Image Versus Information:
Changing Societal Norms and Optimal Privacy

S. Nageeb Ali[1]   Roland Bénabou[2]

This version: December 2015 [3]

[1] Pennsylvania State University. Email: nageeb@psa.edu.

[2] Princeton University, BREAD, CIFAR, CEPR IZA, and NBER. Email: rbenabou@princeton.edu.

[3] We are grateful for helpful comments to Jim Andreoni, Raphael Levy, Navin Kartik, Stephen Morris and Joel Sobel. We thank Edoardo Grillo and Charles Lin for excellent research assistance. Bénabou gratefully acknowledges financial support from the Canadian Institute for Advanced Research.

Abstract

We study a Principal interacting with image-conscious agents in a general public-goods provision context. Agents have private signals about the quality of the public good which, when suitably aggregated, precisely reveal its social value. Each chooses how much to contribute, based on his own mix of public-spiritedness and reputational concern for appearing prosocial. The Principal can amplify or dampen these reputational payoffs by making individual behavior more or less visible to the community. While this entails no direct cost, it entails an endogenous, informational one: because societal preferences evolve, the Principal knows only imperfectly the social value of the public good and the importance attached by agents to social esteem or sanctions. Learning about public good quality is important for choosing her own contribution, matching rate or other policy (e.g., law). If the Principal suppresses image motivations by making contributions or compliance anonymous, she can precisely infer quality from agents' aggregate behavior. However, each of them will free-ride to a greater extent, leaving her with a greater burden in achieving optimal provision. If she leverages social image to encourage prosocial behavior, on the other hand, she faces a signal-extraction problem, not knowing whether to attribute compliance to image concerns or to information. We analyze how the socially optimal degree of privacy/publicity and Principal's matching rate vary with her and agents' information structures, as well as with the variability of aggregate and individual preferences. We show in particular that in a fast-changing society (greater variability in the "fundamental" or the image-motivated component of average preferences), privacy should be greater than in a more static or traditional one, where preferences over public goods vary mostly across individuals.

Keywords: social norms, privacy, transparency, incentives, esteem, reputation, shaming punishments, conformity, societal change

JEL Classification: D62, D64, D82, H41, K42, Z13.

"If you have something that you don't want anyone to know, maybe you shouldn't be doing it in the first place."

(Google CEO Eric Schmidt, CNBC, 2009).

1 Introduction

1.1 Why Privacy?

Visibility is a powerful incentive. When people know that others will learn of their actions, they contribute more to public goods and charities, are more likely to vote, give blood or save energy. Conversely, they are less likely to lie, cheat, pollute, make offensive jokes or engage in other antisocial behaviors.[1] Compared to other incentives such as financial rewards, fines and incarceration, publicity (good or bad) is also extremely cheap. So indeed, following the implicit logic of Google's CEO (and a number of scholars), why not publicize all aspects of individuals behavior that have important external effects, leveraging the ubiquitous desire for social esteem to achieve better social outcomes?

Many public and private institutions already use esteem as a motivator, including the military, which offers medals for valor; businesses, which recognize the "employee-of-the-month"; charities publicizing donors' names on buildings and plaques. On the sanctions side, a number of U.S. states and localities use updated forms of the pillory: televised "perp walks", internet posting of the identities, photos and addresses of people convicted or even simply arrested for a host of offences (tax evasion, child support delinquency, spousal abuse, drunk driving); publishing the pictures of people and the licence plates of cars photographed in areas known for drug trafficking or prostitution; and sentencing offenders to "advertise" their deeds by means of special clothing, signs in front of their houses or purchased ads in the newspaper. While less common in other advanced countries, public shaming is on the rise there as well, as tax authorities, regulators and the public come to perceive the judicial system as unable to adequately discipline major tax evaders and rogue financiers.[2]

With advances in "big data," face recognition and other tracking technologies , the cost of widely disseminating what someone did, gave, took or even just said is rapidly falling to

[1] On public goods contributions see, e.g., Croson and Marks (1998), Ariely, Bracha, and Meier (2009), Della Vigna, List, and Lalmendier (2012) Ashraf, Bandiera, and Jack (2012), Algan, Benkler, Morell, and Hergueux (2013). On voting, see Gerber, Green, and Larimer (2008), and on blood donations Lacetera and Macis (2010).

[2] In Greece, tax authorities have released lists of major corporate and individual tax evaders. In Peru, businesses convicted of tax evasion can be shut down, with a sign plastered in front; conversely, municipalities maintain and publish an "honor list" of households who have always paid their property taxes on time (Del Carpio (2014)). In France, a recent law (July 2014) allows judges finding a firm or an individual guilty of illegal (undeclared) employment to post, for up to two years, their names and professional addresses on an internet "black list" hosted by the Ministry of Labor. Shaming can also be spontaneously organized by activists, as with the "Occupy Wall Street" movement, or the hacking of Ashley Maddison's list of user identities. There is even a growing movement of frustrated parents posting videos on the internet and social media to publicly shame their "misbehaving" children.

zero –it is in fact maintaining privacy and anonymity that is becoming increasingly expensive.[3] The trends described above are therefore likely to accentuate, whether impulsed by budget-constrained public authorities, activist groups or individual whistleblowers and "concerned citizens." A number of scholars in law, economics and philosophy have in fact long argued for a systematic recourse to public marks of honor (e.g., Brennan and Pettit (2004), Frey and Neckermann (2013)) and shame (Kahan (1996), Kahan and Posner (1999), Reeves (2013), Jacquet (2015))), on grounds of both efficiency and expressive justice.

