tilde{\theta}_i) \\
& \Leftrightarrow 0 \geq Pr(s_j = 0 | s_i) \mu E[\tilde{\theta}_i(\theta_j, 0, 1, 0)] + Pr(s_j = 1 | s_i) [\pi E[W | s_i, s_j = 1] + \mu E[\tilde{\theta}_i(\theta_j, 1, 1, 0)]],
\end{aligned}$$

which does not hold as the right-hand-side is strictly positive. Hence, such an equilibrium does not exist, which finishes the second part.

- (iii) Suppose that the low type is alarmist and the high type is truthful and contributes conditional on a high signal or/and a high report.

First, suppose that the high type contributes conditional on a high signal or a high report. Consider the second stage. If  $\hat{m}_j = 1$  and  $\hat{m}_i = 0$ , then contributing yields a high image, while not contributing is off-equilibrium. Denote the image in this case by  $\hat{\theta}_i$  and let  $\tilde{\theta}_i$  denote any consistent belief system such that  $\tilde{\theta}_i(\theta_j, s_j, 0, 1) \equiv \hat{\theta}_i$ . Take a high type with  $s_i = 1$ . Agent  $i$  does not have incentives to take action  $\hat{a}_i = 0$  iff

$$\begin{aligned}
U_i^*(\hat{m}_i, \hat{a}_i = 1 | \theta_i, s_i, \hat{m}_j, \tilde{\theta}_i) \geq U_i^*(\hat{m}_i, \hat{a}_i = 0 | \theta_i, s_i, \hat{m}_j, \tilde{\theta}_i) & \Leftrightarrow 2E[W | s_i, s_j = 1] - c + \mu \geq \mu \hat{\theta}_i \\
& \Leftrightarrow \hat{\theta}_i \leq \frac{3/2 - (c - \mu)}{\mu} \equiv \theta^*(1, 1).
\end{aligned}$$

Notice that Assumption 1 implies that  $\theta^*(1, 1) > 0$ , i.e., if the assigned image is at most  $\theta^*(1, 1)$ , then the high type with  $s_i = 1$  has no incentives to take action  $\hat{a}_i = 0$ . Next, consider a low type with any  $s_i$ . Agent  $i$  does not have incentives to take action  $\hat{a}_i = 0$  iff

$$U_i^*(\hat{m}_i, \hat{a}_i = 1 | \theta_i, s_i, \hat{m}_j, \tilde{\theta}_i) \geq U_i^*(\hat{m}_i, \hat{a}_i = 0 | \theta_i, s_i, \hat{m}_j, \tilde{\theta}_i) \Leftrightarrow \hat{\theta}_i \leq \theta^*(0, s_i),$$

where  $\theta^*(0, s_i) < 0$  since by Assumption 1 low types never contribute. Now, we need to apply the DD1-criterion to determine the type that is most likely to take action  $\hat{a}_i = 0$

when  $\hat{m}_j = 1$  and  $\hat{m}_i = 0$ . Notice that  $\theta^*(\theta_i, s_i) < 0$  implies  $D^*(\theta_i, s_i, 0, 0, 1) = [0, 1]$ , which therefore holds for low types, while  $D^*(1, 1, 0, 0, 1) \neq \emptyset$  as  $\theta^*(1, 1) > 0$ . Therefore,  $\Theta^*(0, 0, 1) = \{0\}$ , i.e., the criterion assigns a low image,  $\hat{\theta}_i = 0$ , to action  $\hat{a}_i = 0$  when  $\hat{m}_j = 1$  and  $\hat{m}_i = 0$ . Therefore, let  $\tilde{\theta}_i$  denote any consistent belief system such that  $\tilde{\theta}_i(\theta_j, s_j, 0, 0) = 0$  for  $(\theta_j, s_j) \neq (1, 0)$  and consider the first stage. Take a high type with  $s_i = 0$  and consider a deviation to  $(m'_i, a'_i) = (1, \mathbf{1}_{\{\hat{m}_j=1\}})$ . Agent  $i$  does not have incentives to do this deviation iff

$$\begin{aligned}
& U_i^*(m_i^*, a_i^* | \theta_i, s_i, \tilde{\theta}_i) \geq U_i^*(m'_i, a'_i | \theta_i, s_i, \tilde{\theta}_i) \\
& \Leftrightarrow Pr(s_j = 0 | s_i) [\mu + (1 - \pi)(2E[W | s_i, s_j = 0] - c)] \\
& \quad + Pr(s_j = 1 | s_i) [(1 + \pi)2E[W | s_i, s_j = 1] + \mu - c] \\
& \geq Pr(s_j = 0 | s_i) [2E[W | s_i, s_j = 0] + (1 - \pi)(\mu - c)] \\
& \quad + Pr(s_j = 1 | s_i) [(1 + \pi)2E[W | s_i, s_j = 1] + \mu - c] \\
& \Leftrightarrow \mu + (1 - \pi)2E[W | s_i, s_j = 0] \geq 2E[W | s_i, s_j = 0] + (1 - \pi)\mu \\
& \Leftrightarrow \mu \geq 1/2.
\end{aligned} \tag{24}$$

Next, take a low type with  $s_i = 0$  and consider a deviation to  $(m'_i, a'_i) = (0, 0)$ . Recall that this deviation yields a low image if  $\hat{m}_j = 1$  by the DD1-criterion. Agent  $i$  does not have incentives to do this deviation iff

$$\begin{aligned}
& U_i^*(m_i^*, a_i^* | \theta_i, s_i, \tilde{\theta}_i) \geq U_i^*(m'_i, a'_i | \theta_i, s_i, \tilde{\theta}_i) \\
& \Leftrightarrow Pr(s_j = 0 | s_i) \pi E[W | s_i, s_j = 0] + Pr(s_j = 1 | s_i) \pi E[W | s_i, s_j = 1] \\
& \geq Pr(s_j = 0 | s_i) \pi \mu + Pr(s_j = 1 | s_i) \pi E[W | s_i, s_j = 1] \\
& \Leftrightarrow \mu \leq 1/4.
\end{aligned} \tag{25}$$

As (24) and (25) cannot both hold, such an equilibrium does not exist.

Second, suppose that the high type contributes conditional on a high signal and a high report. Then, submitting a low report and not contributing yields a high image. Take a low type with  $s_i = 0$  and consider a deviation to  $(m'_i, a'_i) = (0, 0)$ . Agent  $i$  does not have incentives to do this deviation iff

$$\begin{aligned}
& U_i^*(m_i^*, a_i^* | \theta_i, s_i, \tilde{\theta}_i) \geq U_i^*(m'_i, a'_i | \theta_i, s_i, \tilde{\theta}_i) \\
& \Leftrightarrow Pr(s_j = 0 | s_i) \pi \mu E[\tilde{\theta}_i(\theta_j, 0, 1, 0)] + Pr(s_j = 1 | s_i) \pi E[W | s_i, s_j = 1] \geq \mu \\
& \Leftrightarrow 2\pi \mu E[\tilde{\theta}_i(\theta_j, 0, 1, 0)] + \pi/2 \geq 3\mu.
\end{aligned}$$

Notice that

$$E[\tilde{\theta}_i(\theta_j, 0, 1, 0)] = \frac{\pi}{3 - 2\pi}, \tag{26}$$

which yields

$$\mu \frac{2\pi^2}{3 - 2\pi} + \pi/2 \geq 3\mu \Leftrightarrow \pi/2 \geq \mu \frac{9 - 6\pi - 2\pi^2}{3 - 2\pi} \Leftrightarrow \mu \leq \frac{3\pi - 2\pi^2}{2(9 - 6\pi - 2\pi^2)}. \tag{27}$$

Next, take a high type with  $s_i = 0$  and consider a deviation to  $(m'_i, a'_i) = (1, 0)$ . Agent  $i$

does not have incentives to do this deviation iff

$$\begin{aligned}
& U_i^*(m_i^*, a_i^* | \theta_i, s_i, \tilde{\theta}_i) \geq U_i^*(m_i', a_i' | \theta_i, s_i, \tilde{\theta}_i) \\
& \Leftrightarrow \mu \geq Pr(s_j = 0 | s_i) \pi \mu E[\tilde{\theta}_i(\theta_j, 0, 1, 0)] + Pr(s_j = 1 | s_i) \pi 2E[W | s_i, s_j = 1] \\
& \Leftrightarrow 3\mu \geq 2\pi \mu E[\tilde{\theta}_i(\theta_j, 0, 1, 0)] + \pi.
\end{aligned}$$

Substituting (26) yields

$$3\mu \geq \mu \frac{2\pi^2}{3-2\pi} + \pi \Leftrightarrow \mu \geq \frac{3\pi - 2\pi^2}{9 - 6\pi - 2\pi^2}. \quad (28)$$

As (27) and (28) cannot both hold, such an equilibrium does not exist, which finishes the proof.

### A.3 Proof of Proposition 3

Suppose  $\mu < 1/2$ . By Proposition 2, we only need to show that the communication strategy profile in which the low type is a denialist and the high type is truthful is not part of an equilibrium with consequential communication.