At the same time there is also substantial unease at the idea of shaming as a policy tool, and more generally a widespread view that a society with zero privacy would be "unlivable". Besides the universal attachment to anonymous voting as a pillar of democracy, there are many other instances where social institutions preserve privacy, even though publicity could offer a powerful tool to curb free-riding and other "irresponsible" behaviors. During episodes of energy or water rationing, local authorities typically do not publish lists of overusers (the media, on the other hand, often reports on the most egregious cases). In the consumption of publicly provided or funded health care there is no policy to "out" those who impose the highest costs as the result of partially controllable behaviors such as smoking, poor diet, or addictions. On the contrary, there are strong legal protections for patient confidentiality. A general right to privacy is also enshrined in many constitutions, even if its practical content varies across places, times and judicial interpretations.

There is of course a strong case for protecting individuals' information from the eyes of parties with potentially malicious intent or conflicting interests: undemocratic government seeking to repress dissenters, firms using data about consumer's habits and spending patterns to engage in price discrimination or exploitation, hackers intent no identity theft, rivals seeking to steal trade secrets, etc. (see Acquisti, Taylor, and Wagman (2016) for a survey). While these issues are undeniably important, we focus here on identifying very different costs of transparency, related to evolving social norms and the adaptation of formal institutions,. As we shall see these imply that *even when the principal is fully benevolent,* incurs no direct cost to publicizing behaviors, and doing so always leads agents to provide more public goods, it is optimal to maintain or protect a certain degree of privacy. This remains a fortiori true under less ideal conditions.

1.2 Our Framework

The key idea is that while publicity is a powerful and cheap instrument of control, it is also a *blunt* one, generating substantial uncertainty both for those *subject to it* and, most importantly, for those who *wield it.* Our argument builds on two complementary mechanisms:

[3] A flourishing (semi-legal) image-ransoming industry is even developing in the United States These "shame entrepreneurs" operate by re-posting on high-visibility websites the official arrest "mugshots" from police departments and municipalities all across the country, then asking the people involved for a hefty fee in order to take down the post concerning them. (Segal (2013)). There are also more established companies serving businesses by "managing" their on-line reputations in consumer forums, blogs, etc.

1. *Inefficient variability.*

   The rewards and sanctions generated by publicizing an individual's actions stem from the reactions this elicits from his family, peers or neighbors. These social incentives thus involve the *emotional responses* of many people as well as their degree of *coordination,* which makes their severity hard to predict and fine-tune a priori (Posner (2009)). Depending on place, time, group, offense and individual contingencies, the feared response may go from mild ostracism to mob action, be easy or hard to escape, etc.[4] Variability in the strength of agents' concerns about social image and sanctions will, in turn, generate inefficient variations in compliance (not reflecting true variations in social value), which become amplified as individual behavior is made more visible or salient.[5]

2. *Rigid and maladaptive public policy.*

   Public stigmatization and oppressive "community standards" are often criticized for having been extensively used to repress non-believers, mixed-race relationships, divorcees, single-mothers, adulterers, homosexuals, etc. But, of course, their purpose at the time was precisely to discourage such behaviors, widely considered immoral and socially nefarious, and accordingly also punished by the law. The real problem is that *societal preferences change* unpredictably due to technology, migration, trade, enlightenment, etc. *In order to learn* how policy –the law and other institutions, taxes and subsidies, etc.– should be adapted to recent evolutions, an imperfectly informed principal must assess societal preferences from prevailing behaviors and mores. If individuals feel too constrained by the fear of social stigma and sanctions from others, these preference shifts will remain hidden or be revealed too slowly. The result will be a rigidification and maladaptation not only in *private conduct* –excessive conformity– but also in *public policy*, doubly impacting the efficiency of resource allocation.

We model this set of issues by studying a Principal and a continuum of agents interacting in a canonical context of public-goods-provision or externalities. Agents receive private signals about the quality of the public good, and their collective information, suitably aggregated, is a precise signal of its social value. Each chooses how much to contribute, based on his own mix of public-spiritedness, information and reputational concern for appearing prosocial. The Principal can amplify or dampen these reputational payoffs, and hence total contributions, by making individual behavior more or less visible to the community. While this entails little cost (none, for simplicity), she faces an informational problem: because societal preferences change,

[4] On such instability and even multiplicity in collective-action outcomes, see Lohmann (1994) and Kuran (1997). The literal explosion of (planet-wide) shaming via social media over the last few years is a good example of this variability. In many instances, the resulting costs to the "punished" party have turned out to be wildly disproportionate (loss of job and family, suicide) to the perceived offense. Sometimes there is even a backlash, where individuals who played a key role in coordinating the shaming are themselves publicly shamed on the same media (see Ronson (2015)).

[5] Similar variance effects occur if social sanctioning involves (convex) resource costs, or if agents are risk averse. We abstract from these channels, since they would lead to very similar results as the one we focus on.

she knows only imperfectly the social value of the public good and the importance attached by agents to social esteem or sanctions. Learning about public good quality or externalities is important for choosing her own (e.g., tax-financed) contribution, matching rate or other policy, such as the law. If the Principal suppresses image motivations by making contributions or compliance anonymous, she can perfectly infer societal preferences from agents' aggregate behavior. However, each agent will then free-ride on the efforts of others to a greater extent, leaving her with a greater share of the burden in achieving the desired level of public good provision. On the other hand, if she uses social image as a tool to encourage prosocial behavior, she exacerbates her own signal-extraction problem.[6] The Principal thus faces a direct tradeoff between using *image as incentives* and gaining better *information* on societal preferences.

We analyze this tradeoff and show that the optimal degree of publicity is always bounded –equivalently, some positive level of privacy must be maintained. We then characterize its comparative-statics, as well as those of the principal's second-stage policy (contribution or matching rate) with respect to her direct cost of provision, the degree of informational heterogeneity among agents, the noisiness of both sides' signals, and the aggregate variabilities of societal preferences and reputational concerns. We show in particular that *in a fast-changing society* (greater variability in the "fundamental" or the image-motivated component of average preferences), *privacy should be greater* than in a more static or "traditional" one, where preferences over public goods vary mostly across individuals but are stable in the aggregate.

1.3 Applications

The tradeoff between publicity's incentive and information-garbling effects arises in many organizational and institutional settings.

Public good provision and charitable donations. We frame the model in terms of this classical benchmark, as the issues we analyze are central to the provision of the "right kind" of public goods in a cost-effective manner. This also facilitates comparison with previous work.

Community leaders, private philanthropists and foundations must often rely on constituents's and activists' degree of involvement to identify the social value of potential public investments, such as improvements in local schools, parks, transportation, or development projects in poorer and remote parts of the world. This is also why the practice of matching voluntary contributions is so common among donors. Publicly "recognizing" and honoring individuals' or NGO's efforts encourages contributions, but also makes it a less precise signal of the true social value of

[6] The point applies more generally to any incentive to which agents respond strongly on average (effectiveness) but to a degree that is hard to predict ex-ante and parse out ex-post (uncertainty). For the reasons discussed above, this is much more a feature of social norms and peer pressure than of monetary incentives, on which numerous tradeoffs are observable. For instance, it is arguably easier for a government to estimate a stable response of tax compliance to different auditing probabilities or evasion penalties than to posting the names of evaders on-line. If one does not subscribe to such an asymmetry between formal and informal incentives, our model can also be reinterpreted as providing one more reason (learning by the Principal) why strong incentives, of any kind, can be counterproductive.

these goods. The same tradeoff involved in celebrating "leadership" contributions (Vesterlund (2003), Andreoni (2006)), which are meant to serve as signals of worth to subsequent donors.

From social norms to formal institutions. Formal laws and institutions most often crystallize from preexisting community standards, social norms and common-law practices, which inform designers about what behaviors are generally deemed to be sources of positive or negative externalities. As mentioned earlier these change over time, sometimes quite radically and very fast. In a context where behavior is highly constrained by the fear of social stigma, assessing social preferences and shaping laws by what people do (*"descriptive norm"*) can be a very poor indicator of what they really value (*"prescriptive norm"*).

Consumer and corporate social responsibility. Firms are increasingly pressured or even explicitly shamed by activists into behaving "responsibly" on issues of environmental impact, child labor, workplace safety, treatment of animals, etc. To the extent that these reputational incentives make up for deficient regulation or Pigovian taxation they are beneficial, but at the same time they lead to strong conformity effects that make it hard for consumers and investors to know which production practices (and producers) are truly socially valuable and which ones simply reflect "greenwashing". The same applies to "green" and "fair trade" consumer goods, typically heavily advertised and often conspicuously consumed.

Political correctness. Social pressure leads people to refrain from engaging in behavior or speech considered to be "offensive" or, in other places, sacrilegious (e.g., Loury (1994), Morris (2001)). Governments, university administrations and media outlets also seek to encourage "desirable" behaviors and sanction "undesirable" ones, using publicity plus perhaps other incentives such as rules and contracts.[7] Here again, the danger is that insufficient individual privacy will prevent the institution-designer from learning what people have come to really value and think.

Agency incentives. Consider the management of a sales team in charge of a given product. Individual sales representatives are likely to be privately informed about how well suited the product is to customer needs; they also face a choice in terms of how much effort to exert in promoting it. Publicizing the sales records of each sales associate, which leads them to compete harder for status, can alleviate the moral-hazard-in-teams problem (e.g., Larkin (2011)). However, it can also deprive the firm of valuable information: seeing high sales, it may not realize that its product needs further development without which its success will be short-lived, or involves hidden risks.

Leadership. As emphasized in the literature on corporate culture, a key role of leadership in organizations, corporations, and societies is to coordinate expectations and efforts toward goals that reflect shared objectives and beliefs (see Kreps (1990), Hermalin and Katz (2006), Bolton, Brunnermeier, and Veldkamp (2013)). Our analysis highlights how a leader also faces

[7] Simialrly, a recent activist campaign in Brazil tracks down the "geotagged" locations of people who post racist comments on social media, then reposts them on giant billboards and buses in the immediate neighborhood of the source (with names and profile pictures blurred, however).

the dual challenge of using publicity to encourage agents to serve the organization's goals and values, while allowing enough dissent and contrarian behavior for her (and others) to observe how those values change over time.

Political activism. The Principal can also stand for an electorate, while agents are activists and informational lobbies exerting effort to persuade voters of the importance of some drastic reform. When the media makes their actions more visible activists are willing to take more costly steps, so publicity again provides incentives. At the same time, activism is discounted to a further extent as being "attention-seeking," and indeed may not offer much useful information.[8]

1.4 Related Literature

Our study relates to several parts of the large literature examining the impact of publicity or transparency on individual and collective decision-making.

A first strand focuses on signaling in a public-goods context.[9] Our setting builds on Bénabou and Tirole (2003, 2006) who study how incentives, whether material or social, can undermine agents' intrinsic motivation or the reputational value derived from a prosocial activity. We develop this basic framework in two important directions. First, a Principal explicitly chooses how much agents know about each other's behavior, internalizing their equilibrium responses. Second, she is imperfectly informed about the social value of the activity, generating a tradeoff between image incentives and information aggregation that is a novel feature of our model.[10] Also closely related is Daughety and Reinganum (2010), who study how making actions fully public can result in the overprovision of public goods, whereas making them fully private can result in underprovision, and determine conditions under which either one is preferable. We consider the problem of a Principal who can adjust continuously how much privacy to accord individuals, faces uncertainty about they will respond to it and, most importantly, cares about the informational content of their behavior.[11]

Transparency is also a central issue when experts, judges, or committee members have career concerns over the quality of their information (rather than their prosociality), as they may distort their advice or actions in order to appear more "competent". A first effect, working

[8] Lorentzen (2008), for instance, studies how China's government relies on public protests as a signal of local corruption. Our point is that media attention to these protests helps mitigate collective action problems, but also interferes with information transmission when activism becomes attention-seeking. Battaglini and Bénabou (2003) study a complementary problem of how multiple strategic activists seek to persuade a decisionmaker.