First, suppose that the high type contributes conditional on a high signal or a high report. Notice that

$$\frac{(3-\pi)(3-2\pi)}{\pi^2 - 12\pi + 15} > \frac{1}{2} > \mu \text{ for all } \pi \in (0, 1),$$

i.e., Theorem 1 (presented later in Section 4) implies that this strategy profile is not an equilibrium. Second, suppose that the high type contributes conditional on a high signal and a high report. Let  $B(\pi) \equiv \pi(2\pi^2 - 24\pi + 51) - 27$  and notice that if  $B(\pi) > 0$ , then

$$\frac{\pi(3-\pi)(3-2\pi)}{B(\pi)} > 1 > \mu \text{ for all } \pi \in (0, 1),$$

i.e., Proposition 4 (presented later in Section 4) implies that also this strategy profile is not an equilibrium, which finishes the proof.

### A.4 Proof of Theorem 1 and Proposition 6

Proposition 6 in Section 7 extends Theorem 1 to impure public goods and lower intrinsic motivation of high types to contribute to the public good. In the following, we first prove Proposition 6 and then derive Theorem 1 by setting  $\beta = 1$  and  $\bar{\theta} = 1$ .

Suppose that low types are denialist and high types are truthful and contribute conditional on a high signal or a high report. Consider first the second stage. If  $\hat{m}_j = 1$  and  $\hat{m}_i = 0$ , then not contributing yields a low image and contributing a high image. Furthermore, notice that  $\hat{m}_j = 1$  implies that  $j$  is a high type with  $s_j = 1$ . Hence, contributing independent of the signal  $s_i$  is optimal for the high type iff

$$\begin{aligned}
& c - \mu \leq (1 + \bar{\theta}) \min_{s_i} E[W | s_i, \hat{m}_j = 1] = (1 + \bar{\theta}) E[W | s_i = 0, s_j = 1] = (1 + \bar{\theta})/2 \\
& \Leftrightarrow \mu \geq c - (1 + \bar{\theta})/2,
\end{aligned} \quad (29)$$

which is the first desired lower bound on  $\mu$ .

If  $\hat{m}_j = 1$  and  $\hat{m}_i = 1$ , then contributing yields a high image, while not contributing is off-equilibrium. Notice that not contributing may be beneficial if the other player assigns a sufficiently high image to this deviation. Denote this image by  $\hat{\theta}_i$  and let  $\tilde{\theta}_i$  denote any consistent belief system such that  $\tilde{\theta}_i(\theta_j, s_j, 1, 0) \equiv \hat{\theta}_i$ . Take a high type with  $s_i = 1$ . Agent  $i$  does not have incentives to take action  $\hat{a}_i = 0$  iff

$$\begin{aligned} U_i^*(\hat{m}_i, \hat{a}_i = 1 | \theta_i, s_i, \hat{m}_j, \tilde{\theta}_i) \geq U_i^*(\hat{m}_i, \hat{a}_i = 0 | \theta_i, s_i, \hat{m}_j, \tilde{\theta}_i) &\Leftrightarrow (1 + \bar{\theta})E[W | s_i, s_j = 1] - c + \mu \geq \mu \hat{\theta}_i \\ &\Leftrightarrow \hat{\theta}_i \leq \frac{3(1 + \bar{\theta})/4 - (c - \mu)}{\mu} \equiv \theta^*(1, 1). \end{aligned}$$

Notice that (29) implies that

$$\theta^*(1, 1) \geq \frac{1 + \bar{\theta}}{4\mu} > 0,$$

i.e., if the assigned image is at most  $\theta^*(1, 1) > 0$ , then the high type with  $s_i = 1$  does not want to take action  $\hat{a}_i = 0$ . Next, consider a low type with any  $s_i$ . Similarly, agent  $i$  does not have incentives to take action  $\hat{a}_i = 0$  iff

$$U_i^*(\hat{m}_i, \hat{a}_i = 1 | \theta_i, s_i, \hat{m}_j, \tilde{\theta}_i) \geq U_i^*(\hat{m}_i, \hat{a}_i = 0 | \theta_i, s_i, \hat{m}_j, \tilde{\theta}_i) \Leftrightarrow \hat{\theta}_i \leq \theta^*(0, s_i),$$

where  $\theta^*(0, s_i) < 0$  since by Assumption 1 low types never contribute. Thus, low types even have an incentive to take action  $\hat{a}_i = 0$  if this yields a low image. Now, we need to apply the DD1-criterion to determine the type that is most likely to take action  $\hat{a}_i = 0$  when  $\hat{m}_j = 1$  and  $\hat{m}_i = 1$ . Notice that  $\theta^*(\theta_i, s_i) < 0$  implies  $D^*(\theta_i, s_i, 1, 0, 1) = [0, 1]$ , which therefore holds for low types, while  $D^*(1, 1, 1, 0, 1) \subsetneq [0, 1]$  as  $\theta^*(1, 1) > 0$ . Therefore,  $\Theta^*(1, 0, 1) = \{0\}$ , i.e., the criterion assigns a low image,  $\hat{\theta}_i = 0$ , to action  $\hat{a}_i = 0$  when  $\hat{m}_j = 1$  and  $\hat{m}_i = 1$ .

If  $\hat{m}_j = 0$  and  $\hat{m}_i = 1$ , again contributing yields a high image, while not contributing is off-equilibrium. Let  $\tilde{\theta}_i$  denote any consistent belief system such that  $\tilde{\theta}_i(\theta_j, s_j, 1, 0) \equiv \hat{\theta}_i$ . Take a high type with  $s_i = 0$ . Agent  $i$  does not have incentives to take action  $\hat{a}_i = 0$  iff

$$\begin{aligned} U_i^*(\hat{m}_i, \hat{a}_i = 1 | \theta_i, s_i, \hat{m}_j, \tilde{\theta}_i) &\geq U_i^*(\hat{m}_i, \hat{a}_i = 0 | \theta_i, s_i, \hat{m}_j, \tilde{\theta}_i) \\ \Leftrightarrow Pr(s_j = 0 | s_i) &((1 + \bar{\theta})E[W | s_i, s_j = 0] - c + \mu) + Pr(s_j = 1 | s_i)(1 - \pi)((1 + \bar{\theta})E[W | s_i, s_j = 1] \\ &- c + \mu) \geq Pr(s_j = 0 | s_i)\mu\hat{\theta}_i + Pr(s_j = 1 | s_i)(1 - \pi)\mu\hat{\theta}_i \\ \Leftrightarrow (1 + \bar{\theta})Pr(s_j = 0 | s_i)E[W | s_i, s_j = 0] &+ (1 + \bar{\theta})Pr(s_j = 1 | s_i)(1 - \pi)E[W | s_i, s_j = 1] \\ &+ (Pr(s_j = 0 | s_i) + Pr(s_j = 1 | s_i)(1 - \pi))(\mu(1 - \hat{\theta}_i) - c) \geq 0 \\ \Leftrightarrow (1 + \bar{\theta})(1 - \pi/2) - (3 - \pi)(c - \mu) &\geq (3 - \pi)\mu\hat{\theta}_i \\ \Leftrightarrow \hat{\theta}_i \leq \frac{(1 + \bar{\theta})(1 - \pi/2)}{\mu(3 - \pi)} - \frac{c - \mu}{\mu} &\equiv \theta^{**}(1, 0). \end{aligned}$$

Notice that  $(1 + \bar{\theta})(1 - \pi/2)/(3 - \pi) < 2/3$  for all  $\pi \in (0, 1)$  and  $\bar{\theta} \in (0, 1]$ , which yields, together with Assumption 1,

$$\theta^{**}(1, 0) < \frac{2}{3\mu} - \frac{3}{4\mu} < 0,$$

i.e., the high type with  $s_i = 0$  even has an incentive to take action  $\hat{a}_i = 0$  if this yields a low

image. If  $s_i = 1$ , then agent  $i$  does not have incentives to take action  $\hat{a}_i = 0$  iff

$$\begin{aligned}
& U_i^*(\hat{m}_i, \hat{a}_i = 1 | \theta_i, s_i, \hat{m}_j, \tilde{\theta}_i) \geq U_i^*(\hat{m}_i, \hat{a}_i = 0 | \theta_i, s_i, \hat{m}_j, \tilde{\theta}_i) \\
& \Leftrightarrow Pr(s_j = 0 | s_i) ((1 + \bar{\theta})E[W | s_i, s_j = 0] - c + \mu) + Pr(s_j = 1 | s_i) (1 - \pi) ((1 + \bar{\theta})E[W | s_i, s_j = 1] \\
& \quad - c + \mu) \geq Pr(s_j = 0 | s_i) \mu \hat{\theta}_i + Pr(s_j = 1 | s_i) (1 - \pi) \mu \hat{\theta}_i \\
& \Leftrightarrow (1 + \bar{\theta}) Pr(s_j = 0 | s_i) E[W | s_i, s_j = 0] + (1 + \bar{\theta}) Pr(s_j = 1 | s_i) (1 - \pi) E[W | s_i, s_j = 1] \\
& \quad + (Pr(s_j = 0 | s_i) + Pr(s_j = 1 | s_i) (1 - \pi)) (\mu (1 - \hat{\theta}_i) - c) \geq 0 \\
& \Leftrightarrow (1 + \bar{\theta}) (2 - 3\pi/2) - (3 - 2\pi)(c - \mu) \geq (3 - 2\pi) \mu \hat{\theta}_i \\
& \Leftrightarrow \hat{\theta}_i \leq \frac{(1 + \bar{\theta})(2 - 3\pi/2) - c + \mu}{\mu(3 - 2\pi)} \equiv \theta^{**}(1, 1).
\end{aligned}$$

Notice that  $(2 - 3\pi/2)/(3 - 2\pi) > 1/2$  for all  $\pi \in (0, 1)$ , which yields, together with (29),

$$\theta^{**}(1, 1) > \frac{1 + \bar{\theta}}{2\mu} - \frac{1 + \bar{\theta}}{2\mu} = 0,$$

i.e., if the assigned image is at most  $\theta^{**}(1, 1) > 0$ , then the high type with  $s_i = 1$  does not want to take action  $\hat{a}_i = 0$ . Next, consider a low type with any  $s_i$ . As in the previous case,  $\theta^{**}(0, s_i) < 0$  by Assumption 1 and therefore,  $\Theta^*(1, 0, 0) = \{0\}$ , i.e., the criterion assigns a low image,  $\hat{\theta}_i = 0$ , to action  $\hat{a}_i = 0$  when  $\hat{m}_j = 0$  and  $\hat{m}_i = 1$ .

Finally, if  $\hat{m}_j = 0$  and  $\hat{m}_i = 0$ , then not contributing yields an intermediate image of  $\tilde{\theta}_i(\theta_j, s_j, 0, 0)$  depending on the type and signal of the other agent, while contributing is off-equilibrium. By Assumption 1, the low type never contributes, while contributing after submitting a low report may be beneficial for the high type if the other player assigns a sufficiently high image to this deviation. As before, denote this image by  $\hat{\theta}_i$  and let  $\tilde{\theta}_i$  denote any consistent belief system such that  $\tilde{\theta}_i(\theta_j, s_j, 0, 1) \equiv \hat{\theta}_i$ . Take a high type with  $s_i = 0$ . Agent  $i$  does not have incentives to take action  $\hat{a}_i = 1$  iff

$$\begin{aligned}
& U_i^*(\hat{m}_i, \hat{a}_i = 0 | \theta_i, s_i, \hat{m}_j, \tilde{\theta}_i) \geq U_i^*(\hat{m}_i, \hat{a}_i = 1 | \theta_i, s_i, \hat{m}_j, \tilde{\theta}_i) \\
& \Leftrightarrow Pr(s_j = 0 | s_i) \mu E[\tilde{\theta}_i(\theta_j, 0, 0, 0)] + Pr(s_j = 1 | s_i) (1 - \pi) \mu \tilde{\theta}_i(0, 1, 0, 0) \geq Pr(s_j = 0 | s_i) \cdot \\
& \quad ((1 + \bar{\theta})E[W | s_i, s_j = 0] - c + \mu \hat{\theta}_i) + Pr(s_j = 1 | s_i) (1 - \pi) ((1 + \bar{\theta})E[W | s_i, s_j = 1] - c + \mu \hat{\theta}_i) \\
& \Leftrightarrow \mu \frac{2E[\tilde{\theta}_i(\theta_j, 0, 0, 0)] + (1 - \pi) \tilde{\theta}_i(0, 1, 0, 0)}{3 - \pi} \geq \frac{(1 + \bar{\theta})(1 - \pi/2)}{3 - \pi} - c + \mu \hat{\theta}_i. \tag{30}
\end{aligned}$$

By Assumption 1, the right-hand-side of (30) is negative for any belief  $\hat{\theta}_i$  and all  $\pi \in (0, 1)$  and  $\bar{\theta} \in (0, 1]$ , i.e., the high type with  $s_i = 0$  does not want to take action  $\hat{a}_i = 1$ . Hence,  $D^*(\theta_i, s_i, 0, 1, 0) = \emptyset$  for  $(\theta_i, s_i) \neq (1, 1)$ , which implies that if the high type with  $s_i = 1$  benefits from contributing for some beliefs of the other agent,  $D^*(1, 1, 0, 1, 0) \neq \emptyset$ , then  $\Theta^*(0, 1, 0) = \{1\}$ , i.e., the criterion assigns a high image,  $\hat{\theta}_i = 1$ , to action  $\hat{a}_i = 1$  when  $\hat{m}_j = 0$  and  $\hat{m}_i = 0$  in this case.

Next, we consider the first stage. Take a high type with  $s_i = 0$ . Notice that deviations that involve contributing independently of  $\hat{m}_j$  are ruled out by the second stage (contributing is not optimal if  $\hat{m}_j = 0$ ). Truthful reporting and never contributing is also ruled out by the second stage if (29) holds (see case  $(\hat{m}_i = 0, \hat{m}_j = 1)$ ). Furthermore, overreporting and never contributing yields the same low image as the previous deviation by the DD1-criterion if  $\hat{m}_j = 1$  (see case  $(\hat{m}_i = 1, \hat{m}_j = 1)$ ) and is therefore also ruled out. Hence, the only deviation left is

overreporting and contributing if the other agent sends a high report, i.e., a deviation to the strategy  $(m'_i, a'_i) = (1, \mathbf{1}_{\{\hat{m}_j=1\}})$ . Recall that this strategy yields a low image if  $\hat{m}_j = 0$  by the DD1-criterion. Therefore, let  $\tilde{\theta}_i$  denote any consistent belief system such that  $\tilde{\theta}_i(\theta_j, s_j, 1, 0) = 0$  for all  $(\theta_j, s_j)$ . Agent  $i$  does not have incentives to do this deviation iff

$$\begin{aligned}
& U_i^*(m_i^*, a_i^* | \theta_i, s_i, \tilde{\theta}_i) \geq U_i^*(m'_i, a'_i | \theta_i, s_i, \tilde{\theta}_i) \\
& \Leftrightarrow Pr(s_j = 0 | s_i) \mu E[\tilde{\theta}_i(\theta_j, 0, 0, 0)] + Pr(s_j = 1 | s_i) [\pi((1 + \bar{\theta})(1 + \beta)E[W | s_i, s_j = 1] - c + \mu) \\
& \quad + (1 - \pi)\mu\tilde{\theta}_i(0, 1, 0, 0)] \geq Pr(s_j = 0 | s_i) \pi(1 + \bar{\theta})\beta E[W | s_i, s_j = 0] \\
& \quad + Pr(s_j = 1 | s_i) \pi((1 + \bar{\theta})(1 + \beta)E[W | s_i, s_j = 1] - c + \mu) \\
& \Leftrightarrow Pr(s_j = 0 | s_i) \mu E[\tilde{\theta}_i(\theta_j, 0, 0, 0)] + Pr(s_j = 1 | s_i) (1 - \pi)\mu\tilde{\theta}_i(0, 1, 0, 0) \\
& \quad \geq Pr(s_j = 0 | s_i) \pi(1 + \bar{\theta})\beta E[W | s_i, s_j = 0] \\
& \Leftrightarrow 2\mu E[\tilde{\theta}_i(\theta_j, 0, 0, 0)] + (1 - \pi)\mu\tilde{\theta}_i(0, 1, 0, 0) \geq \pi(1 + \bar{\theta})\beta/2. \tag{31}
\end{aligned}$$

Next, notice that

$$E[\tilde{\theta}_i(\theta_j, 0, 0, 0)] = \frac{2\pi}{3 - \pi} \text{ and } \tilde{\theta}_i(0, 1, 0, 0) = \frac{\pi}{3 - 2\pi}. \tag{32}$$

Substituting (32) into (31) yields

$$\mu \left( \frac{4}{3 - \pi} + \frac{1 - \pi}{3 - 2\pi} \right) \geq (1 + \bar{\theta})\beta/2 \Leftrightarrow \mu \geq \frac{(1 + \bar{\theta})\beta(3 - \pi)(3 - 2\pi)}{2(\pi^2 - 12\pi + 15)}, \tag{33}$$

which is the second desired lower bound on  $\mu$ . If  $s_i = 1$ , then deviations that involve never contributing are ruled out by the second stage if (29) holds (see cases  $(\hat{m}_i = 0, \hat{m}_j = 1)$  and  $(\hat{m}_i = 1, \hat{m}_j = 1)$ ). Underreporting and always contributing is also ruled out as it only lowers the likelihood of a contribution by the other player compared to the considered equilibrium candidate (high report and always contributing). Hence, the only deviations left are those that involve contributing if the other agent sends a high report, i.e., deviations to the strategies  $(m'_i, a'_i) = (0, \mathbf{1}_{\{\hat{m}_j=1\}})$  and  $(m''_i, a''_i) = (1, \mathbf{1}_{\{\hat{m}_j=1\}})$ . Consider first a deviation to  $(m'_i, a'_i) = (0, \mathbf{1}_{\{\hat{m}_j=1\}})$ . Let  $\tilde{\theta}_i$  denote any consistent belief system and recall that this deviation yields an intermediate image of  $\tilde{\theta}_i(\theta_j, s_j, 0, 0)$  depending on the type and signal of the other agent if  $\hat{m}_j = 0$ . Agent  $i$  does not have incentives to do this deviation iff