[9] See, e.g., Bernheim (1994)Glazer and Konrad (1996a), Bagwell and Bernheim (1996), Corneo (1997), Corneo and Jeanne (1997), Harbaugh (1998), Ellingsen and Johannesson (2008) and Andreoni and Bernheim (2009).

[10] Excessive constraints on behavior (commitment devices, monitoring with threats of punishment) can also interfere with learning (or self-learning) about agents' types, rather than with information aggregation concerning the state of the world; see, e.g., Bénabou and Tirole (2004) and Ichino and Muehlheusser (2008).

[11] Daughety and Reinganum (2010) also show that waivable privacy rights do not help reduce wasteful signaling. Bénabou and Tirole (2006, 2011) show, on the other hand, that as long as the value of image (e.g., the "going rate" to have one's name on a university or hospital building, or a sponsor's name on an event) is known by the Principal, material incentives such as tax deductibility can be adjusted to offset any reputation-motivated distortions in the level of contributions or their allocation toward more highly visible public goods. This is another reason why, in the present model, the fact the Principal does not know the exact value of image is important.

in the direction of conformity or "conservatism," arises when agents have no private knowledge of their own ability: they will then make forecasts and choices that tend to agree with (look plausible given) the Principals' prior (e.g., Prendergast (1993), Prat (2005), Bar-Isaac (2012)), or with the views expressed by more "senior" agents thought to be a priori more knowledgeable. (Ottaviani and Sørensen (2001)). When competence is a private type, on the other hand, the incentive to signal it generates "anti-conformist" or activist tendencies: agents will overreact to their private signals , excessively contradict seniors or reverse precedents , etc. Which of the two forces dominates depends on the game's information structure and may in particular vary: (i) over time, in the case of repeated decisions (Prendergast and Stole (1996)); (ii) across equilibria, when the Principal has access to a verification technology that makes her information endogenous (Levy (2005)); (ii) between a single expert and a committee that provides multiple but strategically interdependent reports (Levy (2007)); (iv) with committee members' ability to communicate privately among themselves (Visser and Swank (2007)). On the normative side, which of the two distortions –conformity or exaggeration– is worse for the Principal, and whether she prefers transparency or anonymity for the agents, depends intuitively on how her loss function weighs "getting things wrong" in the more likely states of the world versus the more rare ones (e.g., Fox and Weelden (2012), Fehrler and Hughes (2014)). In our framework, agents' incentives to signal their types increase rather than decrease conformity, and the latter has simultaneously positive (mean-contribution) and negative (excessive variance and information-garbling) effects. Another key difference is that the strength of image concerns, which is common knowledge in nearly all of the signaling literature, constitutes here one of the key sources of uncertainty.[12]

Privacy is a vast subject, so it may be useful to also state what the paper is *not* about. We do not deal here with issues linked to government snooping for political-control purposes, corporate targeted advertising and consumer exploitation, identity theft or the protection of trade secrets, which all involve principals seeking to "misuse" agents' data. Our focus is instead on what private citizens know about each other's behaviors, on the social value of privacy even when the Principals is benevolent, and more generally on how her learning problem (whatever her preferences might be) affects its optimal level. For the same reasons we abstract from concerns that public shaming amounts to cruel humiliation that negates other important societal values, such as general human dignity.[13] This is a genuine concern for extreme and personalized forms of stigmatization such as special clothing, lawn signs or parades of prisoners or adulterers, but arguably much less so (especially when compared to prison) for making judicial records uniformly accessible (e.g., of tax evasion, drunk driving, child support delinquency, spousal abuse, hate speech), and not at all for creating a public registry of taxpayers' charitable contributions (Cooter (2003)) and honoring other forms of "exemplary" compliance.

[12] Other exceptions are Bénabou and Tirole (2006) and, through a mapping from heterogenous costs of misrepresentation to heterogeneous image concerns, Fischer and Verrecchia (2000) and Frankel and Kartik (2014).

[13] See, e.g., R. Posner (1998) for such arguments and Bénabou and Tirole (2011) for an analysis of expressive law, including the case of "cruel and unusual punishments".

The paper is organized as follows. Section 2 develops the basic model. Section 3.1 solves for agents' equilibrium response to a given level of publicity (observability of one's actions by others) in the presence of reputational or social-enforcement concerns. Section 3.2 derives the Principals' optimal choices of this publicity level and her own contribution. Section 4 then analyzes their comparative statics with respect to the nature of the publics-goods problem, the variability of individual and societal preferences, and the information structures of agents and principal. Section 5 outlines further extensions and concludes. Main proofs are gathered in Appendix A, while Appendix B presents extensions of the benchmark model.

2 Model

We study the interaction between a continuum of small agents ($i \in [0,1]$) and a single large Principal (P), each of whom chooses how much to contribute (in time, effort or money) to a public good. Depending on the context, these actors may correspond to: (i) a government and its citizens; (ii) a charitable organization and potential donors; (iii) a profit-maximizing firm and workers who care to some degree about how well it is doing, whether out of pure loyalty or because they have a stake in its long-run survival.