$$\begin{aligned}
& U_i^*(m'_i, a'_i | \theta_i, s_i, \tilde{\theta}_i) \geq U_i^*(m_i^*, a_i^* | \theta_i, s_i, \tilde{\theta}_i) \\
& \Leftrightarrow (1 + \pi\beta)(1 + \bar{\theta})(Pr(s_j = 0 | s_i)E[W | s_i, s_j = 0] + Pr(s_j = 1 | s_i)E[W | s_i, s_j = 1]) - c + \mu \\
& \quad \geq Pr(s_j = 0 | s_i) \mu E[\tilde{\theta}_i(\theta_j, 0, 0, 0)] + Pr(s_j = 1 | s_i) [\pi((1 + \bar{\theta})(1 + \beta)E[W | s_i, s_j = 1] \\
& \quad - c + \mu) + (1 - \pi)\mu\tilde{\theta}_i(0, 1, 0, 0)] \\
& \Leftrightarrow (1 + \bar{\theta})(2(1 + \pi\beta) - 3\pi(1 + \beta)/2) - (3 - 2\pi)c \\
& \quad \geq \mu E[\tilde{\theta}_i(\theta_j, 0, 0, 0)] + 2(1 - \pi)\mu\tilde{\theta}_i(0, 1, 0, 0) - (3 - 2\pi)\mu. \tag{34}
\end{aligned}$$



Substituting (32) into (34) yields

$$\begin{aligned}
& (1 + \bar{\theta})(2(1 + \pi\beta) - 3\pi(1 + \beta)/2) - (3 - 2\pi)c \geq \mu \left( \frac{2\pi}{3 - \pi} + \frac{2\pi(1 - \pi)}{3 - 2\pi} - (3 - 2\pi) \right) \\
& \Leftrightarrow (1 + \bar{\theta})(2(1 + \pi\beta) - 3\pi(1 + \beta)/2) - (3 - 2\pi)c \geq \mu \frac{-3(1 - \pi)(2\pi^2 - 10\pi + 9)}{(3 - \pi)(3 - 2\pi)} \\
& \Leftrightarrow \mu \geq \frac{(2(3 - 2\pi)c - (1 + \bar{\theta})(4 - \pi(3 - \beta)))(3 - \pi)(3 - 2\pi)}{6(1 - \pi)(2\pi^2 - 10\pi + 9)}, \tag{35}
\end{aligned}$$

which is the third desired lower bound on  $\mu$ . Second, consider a deviation to  $(m_i'', a_i'') = (1, \mathbf{1}_{\{\hat{m}_j=1\}})$  and recall that this strategy yields a low image if  $\hat{m}_j = 0$  by the DD1-criterion. Therefore, let  $\tilde{\theta}_i$  denote any consistent belief system such that  $\tilde{\theta}_i(\theta_j, s_j, 1, 0) = 0$  for all  $(\theta_j, s_j)$ . Agent  $i$  does not have incentives to do this deviation iff

$$\begin{aligned}
& U_i^*(m_i^*, a_i^* | \theta_i, s_i, \tilde{\theta}_i) \geq U_i^*(m_i'', a_i'' | \theta_i, s_i, \tilde{\theta}_i) \\
& \Leftrightarrow (1 + \pi\beta)(1 + \bar{\theta})(Pr(s_j = 0 | s_i)E[W | s_i, s_j = 0] + Pr(s_j = 1 | s_i)E[W | s_i, s_j = 1]) - c + \mu \\
& \quad \geq Pr(s_j = 0 | s_i)\pi\beta(1 + \bar{\theta})E[W | s_i, s_j = 0] + Pr(s_j = 1 | s_i)\pi((1 + \beta)(1 + \bar{\theta})E[W | s_i, s_j = 1] - c + \mu) \\
& \Leftrightarrow (1 + \bar{\theta})(Pr(s_j = 0 | s_i)E[W | s_i, s_j = 0] + Pr(s_j = 1 | s_i)E[W | s_i, s_j = 1]) - c + \mu \\
& \quad \geq Pr(s_j = 1 | s_i)\pi((1 + \bar{\theta})E[W | s_i, s_j = 1] - c + \mu) \\
& \Leftrightarrow \frac{(1 + \bar{\theta})(2 - 3\pi/2)}{3 - 2\pi} \geq c - \mu
\end{aligned}$$

which is implied by (29) as  $(2 - 3\pi/2)/(3 - 2\pi) > 1/2$  for all  $\pi \in (0, 1)$ . Finally, consider a low type. As the low type never contributes, the only possible deviation is submitting a high report but not contributing,  $(m_i', a_i') = (1, 0)$ . Recall that this strategy yields a low image by the DD1-criterion and let  $\tilde{\theta}_i$  denote any consistent belief system such that  $\tilde{\theta}_i(\theta_j, s_j, 1, 0) = 0$  for all  $(\theta_j, s_j)$ . If  $s_i = 0$ , then agent  $i$  does not have incentives to do this deviation iff

$$\begin{aligned}
& U_i^*(m_i^*, a_i^* | \theta_i, s_i, \tilde{\theta}_i) \geq U_i^*(m_i', a_i' | \theta_i, s_i, \tilde{\theta}_i) \\
& \Leftrightarrow Pr(s_j = 0 | s_i)\mu E[\tilde{\theta}_i(\theta_j, 0, 0, 0)] + Pr(s_j = 1 | s_i)[\pi\beta E[W | s_i, s_j = 1] + (1 - \pi)\mu\tilde{\theta}_i(0, 1, 0, 0)] \\
& \quad \geq Pr(s_j = 0 | s_i)\pi\beta E[W | s_i, s_j = 0] + Pr(s_j = 1 | s_i)\pi\beta E[W | s_i, s_j = 1] \\
& \Leftrightarrow 2\mu E[\tilde{\theta}_i(\theta_j, 0, 0, 0)] + (1 - \pi)\mu\tilde{\theta}_i(0, 1, 0, 0) \geq \pi\beta/2. \tag{36}
\end{aligned}$$

Substituting (32) into (36) yields

$$\mu \left( \frac{4}{3 - \pi} + \frac{1 - \pi}{3 - 2\pi} \right) \geq \frac{\beta}{2} \Leftrightarrow \mu \geq \frac{\beta(3 - \pi)(3 - 2\pi)}{2(\pi^2 - 12\pi + 15)},$$

which is implied by (33). If  $s_i = 1$ , then agent  $i$  does not have incentives to do this deviation iff

$$\begin{aligned}
& U_i^*(m_i^*, a_i^* | \theta_i, s_i, \tilde{\theta}_i) \geq U_i^*(m_i', a_i' | \theta_i, s_i, \tilde{\theta}_i) \\
& \Leftrightarrow Pr(s_j = 0 | s_i)\mu E[\tilde{\theta}_i(\theta_j, 0, 0, 0)] + Pr(s_j = 1 | s_i)[\pi\beta E[W | s_i, s_j = 1] + (1 - \pi)\mu\tilde{\theta}_i(0, 1, 0, 0)] \\
& \quad \geq Pr(s_j = 0 | s_i)\pi\beta E[W | s_i, s_j = 0] + Pr(s_j = 1 | s_i)\pi\beta E[W | s_i, s_j = 1] \\
& \Leftrightarrow \mu E[\tilde{\theta}_i(\theta_j, 0, 0, 0)] + 2(1 - \pi)\mu\tilde{\theta}_i(0, 1, 0, 0) \geq \pi\beta/2. \tag{37}
\end{aligned}$$

Substituting (32) into (37) yields

$$\mu \left( \frac{1}{3-\pi} + \frac{1-\pi}{3-2\pi} \right) \geq \frac{\beta}{4} \Leftrightarrow \mu \geq \frac{\beta(3-\pi)(3-2\pi)}{4(\pi^2-6\pi+6)}, \quad (38)$$

which is the fourth desired lower bound on  $\mu$ . Hence, we have established that no type-signal pair has incentives to deviate if (29), (33), (35) and (38) hold. Notice also that if one of the four conditions does not hold, this claim is no longer valid. Accounting for the bounds from Assumption 1 finishes the proof of Proposition 6.

Next, we derive Theorem 1. Suppose  $\beta = 1$  and  $\bar{\theta} = 1$ . Using (33), we obtain

$$\frac{(3-\pi)(3-2\pi)}{4(\pi^2-6\pi+6)} = \frac{\pi^2-12\pi+15}{4(\pi^2-6\pi+6)} \cdot \frac{(3-\pi)(3-2\pi)}{\pi^2-12\pi+15} < \frac{(3-\pi)(3-2\pi)}{\pi^2-12\pi+15} \leq \mu$$

for all  $\pi \in (0, 1)$ , i.e., (33) implies (38). Hence, no type-signal pair has incentives to deviate if (29), (33) and (35) hold. Again, notice that if one of these conditions does not hold, this claim is no longer valid. Accounting for the bounds from Assumption 1 finishes the proof of Theorem 1.

## A.5 Proof of Proposition 4

Suppose that low types are denialist and high types are truthful and contribute conditional on a high signal and a high report. Consider first the second stage. If  $\hat{m}_j = 1$  and  $\hat{m}_i = 1$ , then contributing yields a high image, while not contributing is off-equilibrium. Notice that not contributing may be beneficial if the other player assigns a sufficiently high image to this deviation. Denote this image by  $\hat{\theta}_i$  and let  $\tilde{\theta}_i$  denote any consistent belief system such that  $\tilde{\theta}_i(\theta_j, s_j, 1, 0) \equiv \hat{\theta}_i$ . Take a high type with  $s_i = 1$ . Agent  $i$  does not have incentives to take action  $\hat{a}_i = 0$  iff

$$\begin{aligned} U_i^*(\hat{m}_i, \hat{a}_i = 1 | \theta_i, s_i, \hat{m}_j, \tilde{\theta}_i) &\geq U_i^*(\hat{m}_i, \hat{a}_i = 0 | \theta_i, s_i, \hat{m}_j, \tilde{\theta}_i) \Leftrightarrow 2E[W | s_i, s_j = 1] - c + \mu \geq \mu \hat{\theta}_i \\ &\Leftrightarrow \hat{\theta}_i \leq \frac{3/2 - (c - \mu)}{\mu} \equiv \theta^*(1, 1). \end{aligned}$$

Notice that Assumption 1 implies that  $\theta^*(1, 1) > 0$ , i.e., if the assigned image is at most  $\theta^*(1, 1) > 0$ , then the high type with  $s_i = 1$  does not want to take action  $\hat{a}_i = 0$ . Next, consider a low type with any  $s_i$ . Recall that by Assumption 1, low types do not contribute even if this yields a low image. Thus,  $\Theta^*(1, 0, 1) = \{0\}$ , i.e., the DD1-criterion assigns a low image,  $\hat{\theta}_i = 0$ , to action  $\hat{a}_i = 0$  when  $\hat{m}_j = 1$  and  $\hat{m}_i = 1$ . In particular, the deviation is not profitable for the high type with  $s_i = 1$ . Take a high type with  $s_i = 0$ . Agent  $i$  does not have incentives to take action  $\hat{a}_i = 1$  iff

$$\begin{aligned} U_i^*(\hat{m}_i, \hat{a}_i = 0 | \theta_i, s_i, \hat{m}_j, \tilde{\theta}_i) &\geq U_i^*(\hat{m}_i, \hat{a}_i = 1 | \theta_i, s_i, \hat{m}_j, \tilde{\theta}_i) \Leftrightarrow 0 \geq 2E[W | s_i, s_j = 1] - c + \mu \\ &\Leftrightarrow \mu \leq c - 1, \end{aligned} \quad (39)$$

which is the desired upper bound on  $\mu$ .

If  $\hat{m}_j = 0$  or  $\hat{m}_i = 0$ , then not contributing yields some positive image  $\tilde{\theta}_i(\theta_j, s_j, \hat{m}_i, 0) > 0$  depending on the type and signal of the other agent, while contributing is off-equilibrium. As the low type never contributes,  $D^*(0, s_i, \hat{m}_i, 1, \hat{m}_j) = \emptyset$  for all  $s_i$ . Hence, if the high type

benefits from contributing for some  $s_i$  and beliefs of the other agent,  $D^*(1, s_i, \hat{m}_i, 1, \hat{m}_j) \neq \emptyset$ , then  $\Theta^*(\hat{m}_i, 1, \hat{m}_j) = \{1\}$ , i.e., the criterion assigns a high image,  $\hat{\theta}_i = 1$ , to action  $\hat{a}_i = 1$  when  $\hat{m}_j = 0$  or  $\hat{m}_i = 0$  in this case. Therefore, let  $\tilde{\theta}_i$  denote any consistent belief system such that  $\tilde{\theta}_i(\theta_j, s_j, \hat{m}_i, 1) \equiv 1$  and consider the case  $\hat{m}_j = 0$  and  $\hat{m}_i = 1$ ; notice that contributing and not contributing both yields a high image. Take a high type with  $s_i = 0$ . Agent  $i$  does not have incentives to take action  $\hat{a}_i = 1$  iff

$$\begin{aligned} U_i^*(\hat{m}_i, \hat{a}_i = 0 | \theta_i, s_i, \hat{m}_j, \tilde{\theta}_i) &\geq U_i^*(\hat{m}_i, \hat{a}_i = 1 | \theta_i, s_i, \hat{m}_j, \tilde{\theta}_i) \\ \Leftrightarrow \mu &\geq Pr(s_j = 0 | s_i) 2E[W | s_i, s_j = 0] + Pr(s_j = 1 | s_i) 2E[W | s_i, s_j = 1] - c + \mu \\ \Leftrightarrow c &\geq 2/3, \end{aligned}$$

which is implied by (39). Take a high type with  $s_i = 1$ . Agent  $i$  does not have incentives to take action  $\hat{a}_i = 1$  iff

$$\begin{aligned} U_i^*(\hat{m}_i, \hat{a}_i = 0 | \theta_i, s_i, \hat{m}_j, \tilde{\theta}_i) &\geq U_i^*(\hat{m}_i, \hat{a}_i = 1 | \theta_i, s_i, \hat{m}_j, \tilde{\theta}_i) \\ \Leftrightarrow \mu &\geq Pr(s_j = 0 | s_i) 2E[W | s_i, s_j = 0] + Pr(s_j = 1 | s_i) 2E[W | s_i, s_j = 1] - c + \mu \\ \Leftrightarrow c &\geq 4/3, \end{aligned} \tag{40}$$

which will later be shown to be implied by another condition. This shows that always contributing after submitting a high report is not optimal.

Next, we consider the first stage. Take a high type with  $s_i = 0$ . Notice that deviations that involve contributing independently of  $\hat{m}_j$  yield a high image regardless of  $\hat{m}_i$ . Hence, as  $\hat{m}_i = 1$  yields a higher likelihood of a contribution by the other player, strategy  $(0, 1)$  is strictly dominated by strategy  $(1, 1)$ . Furthermore, a deviation to the latter strategy is ruled out by the second stage (case  $\hat{m}_j = 0$  and  $\hat{m}_i = 1$ ). Similarly, deviations that involve contributing if  $\hat{m}_j = 1$  yield a weakly higher image and a higher likelihood of a contribution by the other player if  $\hat{m}_i = 1$ . Hence, strategy  $(0, \mathbf{1}_{\{\hat{m}_j=1\}})$  is strictly dominated by strategy  $(1, \mathbf{1}_{\{\hat{m}_j=1\}})$ . Moreover, also a deviation to the latter strategy is ruled out by the second stage (case  $\hat{m}_j = 1$  and  $\hat{m}_i = 1$ ). Hence, the only deviation left is overreporting and never contributing, i.e., a deviation to the strategy  $(m'_i, a'_i) = (1, 0)$ . This strategy yields a high image if  $\hat{m}_j = 0$  and a low image if  $\hat{m}_j = 1$  by the DD1-criterion. Therefore, let  $\tilde{\theta}_i$  denote any consistent belief system such that  $\tilde{\theta}_i(\theta_j, s_j, 1, 0) = 1$  for all  $(\theta_j, s_j) \neq (1, 1)$  and  $\tilde{\theta}_i(1, 1, 1, 0) = 0$ . Agent  $i$  does not have incentives to do this deviation iff

$$\begin{aligned} U_i^*(m'_i, a'_i | \theta_i, s_i, \tilde{\theta}_i) &\geq U_i^*(m'_i, a'_i | \theta_i, s_i, \tilde{\theta}_i) \\ \Leftrightarrow Pr(s_j = 0 | s_i) \mu E[\tilde{\theta}_i(\theta_j, 0, 0, 0)] &+ Pr(s_j = 1 | s_i) \mu E[\tilde{\theta}_i(\theta_j, 1, 0, 0)] \\ &\geq Pr(s_j = 0 | s_i) \mu + Pr(s_j = 1 | s_i) [\pi 2E[W | s_i, s_j = 1] + (1 - \pi) \mu] \\ \Leftrightarrow 2\mu E[\tilde{\theta}_i(\theta_j, 0, 0, 0)] &+ \mu E[\tilde{\theta}_i(\theta_j, 1, 0, 0)] \geq \pi + (3 - \pi) \mu. \end{aligned} \tag{41}$$

Next, notice that

$$E[\tilde{\theta}_i(\theta_j, 0, 0, 0)] = \frac{2\pi}{3 - \pi} \text{ and } E[\tilde{\theta}_i(\theta_j, 1, 0, 0)] = \frac{\pi}{3 - 2\pi}. \tag{42}$$

Substituting (42) into (41) yields

$$\mu \left( \frac{4\pi}{3-\pi} + \frac{\pi}{3-2\pi} \right) \geq \pi + (3-\pi)\mu \Leftrightarrow \mu \frac{\pi(2\pi^2 - 24\pi + 51) - 27}{(3-\pi)(3-2\pi)} \geq \pi. \quad (43)$$

For (43) to hold it is necessary that

$$B(\pi) \equiv \pi(2\pi^2 - 24\pi + 51) - 27 > 0, \quad (44)$$

which is the desired lower bound on  $\pi$ .<sup>14</sup> Given (44) holds, we obtain

$$\mu \geq \frac{\pi(3-\pi)(3-2\pi)}{B(\pi)}, \quad (45)$$

which is the first desired lower bound on  $\mu$ . Notice that the right-hand-side of (45) is strictly decreasing in  $\pi$  on the range where (44) holds and attains 1 at  $\pi = 1$ , which implies  $\mu \geq 1$ . Together with (39), we obtain  $c \geq 2$ , which implies (40) as claimed above.