A. Agents' Choices and Payoffs. Each agent i selects a contribution level $a_i \in \mathbb{R}$, at cost $C(a_i) \equiv a_i^2/2$. An individual's utility depends on his own contribution, from which he derives some intrinsic satisfaction (or "joy of giving"), on the total provision of the public good, which has quality or social usefulness indexed by θ, and on the reputational rewards attached to contributing. Given total private contributions $\bar{a}$ and the Principal contributing a_P, Agent i's direct (non-reputational) payoff is

$$U_i(v_i, \theta, w; a_i, \bar{a}, a_P) \equiv (v_i + \theta)\, a_i + (w + \theta)\,(\bar{a} + a_P) - C\,(a_i)\,. \tag{1}$$

The first term corresponds to his *intrinsic motivation,* which includes both an idiosyncratic component v_i and the common shift factor θ, reflecting the idea that people like to contribute more to socially valuable projects than to less useful ones.[14] Agent i's baseline valuation v_i is distributed as $N\left(\bar{v}, s_v^2\right)$ and privately known to him. The second term in (1) is the *value derived from the public good*, which we take to be similar across individuals without loss of generality. We assume $\bar{v} < w$, ensuring that intrinsic motivations alone do not solve the free-rider problem.

The quality or social value of the public good is *a priori* uncertain, with agents and the Principal starting with common prior belief that θ is distributed as $N\left(\bar{\theta}, \sigma_\theta^2\right)$. Each agent i receives a private noisy signal, $\theta_i \equiv \theta + \varepsilon_i$, in which the error is distributed as $N\left(0, s_\theta^2\right)$, independently of the signals of others. Agent i's private *type* thus consists of his motivation and information (v_i, θ_i).

[14] We model agents' preferences as separable in intrinsic motivation and quality for analytical tractability, but the basic insights are robust to relaxing this assumption; see Section 2.1. for a discussion.

Each agent cares about the inferences that others—friends, family, members of his social and economic networks— will draw about his intrinsic motivation, v_i : he wishes to appear prosocial, a good citizen rather than a free-rider, dedicated to his work, etc.[15] Another agent j observing a_i does not know to what extent it was motivated intrinsically (high v_i) or by a high signal realization (high θ_i), but he can use his own signal θ_j and the realized average contribution $\bar{a}$ to form his assessment $E[v_i|a_i, \bar{a}, \theta_j]$ of player i. Thinking ahead, Agent i uses his signal θ_i to forecast the benchmark against which he will be judged. The average *social image* that he can anticipate if he contributes $a_i = a$ is thus

$$R(a, \theta_i) \equiv E_{\bar{a},\theta_{-i}} \left[\int_0^1 E\left[v_i | a, \bar{a}, \theta_j\right] dj \;\middle|\; \theta_i \right]. \tag{2}$$

The importance of reputational concerns may vary across individuals, communities and time periods, as well as with institutional choices of how much visibility and recognition to accord individual actions. We abstract here from idiosyncratic differences in how much agents care about their these social payoffs, focusing instead on aggregate variations.[16] For instance, signaling that one values the common good is more important in settings where trust plays a significant role than where most transactions take place through impersonal markets or complete contracts. Social enforcement –punishing or shunning perceived free-riders, rewarding those seen as model citizens– also relies on mobilizing emotional reactions and achieving group coordination, both of which are likely to fluctuate over time and place.

When an Agent i has average image $R(a, \theta_i)$ he thus obtains a *net* reputational payoff of $\mu x[R(a, \theta_i) - \bar{v}]$, where $\mu \sim N(\bar{\mu}, \sigma_\mu^2)$, $\bar{\mu} > 0$, represents the baseline importance of social esteem in the group and $x \geq 0$ parametrizes the degree of visibility and memorability of individual actions, which can be exogenous or under the Principal's control. Accounting for both direct and image-based payoffs, an agent of type (v_i, θ_i) chooses a_i that solves

$$\max_{a_i \in \mathbb{R}} \{ E\left[U_i(v_i, \theta, w; a_i, \bar{a}, a_P) | \, \theta_i\right] + x\mu[R(a_i, \theta_i) - \bar{v}] - C(a_i) \}. \tag{3}$$

B. Principal's Choices and Payoffs. The Principal's ex-post payoff is a convex combination of agents' total utility and her own private benefits and costs from the overall supply of the (quality-adjusted) public good:

$$\begin{aligned} V(\bar{a}, a_P, \theta) \equiv \lambda \left[\alpha \int_0^1 (v_i + \theta)\, a_i \, di + (w + \theta)(\bar{a} + a_P) - \int_0^1 C(a_i) di \right] \\ + (1 - \lambda) \left[\eta (w + \theta)(\bar{a} + a_P) - k_P C(a_P) \right]. \end{aligned} \tag{4}$$

In the first term, $\alpha \in [0, 1]$ captures the extent to which Principal internalizes agents' intrinsic

[15] These concerns may be instrumental (appearing as a more desirable employee, mate, business partner or public official), hedonic (feeling pride rather than shame, basking in social esteem), or a combination of both.

[16] See Section 2.1 for a discussion, and Appendix B for the model's extension to heterogeneous image concerns.

"joy of giving" utility, relative to their material payoffs. As to image gains and losses, by Bayes' rule they sum to zero across all agents ($\int_0^1 R\left(a_i, \theta_i\right) di = \bar{v}$), so whether or not she internalizes them is irrelevant. In the second term, k_P is the Principal's cost of directly contributing, relative to that of agents, while $\eta \in \mathbb{R}$ represents any private benefits she may derive from the total supply of public good. It will be useful to denote

$$\varphi \equiv \lambda + (1-\lambda)\eta, \tag{5}$$

$$\omega \equiv (w+\bar{\theta})\varphi - \lambda(1-\alpha)(\bar{v}+\bar{\theta}). \tag{6}$$

The coefficient φ is the Principal's total gain per (efficiency) unit added to the total supply of public good $\bar{a} + a_P$, whatever its source. The coefficient ω is her *net expected utility* from each marginal unit of the good provided specifically by the agents, taking into account that when $\lambda > 0$ she internalizes: (i) a fraction $\lambda\alpha$ of their intrinsic satisfaction from doing so; (ii) a fraction λ of their marginal contribution cost $\int_0^1 C'(a_i)di = \bar{a}$, which absent reputational incentives they would equate to their intrinsic marginal benefit, $\bar{v} + \theta$.