If  $s_i = 1$ , then the strategies  $(0, 1)$  and  $(0, \mathbf{1}_{\{\hat{m}_j=1\}})$  are again strictly dominated. Furthermore, also a deviation to the strategies  $(1, 1)$  (case  $\hat{m}_j = 0$  and  $\hat{m}_i = 1$ ) and  $(1, 0)$  (case  $\hat{m}_j = 1$  and  $\hat{m}_i = 1$ ) is ruled out by the second stage. Hence, the only deviation left is underreporting and never contributing, i.e., a deviation to the strategy  $(m'_i, a'_i) = (0, 0)$ . This strategy yields an intermediate image of  $\tilde{\theta}_i(\theta_j, s_j, 0, 0)$  depending on the type and signal of the other agent. Agent  $i$  does not have incentives to do this deviation iff

$$\begin{aligned} U_i^*(m_i^*, a_i^* | \theta_i, s_i, \tilde{\theta}_i) &\geq U_i^*(m'_i, a'_i | \theta_i, s_i, \tilde{\theta}_i) \\ \Leftrightarrow Pr(s_j = 0 | s_i)\mu + Pr(s_j = 1 | s_i)[\pi(4E[W | s_i, s_j = 1] - c + \mu) + (1-\pi)\mu] \\ &\geq Pr(s_j = 0 | s_i)\mu E[\tilde{\theta}_i(\theta_j, 0, 0, 0)] + Pr(s_j = 1 | s_i)\mu E[\tilde{\theta}_i(\theta_j, 1, 0, 0)] \\ \Leftrightarrow 3\mu + 2\pi(3-c) &\geq \mu E[\tilde{\theta}_i(\theta_j, 0, 0, 0)] + 2\mu E[\tilde{\theta}_i(\theta_j, 1, 0, 0)]. \end{aligned} \quad (46)$$

Substituting (42) into (46) yields

$$\begin{aligned} 3\mu + 2\pi(3-c) \geq \mu \left( \frac{2\pi}{3-\pi} + \frac{2\pi}{3-2\pi} \right) &\Leftrightarrow 2\pi(3-c) \geq -\mu \frac{3(1-\pi)(9-4\pi)}{(3-\pi)(3-2\pi)} \\ &\Leftrightarrow \mu \geq \frac{2\pi(c-3)(3-\pi)(3-2\pi)}{3(1-\pi)(9-4\pi)}, \end{aligned} \quad (47)$$

which is the second desired lower bound on  $\mu$ . Finally, consider a low type. As the low type never contributes, the only possible deviation is submitting a high report but not contributing,  $(m'_i, a'_i) = (1, 0)$ . Recall that this strategy yields a high image if  $\hat{m}_j = 0$  and a low image if  $\hat{m}_j = 1$  by the DD1-criterion. Therefore, let  $\tilde{\theta}_i$  denote any consistent belief system such that  $\tilde{\theta}_i(\theta_j, s_j, 1, 0) = 1$  for all  $(\theta_j, s_j) \neq (1, 1)$  and  $\tilde{\theta}_i(1, 1, 1, 0) = 0$ . If  $s_i = 0$ , then agent  $i$  does not

<sup>14</sup>Notice that the corresponding cubic equation  $B(\pi) = 0$  has exactly one real solution on  $(0, 1)$ , which is irrational and slightly larger than  $4/5$ . In particular, (44) holds if  $\pi$  exceeds this value.

have incentives to do this deviation iff

$$\begin{aligned}
& U_i^*(m_i^*, a_i^* | \theta_i, s_i, \tilde{\theta}_i) \geq U_i^*(m_i', a_i' | \theta_i, s_i, \tilde{\theta}_i) \\
& \Leftrightarrow Pr(s_j = 0 | s_i) \mu E[\tilde{\theta}_i(\theta_j, 0, 0, 0)] + Pr(s_j = 1 | s_i) \mu E[\tilde{\theta}_i(\theta_j, 1, 0, 0)] \\
& \quad \geq Pr(s_j = 0 | s_i) \mu + Pr(s_j = 1 | s_i) [\pi E[W | s_i, s_j = 1] + (1 - \pi) \mu] \\
& \Leftrightarrow 2\mu E[\tilde{\theta}_i(\theta_j, 0, 0, 0)] + \mu E[\tilde{\theta}_i(\theta_j, 1, 0, 0)] \geq \pi/2 + (3 - \pi)\mu. \tag{48}
\end{aligned}$$

Substituting (42) into (48) yields

$$\mu \left( \frac{4\pi}{3 - \pi} + \frac{\pi}{3 - 2\pi} \right) \geq \pi/2 + (3 - \pi)\mu \Leftrightarrow \mu \frac{\pi(2\pi^2 - 24\pi + 51) - 27}{(3 - \pi)(3 - 2\pi)} \geq \pi/2,$$

which is implied by (44) and (45). If  $s_i = 1$ , then agent  $i$  does not have incentives to do this deviation iff

$$\begin{aligned}
& U_i^*(m_i^*, a_i^* | \theta_i, s_i, \tilde{\theta}_i) \geq U_i^*(m_i', a_i' | \theta_i, s_i, \tilde{\theta}_i) \\
& \Leftrightarrow Pr(s_j = 0 | s_i) \mu E[\tilde{\theta}_i(\theta_j, 0, 0, 0)] + Pr(s_j = 1 | s_i) \mu E[\tilde{\theta}_i(\theta_j, 1, 0, 0)] \\
& \quad \geq Pr(s_j = 0 | s_i) \mu + Pr(s_j = 1 | s_i) [\pi E[W | s_i, s_j = 1] + (1 - \pi) \mu] \\
& \Leftrightarrow \mu E[\tilde{\theta}_i(\theta_j, 0, 0, 0)] + 2\mu E[\tilde{\theta}_i(\theta_j, 1, 0, 0)] \geq 3\pi/2 + (3 - 2\pi)\mu. \tag{49}
\end{aligned}$$

Substituting (42) into (49) yields

$$\mu \left( \frac{2\pi}{3 - \pi} + \frac{2\pi}{3 - 2\pi} \right) \geq 3\pi/2 + (3 - 2\pi)\mu \Leftrightarrow \mu \frac{\pi(4\pi^2 - 30\pi + 57) - 27}{(3 - \pi)(3 - 2\pi)} \geq 3\pi/2, \tag{50}$$

which is also implied by (44) and (45).<sup>15</sup> Hence, we have established that no type-signal pair has incentives to deviate if (39), (44), (45) and (47) hold. Notice also that if one of these four conditions does not hold, this claim is no longer valid. Accounting for the bounds from Assumption 1 finishes the proof.

## A.6 Proof of Proposition 5

We first determine welfare of strategy profiles in  $\mathcal{S}^*$ .

**Lemma 1.** *The strategy profile in which*

- (i) *there is no information transmission (both types are truthful) and the high type never contributes yields welfare  $2\pi\mu$ .*
- (ii) *there is no information transmission and the high type contributes conditional on a high signal yields welfare  $2\pi(1 + \pi/3 - c/2 + \mu)$ .*
- (iii) *both types are truthful and the high type contributes conditional on a high signal or a high report yields welfare  $2\pi(5(3 + \pi)/12 - 2c/3 + \mu)$ .*
- (iv) *both types are truthful and the high type contributes conditional on a high signal and a high report yields welfare  $2\pi((3 + \pi)/4 - c/3 + \mu)$ .*

<sup>15</sup>To see why, notice that  $\pi(4\pi^2 - 30\pi + 57) - 27 \geq 3(\pi(2\pi^2 - 24\pi + 51) - 27)/2 > 0$  for all  $\pi$  on the range where (44) holds. Combined with (45), this yields (50).

(v) the low type is a denialist and the high type is truthful and contributes conditional on a high signal or a high report yields welfare  $2\pi(1 + 2\pi/3 - (3 + \pi)c/6 + \mu)$ .

(vi) the low type is a denialist and the high type is truthful and contributes conditional on a high signal and a high report yields welfare  $2\pi(\pi - \pi c/3 + \mu)$ .

In particular, the ex-ante expected total image in society is equal to  $2\pi$ , independent of the agents' strategies.

*Proof.* (i) As both types employ the same communication strategy, their ex-ante expected utility is  $\mu\pi$ , which yields welfare  $2(\pi\mu\pi + (1 - \pi)\mu\pi) = 2\pi\mu$ .

(ii) The ex-ante expected utility of a high type is

$$\begin{aligned} & Pr(s_i = 1) (2E[W|s_i = 1] - c + \mu + Pr(s_j = 1|s_i = 1)\pi 2E[W|s_i = 1, s_j = 1]) \\ & + Pr(s_i = 0) \left( Pr(s_j = 1|s_i = 0) (\pi 2E[W|s_i = 0, s_j = 1] + \mu E[\tilde{\theta}_i(\theta_j, 1, 1, 0)]) \right. \\ & \left. + Pr(s_j = 0|s_i = 0) \mu E[\tilde{\theta}_i(\theta_j, 0, 1, 0)] \right) \\ & = \frac{1}{2} \left( \frac{4}{3}(1 + \pi) - c + \mu \left( 1 + \frac{1}{3}E[\tilde{\theta}_i(\theta_j, 1, 1, 0)] + \frac{2}{3}E[\tilde{\theta}_i(\theta_j, 0, 1, 0)] \right) \right). \end{aligned} \quad (51)$$

Notice that

$$E[\tilde{\theta}_i(\theta_j, 0, 1, 0)] = \frac{2\pi}{3 - \pi} \text{ and } E[\tilde{\theta}_i(\theta_j, 1, 1, 0)] = \frac{\pi}{3 - 2\pi}. \quad (52)$$

Substituting (52) into (51) yields

$$\frac{1}{2} \left( \frac{4}{3}(1 + \pi) - c + \frac{\mu}{3} \left( 3 + \frac{\pi}{3 - 2\pi} + \frac{4\pi}{3 - \pi} \right) \right) = \frac{1}{2} \left( \frac{4}{3}(1 + \pi) - c - \mu \frac{\pi^2 + 4\pi - 9}{(3 - \pi)(3 - 2\pi)} \right).$$

The ex-ante expected utility of a low type is

$$\begin{aligned} & Pr(s_i = 1) \left( Pr(s_j = 1|s_i = 1) (\pi E[W|s_i = 1, s_j = 1] + \mu E[\tilde{\theta}_i(\theta_j, 1, 1, 0)]) \right. \\ & \left. + Pr(s_j = 0|s_i = 1) \mu E[\tilde{\theta}_i(\theta_j, 0, 1, 0)] \right) \\ & + Pr(s_i = 0) \left( Pr(s_j = 1|s_i = 0) (\pi E[W|s_i = 0, s_j = 1] + \mu E[\tilde{\theta}_i(\theta_j, 1, 1, 0)]) \right. \\ & \left. + Pr(s_j = 0|s_i = 0) \mu E[\tilde{\theta}_i(\theta_j, 0, 1, 0)] \right) \\ & = \frac{1}{2} \left( \frac{2}{3}\pi + \mu \left( E[\tilde{\theta}_i(\theta_j, 1, 1, 0)] + E[\tilde{\theta}_i(\theta_j, 0, 1, 0)] \right) \right). \end{aligned} \quad (53)$$

Substituting (52) into (53) yields

$$\frac{1}{2} \left( \frac{2}{3}\pi + \mu \left( \frac{\pi}{3 - 2\pi} + \frac{2\pi}{3 - \pi} \right) \right) = \frac{1}{2} \left( \frac{2}{3}\pi + \mu \frac{\pi(9 - 5\pi)}{(3 - \pi)(3 - 2\pi)} \right).$$

Together, this yields welfare

$$\begin{aligned} & \pi \left( \frac{4}{3}(1 + \pi) - c - \mu \frac{\pi^2 + 4\pi - 9}{(3 - \pi)(3 - 2\pi)} \right) + (1 - \pi) \left( \frac{2}{3}\pi + \mu \frac{\pi(9 - 5\pi)}{(3 - \pi)(3 - 2\pi)} \right) \\ & = 2\pi \left( 1 + \frac{\pi}{3} - \frac{c}{2} + \mu \right). \end{aligned}$$

(iii) The ex-ante expected utility of a high type is

$$\begin{aligned} & Pr(s_i = 1) (2(1 + \pi)E[W|s_i = 1] - c + \mu) + Pr(s_i = 0) \left( Pr(s_j = 0|s_i = 0)\mu\pi \right. \\ & \quad \left. + Pr(s_j = 1|s_i = 0)(2(1 + \pi)E[W|s_i = 0, s_j = 1] - c + \mu) \right) \\ &= \frac{1}{2} \left( (1 + \pi)\frac{5}{3} + \frac{2}{3}\mu\pi - \frac{4}{3}(c - \mu) \right). \end{aligned}$$

The ex-ante expected utility of a low type is

$$\begin{aligned} & Pr(s_i = 1)\pi E[W|s_i = 1] + Pr(s_i = 0) \left( Pr(s_j = 1|s_i = 0)\pi E[W|s_i = 0, s_j = 1] \right. \\ & \quad \left. + Pr(s_j = 0|s_i = 0)\mu\pi \right) \\ &= \frac{1}{2} \left( \frac{5}{6}\pi + \frac{2}{3}\mu\pi \right). \end{aligned}$$

Together, this yields welfare

$$\pi \left( (1 + \pi)\frac{5}{3} + \frac{2}{3}\mu\pi - \frac{4}{3}(c - \mu) \right) + (1 - \pi) \left( \frac{5}{6}\pi + \frac{2}{3}\mu\pi \right) = 2\pi \left( (3 + \pi)\frac{5}{12} - \frac{2}{3}c + \mu \right).$$

(iv) The ex-ante expected utility of a high type is

$$\begin{aligned} & Pr(s_i = 1) \left( Pr(s_j = 1|s_i = 1)(2(1 + \pi)E[W|s_i = 1, s_j = 1] - c + \mu) \right. \\ & \quad \left. + Pr(s_j = 0|s_i = 1)\mu\pi \right) + Pr(s_i = 0)\mu\pi \\ &= \frac{1}{2} \left( 1 + \pi - \frac{2}{3}(c - \mu) + \frac{4}{3}\mu\pi \right). \end{aligned}$$

The ex-ante expected utility of a low type is

$$\begin{aligned} & Pr(s_i = 1) \left( Pr(s_j = 1|s_i = 1)\pi E[W|s_i = 1, s_j = 1] + Pr(s_j = 0|s_i = 1)\mu\pi \right) \\ & \quad + Pr(s_i = 0)\mu\pi \\ &= \frac{1}{2} \left( \frac{\pi}{2} + \frac{4}{3}\mu\pi \right). \end{aligned}$$

Together, this yields welfare

$$\pi \left( 1 + \pi - \frac{2}{3}(c - \mu) + \frac{4}{3}\mu\pi \right) + (1 - \pi) \left( \frac{\pi}{2} + \frac{4}{3}\mu\pi \right) = 2\pi \left( \frac{3 + \pi}{4} - \frac{c}{3} + \mu \right).$$

(v) The ex-ante expected utility of a high type is

$$\begin{aligned} & Pr(s_i = 1) (2(1 + \pi)E[W|s_i = 1] - c + \mu) + Pr(s_i = 0) \left( Pr(s_j = 1|s_i = 0) \cdot \right. \\ & \quad \left( \pi(4E[W|s_i = 0, s_j = 1] - c + \mu) + (1 - \pi)\mu E[\tilde{\theta}_i(\theta_j, 1, 0, 0)] \right) \right. \\ & \quad \left. + Pr(s_j = 0|s_i = 0)\mu E[\tilde{\theta}_i(\theta_j, 0, 0, 0)] \right) \\ &= \frac{1}{2} \left( \frac{4}{3} + 2\pi - \frac{3 + \pi}{3}c + \mu \left( \frac{3 + \pi}{3} + \frac{1 - \pi}{3}E[\tilde{\theta}_i(\theta_j, 1, 0, 0)] + \frac{2}{3}E[\tilde{\theta}_i(\theta_j, 0, 0, 0)] \right) \right). \quad (54) \end{aligned}$$

Next, notice that

$$E[\tilde{\theta}_i(\theta_j, 0, 0, 0)] = \frac{2\pi}{3-\pi} \text{ and } E[\tilde{\theta}_i(\theta_j, 1, 0, 0)] = \frac{\pi}{3-2\pi}. \quad (55)$$

Substituting (55) into (54) yields

$$\begin{aligned} & \frac{1}{2} \left( \frac{4}{3} + 2\pi - \frac{3+\pi}{3}c + \frac{\mu}{3} \left( 3 + \pi + \frac{\pi(1-\pi)}{3-2\pi} + \frac{4\pi}{3-\pi} \right) \right) \\ &= \frac{1}{2} \left( \frac{4}{3} + 2\pi - \frac{3+\pi}{3}c + \mu \frac{\pi^3 - 5\pi^2 - \pi + 9}{(3-2\pi)(3-\pi)} \right). \end{aligned}$$

The ex-ante expected utility of a low type is

$$\begin{aligned} & Pr(s_i = 1) \left( Pr(s_j = 1|s_i = 1)(\pi E[W|s_i = 1, s_j = 1] + (1-\pi)\mu E[\tilde{\theta}_i(\theta_j, 1, 0, 0)]) \right. \\ & \quad \left. + Pr(s_j = 0|s_i = 1)\mu E[\tilde{\theta}_i(\theta_j, 0, 0, 0)] \right) + Pr(s_i = 0) \left( Pr(s_j = 1|s_i = 0) \right. \\ & \quad \left. (\pi E[W|s_i = 0, s_j = 1] + (1-\pi)\mu E[\tilde{\theta}_i(\theta_j, 1, 0, 0)]) \right. \\ & \quad \left. + Pr(s_j = 0|s_i = 0)\mu E[\tilde{\theta}_i(\theta_j, 0, 0, 0)] \right) \\ &= \frac{1}{2} \left( \frac{2}{3}\pi + \mu \left( (1-\pi)E[\tilde{\theta}_i(\theta_j, 1, 0, 0)] + E[\tilde{\theta}_i(\theta_j, 0, 0, 0)] \right) \right). \quad (56) \end{aligned}$$