Put differently, ω represents the *wedge* between the *Principal's expected value* of agents' contributions and the latter's expected willingness to contribute spontaneously. To make the problem non-trivial we shall assume that $\omega > 0$, so that, on average, the Principal does want to increase compliance.

Our formulation includes as special cases:

(a) For $\lambda = 1$, a purely benevolent, "selfless" Principal.

(a) For $\lambda = 1/2$ and $\eta = 0$, a standard social planner, who values equally agents' and her own costs of provision. (The latter could even be those incurred by the rest of society's, e.g., due to a shadow price of public funds.)

(iii) For $\lambda = 0$, a purely selfish Principal, such as profit-maximizing firm that uses both payments and image to elicit effort provision from its employees.

The Principal would like to *foster public-good contributions,* but she also seeks to *learn about* θ, so as to set her own contribution a_P efficiently. A key piece of data she can observe is the aggregate contribution or compliance rate $\bar{a}$, which embodies information about the average signal received by agents. The difficulty is that she generally uncertain about the realizations of *both* aggregate shocks, θ and μ, and thus faces a *signal-extraction problem:* to what extent does $\bar{a}$ reflect agents' (average) valuing of the public good, or just their pursuit of social esteem?

The Principal shares agents' prior about the quality of the public good and may also obtain an independent signal $\theta_P \equiv \theta + \varepsilon_P$, with error distributed as $N(0, s^2_{\theta,P})$. Her prior for the importance of image is $N(\bar{\mu}, \sigma^2_\mu)$. These beliefs incorporate all the information previously obtained the Principal, for instance by polling agents about the quality of the public good or the importance of social image.[17]

[17] This information is typically limited: polling a large population is costly (see Auriol and Gary-Bobo (2012) on the optimal number of representatives), whereas polling a small population invites strategic responses from agents who would like the Principal to contribute more (see Hummel, Morgan, and Stocken (2013)).

C. Timing. The game unfolds as follows:

1. The Principal chooses the level of observability of individual behavior, x, that will prevail among agents. Conversely, $1/x$ represents the degree of *privacy.*

2. Each agent learns his private signal about quality, θ_i, and the baseline importance of social esteem, μ, then chooses his contribution a_i.

3. The aggregate contribution $\bar{a}$ is publicly observed.

4. The Principal observes her own signal θ_P.

5. The Principal chooses her contribution a_P, and the total supply $\bar{a} + a_P$ is enjoyed by all**.**

We focus, for tractability, on Perfect Bayesian Equilibria that are differentiable and strictly monotone, meaning that: (i) $R(a, \theta_i)$ is continuously differentiable in a, (ii) aggregate compliance $\bar{a}$ is strictly increasing or decreasing in quality θ. This will also be shown to imply that an equivalent formulation of the Principal's decision problem is:

(a) Given any x, optimally choose a baseline investment level she will provide and a *matching rate* on private contributions: $a_P = \underline{a}_P(x, \theta_P) + m(x)\bar{a}$.

(b) Based on ex-ante information only, set x optimally.

2.1 Discussion of the Model

At the core of our model are two related tensions between the benefits of publicity (which, on average, improves provision of public goods and economizes on costly incentives) and the distortions it generates in agents' and the Principal's decisions:

1. An agent contributes more when his actions are publicized, even if he privately believes that the public good is not worth the cost or even socially harmful, because he worries that not doing so would reflect badly upon him.

2. A Principal who does not precisely know the extent to which agents care about social payoffs must use publicity carefully, lest it make agents' behavior excessively conformist –that is, too uncorrelated with the true quality of the public good, and too difficult to for her to learn from.

To identify these strategic forces as cleanly as possible, we have made a number of simplifying assumptions, which we discuss below.

Separability in Intrinsic Motivation and Quality The model features multidimensional signaling with a single-dimensional action space, which leads to pooling between types with favorable information θ_i about the public good and those with high intrinsic motivation v_i.

Moreover, each agent lacks information about others' signals and so cannot perfectly anticipate how they will interpret his actions. Social incentives thus involve both multidimensional signaling and higher-order uncertainty, making the general problem a complex one. Specifying agents' preferences as separable in intrinsic motivation and public-good quality allows us to keep it tractable and derive simple, closed-form equilibria. The basic tradeoff between incentives and information identified here would, however, apply even with complementarity between these dimensions.[18]

Common Social Image We assume in the main analysis that agents differ in their altruism and information but share the same value for social image, μ. This simplifying assumption –almost universal in the literature on signaling– allows us to focus on the dimensions of heterogeneity of interest here and their interactions with aggregate shocks. In Appendix B we incorporate differential image concerns (each i having his own μ_i), which introduce another source of "noise" muddling the inferences that can be drawn about someone's intrinsic motivation from observing their contribution.[19] Our main results on the Principal's motive to limit publicity extend to this setting, but we lose some of the analytical tractability of the benchmark framework.

Timing of Information and Publicity Having the Principal first set the degree of publicity and then observe her signal θ_P allows us to abstract from an "Informed Principal" problem. Were the timing reversed, her choice of x would convey information about the quality of the public good, which is a different strategic force than those of interest here.[20] The choice of publicity / privacy would then also commingle the Principal's motive to learn from agents with her incentive to signal to them.

Principal's Policy We formulate the problem as the Principal choosing her provision level a_P after agents make their decisions, but the results are identical when she commits in advance to a matching rate on private contributions. This invariance reflects the fact that each a_i has a negligible effect on the aggregate, together with the assumption (implicit in how a_P enters (1)) that agents derive intrinsic utility only from their own contribution, and not from the induced

[18] In particular, for small variations in the v_i's and θ_i's, agents' preferences can be locally linearized.