Substituting (55) into (56) yields

$$\frac{1}{2} \left( \frac{2}{3}\pi + \mu \left( \frac{\pi(1-\pi)}{3-2\pi} + \frac{2\pi}{3-\pi} \right) \right) = \frac{1}{2} \left( \frac{2}{3}\pi + \mu \frac{\pi(\pi^2 - 8\pi + 9)}{(3-2\pi)(3-\pi)} \right).$$

Together, this yields welfare

$$\begin{aligned} & \pi \left( \frac{4}{3} + 2\pi - \frac{3+\pi}{3}c + \mu \frac{\pi^3 - 5\pi^2 - \pi + 9}{(3-2\pi)(3-\pi)} \right) + (1-\pi) \left( \frac{2}{3}\pi + \mu \frac{\pi(\pi^2 - 8\pi + 9)}{(3-2\pi)(3-\pi)} \right) \\ &= 2\pi \left( 1 + \frac{2}{3}\pi - \frac{3+\pi}{6}c + \mu \right). \end{aligned}$$

(vi) The ex-ante expected utility of a high type is

$$\begin{aligned} & Pr(s_i = 1) \left( Pr(s_j = 1|s_i = 1)\pi(4E[W|s_i = 1, s_j = 1] - c) + \mu \right) \\ & \quad + Pr(s_i = 0)\mu \left( Pr(s_j = 1|s_i = 0)E[\tilde{\theta}_i(\theta_j, 1, 0, 0)] + Pr(s_j = 0|s_i = 0)E[\tilde{\theta}_i(\theta_j, 0, 0, 0)] \right) \\ &= \frac{1}{2} \left( \frac{2}{3}\pi(3-c) + \mu \left( 1 + \frac{1}{3}E[\tilde{\theta}_i(\theta_j, 1, 0, 0)] + \frac{2}{3}E[\tilde{\theta}_i(\theta_j, 0, 0, 0)] \right) \right). \quad (57) \end{aligned}$$

Substituting (55) into (57) yields

$$\frac{1}{2} \left( \frac{2}{3}\pi(3-c) + \frac{\mu}{3} \left( 3 + \frac{\pi}{3-2\pi} + \frac{4\pi}{3-\pi} \right) \right) = \frac{1}{2} \left( \frac{2}{3}\pi(3-c) - \mu \frac{\pi^2 + 4\pi - 9}{(3-2\pi)(3-\pi)} \right).$$



The ex-ante expected utility of a low type is

$$\begin{aligned}
& Pr(s_i = 1)\mu(Pr(s_j = 1|s_i = 1)E[\tilde{\theta}_i(\theta_j, 1, 0, 0)] + Pr(s_j = 0|s_i = 1)E[\tilde{\theta}_i(\theta_j, 0, 0, 0)]) \\
& + Pr(s_i = 0)\mu(Pr(s_j = 1|s_i = 0)E[\tilde{\theta}_i(\theta_j, 1, 0, 0)] + Pr(s_j = 0|s_i = 0)E[\tilde{\theta}_i(\theta_j, 0, 0, 0)]) \\
& = \frac{1}{2}\mu(E[\tilde{\theta}_i(\theta_j, 1, 0, 0)] + E[\tilde{\theta}_i(\theta_j, 0, 0, 0)]). \tag{58}
\end{aligned}$$

Substituting (55) into (58) yields

$$\frac{1}{2}\mu \left( \frac{\pi}{3-2\pi} + \frac{2\pi}{3-\pi} \right) = \frac{1}{2}\mu \frac{\pi(9-5\pi)}{(3-2\pi)(3-\pi)}.$$

Together, this yields welfare

$$\pi \left( \frac{2}{3}\pi(3-c) - \mu \frac{\pi^2 + 4\pi - 9}{(3-2\pi)(3-\pi)} \right) + (1-\pi)\mu \frac{\pi(9-5\pi)}{(3-2\pi)(3-\pi)} = 2\pi \left( \pi - \frac{\pi c}{3} + \mu \right).$$

Finally, observe that the ex-ante expected total image of strategy profiles in  $\mathcal{S}^*$  is constant at  $2\pi$ , which finishes the proof.  $\square$

In the following, we identify the equilibria in  $\mathcal{S}^*$  with the respective parts of Lemma 1. First, notice that (v) yields higher welfare than (ii) iff

$$2\pi(1 + 2\pi/3 - (3 + \pi)c/6 + \mu) > 2\pi(1 + \pi/3 - c/2 + \mu) \Leftrightarrow c < 2$$

and (iv) yields higher welfare than (ii) iff

$$2\pi((3 + \pi)/4 - c/3 + \mu) > 2\pi(1 + \pi/3 - c/2 + \mu) \Leftrightarrow c > (3 + \pi)/2,$$

i.e., (ii) never yields the highest welfare. Second, (iii) yields at least as high welfare than (v) iff

$$2\pi(5(3 + \pi)/12 - 2c/3 + \mu) \geq 2\pi(1 + 2\pi/3 - (3 + \pi)c/6 + \mu) \Leftrightarrow c \leq 3/2.$$

Third, (v) yields at least as high welfare than (iv) iff

$$2\pi(1 + 2\pi/3 - (3 + \pi)c/6 + \mu) \geq 2\pi((3 + \pi)/4 - c/3 + \mu) \Leftrightarrow c \leq (3 + 5\pi)/(2 + 2\pi).$$

Fourth, (iv) yields at least as high welfare than (vi) iff

$$2\pi((3 + \pi)/4 - c/3 + \mu) \geq 2\pi(\pi - \pi c/3 + \mu) \Leftrightarrow c \leq 9/4.$$

Finally, (vi) yields at least as high welfare than (i) iff

$$2\pi(\pi - \pi c/3 + \mu) \geq 2\pi\mu \Leftrightarrow c \leq 3,$$

which finishes the proof as the thresholds are strictly increasing.

## A.7 Proof of Corollary 1

Suppose  $27/20 < c < 39/20$ . Recall that in Section 4 we have identified equilibrium candidate strategy profiles to be such that either

- (i) there is no information transmission and the high type never contributes,
- (ii) there is no information transmission and the high type contributes conditional on a high signal,
- (iii) the low type is a denialist and the high type is truthful and contributes conditional on a high signal or a high report, or
- (iv) the low type is a denialist and the high type is truthful and contributes conditional on a high signal and a high report.

We have already shown in the proof of Proposition 5 that (iii) yields higher welfare than (ii) and that (iv) yields higher welfare than (i) on this cost range. Furthermore, it follows from Lemma 1 that (iii) yields higher welfare than (iv) iff

$$2\pi(1 + 2\pi/3 - (3 + \pi)c/6 + \mu) > 2\pi(\pi - \pi c/3 + \mu) \Leftrightarrow c < 2,$$

which implies that (iii) yields the highest welfare, and by Remark 1 also the highest net-of-image-welfare, among the equilibrium candidates.

Next, fix any  $\pi \in (0, 1)$  and suppose that  $\mu < 1/2$ , which implies (Proposition 3) that (iii) is not an equilibrium. It is left to show that we can choose  $\mu' > \mu$  such that (iii) is an equilibrium under  $\mu'$  and hence net-of-image-welfare in equilibrium has increased.

Consider  $\mu' = c - 3/4 - \varepsilon > \mu$ , with  $\varepsilon > 0$  small enough. First, notice that by definition,  $c - 3/4 > \mu' > c - 1$ . Second,

$$\mu' = c - 3/4 - \varepsilon > 3/5 - \varepsilon > \frac{(3 - \pi)(3 - 2\pi)}{\pi^2 - 12\pi + 15}.$$

It left to show that the last inequality of (4) is satisfied. Using that  $c = \mu' + 3/4 + \varepsilon$ , we get

$$\begin{aligned} & \mu' - \frac{((3 - 2\pi)(c - 1) - 1)(3 - \pi)(3 - 2\pi)}{3(1 - \pi)(2\pi^2 - 10\pi + 9)} \\ &= \mu' - \frac{((3 - 2\pi)(\mu' - 1/4 + \varepsilon) - 1)(3 - \pi)(3 - 2\pi)}{3(1 - \pi)(2\pi^2 - 10\pi + 9)} \\ &= \frac{2\pi(\pi^2 - 6\pi + 6)}{3(1 - \pi)(2\pi^2 - 10\pi + 9)} \left( \underbrace{\frac{(3 - \pi)(3 - 2\pi)(7 - 2\pi - 4\varepsilon)}{8\pi(\pi^2 - 6\pi + 6)}}_{> 6/5} - \mu' \right) \\ &> 0, \end{aligned}$$

where the last inequality holds since  $\mu' = c - 3/4 - \varepsilon < 6/5 - \varepsilon$ , which finishes the proof.