[19] This "overjustification effect" arising from heterogeneous image concerns is studied by Bénabou and Tirole (2006), but only in a context where overall publicity is x is an exogenous parameter. Another recent model allowing agents to differ in the strength of their signaling concerns is Frankel and Kartik (2014).

[20] "Papers studying an informed-principal problem in related contexts include Bénabou and Tirole (2003), Sliwka (2008), Weele (2013) and Bénabou and Tirole (2011). In the first paper, an agent learns about his own payoffs from the type of contract offered by the Principal. In the last three the Principal's choice of incentives conveys information concerning the distribution of preferences in society, which matters to individual agents due to, respectively, a taste for conformity, conditional reciprocity, or endogenous reputational payoffs of the type considered here.

matching.[21]

3 Equilibrium Behavior and Optimal Privacy

We first analyze how agents respond to a *fixed* level of publicity, given their first-order uncertainty about the quality of the public good and their higher-order uncertainty about the beliefs of others. In a second step we examine how the Principal should optimally set the level of publicity, given the induced behaviors.

3.1 How Agents Respond to Publicity

Maximizing his utility (3), each agent chooses his contribution level a_i to satisfy:

$$C'(a) = v_i + E[\theta|\theta_i] + x\mu \frac{\partial R(a, \theta_i)}{\partial a}. \tag{7}$$

This equation embodies the agent's three basic motivations: his baseline intrinsic utility from contributing, his posterior belief about the quality of the public good, and the impact of contributions on his expected image. To form his optimal estimate of θ, he combines his private signal and prior expectation according to

$$E[\theta|\theta_i] = \rho\theta_i + (1-\rho)\bar{\theta}, \tag{8}$$

where $\rho = \sigma_\theta^2 / \left(\sigma_\theta^2 + s_\theta^2\right)$ is the *signal-to-noise ratio* in his inference. We show that in any equilibrium satisfying (i) and (ii) above, $\partial R(a, \theta_i)/\partial a$ is constant, leading to a unique outcome.

Proposition 1. *The unique differentiable and strictly monotone equilibrium is linear and involves an agent of type (v_i, θ_i) choosing*

$$a_i = v_i + [\rho\theta_i + (1-\rho)\,\bar{\theta}] + \quad x\mu\xi, \tag{9}$$

$$\textit{where} \quad \rho = \frac{\sigma_\theta^2}{\sigma_\theta^2 + s_\theta^2} \quad \textit{and} \quad \xi = \frac{s_v^2}{s_v^2 + \rho^2 s_\theta^2}. \tag{10}$$

The resulting aggregate contribution (or compliance level) is

$$\bar{a} = \bar{v} + \rho\theta + (1-\rho)\bar{\theta} + x\mu\xi. \tag{11}$$

Greater intrinsic motivation and better perceived quality naturally lead agents to contribute more. As to the reputational return ξ, it corresponds to the *signal-to-noise ratio* faced by an

[21] There is no a priori "right answer" on what these preferences should be. The limited experimental evidence on the question (e.g., Harbaugh, Mayr, and Burghart (2007)) suggests that while induced contributions from some outside source do generate some intrinsic satisfaction, it is markedly less than that associated to own contributions.

observer when trying to infer someone's type v_i from their action, knowing that behavior reflects private preferences, private signals and image concerns according to (9).

To better understand the underlying mechanism, note that once agents have observed $\bar{a}$ they can *retrieve the true* θ from (11), since they also know μ.[22] When an agent forms his eventual image of another, he can therefore use the actual θ rather than having to rely on his own noisy signal. In equilibrium, every observer will thus judge a given individual similarly: $E\left[v_i \mid a_i, \bar{a}, \theta_j\right]$ is in fact independent of θ_j. Furthermore, given that i is known to follow the decision rule (9), the only source of attribution error in inferring his motivation v_i from his behavior a_i is the idiosyncratic variation in the private signal θ_i he will have received. Put differently, when a_i is *judged against the benchmark* $\bar{a}$, contributions above average (say) must reflect a better than average preference, or signal, or some of both:

$$a_i - \bar{a} = v_i - \bar{v} + \rho\left(\theta_i - \theta\right). \tag{12}$$

Observers assign to each source of variation a weight proportional to its relative variance, *conditional on* θ (or $\bar{a}$), so that:

$$E\left[v_i \mid a_i, \bar{a}\right] = \left(1-\xi\right)\bar{v} + \xi\left(\bar{v} + a_i - \bar{a}\right) = \bar{v} + \xi\left(a_i - \bar{a}\right), \tag{13}$$

where ξ is given by (10). Consequently $\partial E\left[v_i \,|\, a_i, \bar{a}\right]/\partial a_i = \xi$ measures the marginal improvement in social image that additional contributions will buy.[23]

Comparative statics. Proposition 1 also identifies how equilibrium reputations and behavior vary with both idiosyncratic and aggregate variability in agents' preferences, as well as with the quality of their information. Indeed, we can re-write the marginal image return ξ as

$$\frac{\xi}{1-\xi} = s_v^2\left(\frac{1}{s_\theta} + \frac{s_\theta}{\sigma_\theta^2}\right)^2. \tag{14}$$

Proposition 2. *Reputational incentives and equilibrium contributions are increasing in* s_v^2*, decreasing in* σ_θ^2 *and U-shaped in* s_θ^2*.*[24]

The first two properties are quite intuitive. First, signaling motives are amplified by a greater cross-sectional dispersion s_v^2 in the preferences v_i that observers are trying to infer. Second, decreasing the variance σ_θ^2 of the aggregate shock means that each agent is less responsive to his private information θ_i (as it is more likely to be noise), so individual variations in contribution are again more indicative of differences in intrinsic motivation.

The third comparative static is the most novel: the U-shape in s_θ^2 reflects the idea that reputational effects are strongest when agents expect to agree *at the interim stage* about the

[22] This is where restriction (ii), focusing attention on equilibria where $\bar{a}$ is strictly monotone in θ, is used.

[23] Equation (9) also shows that visibility leads all agents to raise their contributions by the same amount. The source of this invariance is the specification of reputational payoffs as linear in image and independent of type.

[24] Throughout the paper we use the following mnemonics: cross-sectional dispersions are denoted as $s_\cdot^2$, aggregate variabilities as $\sigma_\cdot^2$

quality of the public good. This occurs when their private signals are either very precise ($s_\theta \to 0$) and hence all close to the true θ, or on the contrary very imprecise ($s_\theta \to \infty$), leading them to put a weight close to 1 on the common prior $\bar{\theta}$. In both cases, differences in contributions reflect mostly differences in intrinsic motivation, which intensifies the signaling game and thereby raises contributions.

As $\xi \to 1$ the equilibrium becomes fully revealing, with each agent's social image exactly matching his actual preference: $E\left[v_i \mid a_i\right] = v_i$. Yet everyone's contribution exceeds by $x\mu$ that which he would make, were his type directly observable: the contest for status traps everyone in an expectations game where they cannot afford to contribute less than the equilibrium level.

3.2 Optimal Publicity and Matching Policies

The Principal wants to encourage private provision of the public good but also learn about θ, so as to make efficient decisions. She does not have access to price incentives, but can stimulate contributions (or induce compliance) by publicizing everyone's behavior, thus leveraging their desire for social approval.[25] We model this degree of public visibility and memorability of agents' actions as a parameter $x \in \mathbb{R}_+$ that scales reputational payoffs up or down to $x\mu R\left(a, \theta_i\right)$, as in Bénabou and Tirole (2006). In order to highlight the tradeoffs inherent to publicity –or, conversely, the *social value of privacy* arising endogenously from agents' behavior – we assume that the Principal can vary x costlessly. While the costs of honorific ceremonies, medals, public shame lists, etc., are non-zero, they are trivially small compared to direct spending on public goods, subsidies or the legal enforcement of prohibitions.[26]

To separate the three distinct motivations for the Principal to grant agents some degree of privacy, we consider in turn:

(a) A simple benchmark in which there is no variability in their image motive, $\sigma_\mu^2 = 0$.

(b) The case where $\sigma_\mu^2 > 0$ but the Principal, like the agents, observes the realization of μ, once x has been set.

(c) The main setting of interest, in which the Principal is uncertain about the realizations of both aggregate shocks, θ and μ.

3.2.1 Fine-Tuned Publicity: An Image-Based Pigovian Policy

Consider first the simple benchmark where agents' image motive is invariant: both they and the Principal believe with probability 1 that $\mu = \bar{\mu}$ (so $\sigma_\mu^2 = 0$). Upon observing the aggregate

[25] This is for simplicity. More generally, monetary incentives entail various costs (both direct and indirect) that limit the extent to which they can be used by the Principal. As a result, even when they are feasible it will always be optimal to also use some positive level of publicity as an additional incentive, since the gain from doing so is initially first-order, whereas the induced distortions are second-order.

[26] This cost advantage is one of the main arguments put forward by proponents of both shaming punishments (e.g., Kahan (1996), Jacquet (2015)) and public honors (e.g., Brennan and Pettit (2004)). As mentioned earlier, with developments in information technology it may even be reducing x from its laissez-faire level (protecting privacy) that necessitates costly investments.

contribution $\bar{a}$ the Principal will be able to perfectly infer θ by inverting (11), allowing her to optimally set

$$a_p = \frac{(w+\theta)[\lambda + (1-\lambda)\eta]}{k_P(1-\lambda)} = \frac{(w+\theta)\varphi}{k_P(1-\lambda)}, \tag{15}$$

where φ was defined in (5). This full revelation of θ also makes the Principal's own signal θ_P, received at the interim stage, redundant. Anticipating this at the *ex-ante* stage, the expectations of θ, μ and $\bar{a}$ she uses in choosing x are thus simply her priors $\bar{\theta}, \bar{\mu}$ and $\tilde{a}(x) = \bar{v} + \bar{\theta} + x\xi\bar{\mu}$. Substituting into the objective function (4) and differentiating with respect to x leads to an optimal level of[27]

$$x^{FB} = \frac{(w+\bar{\theta})\varphi - (\bar{v}+\bar{\theta})\lambda(1-\alpha)}{\lambda\xi\bar{\mu}} = \frac{\omega}{\lambda\xi\bar{\mu}} > 0, \tag{16}$$

where the superscripts stands for "First Best" and the wedge $\omega > 0$ was defined in (6).

Image-based Pigovian policy. Consider in particular the case of a Principal who values the public good exactly like the agents but puts no weight on their "warm-glow" utilities from contributing: $\alpha = 0$ and either $\eta = 1$ or $\lambda = 1$. The optimal level of visibility is then

$$x^{FB} = \frac{w - \bar{v}}{\lambda\xi\bar{\mu}}. \tag{17}$$

This corresponds to a "Pigovian" image subsidy which the Principal fine-tunes to exactly offset free-riding, i.e. the gap between the public good's social value w and agents' average willingness to contribute voluntarily, $\bar{v}$. More generally, by using *publicity as an incentive* according to (16) the Principal is able to achieve her preferred overall level of public-good provision (fully offsetting the wedge ω), just as she would with monetary subsidies.

3.2.2 Accounting for Variability in the Image Motive

When there are variations in the importance of social image for, $\sigma_\mu^2 > 0$, the Principal can no longer finely adjust publicity *ex ante* to achieve precise control of agents' compliance and achieve her first-best through (15)-(16). We show below that this leads her, *even if she observes* the realization of μ *ex post*, to moderate her use of visibility as an incentive mechanism.

A principal who learns the realization of μ (once x has been set) is again able, upon observing $\bar{a}$, to infer the true θ by inverting (11). As before, she will thus ignore her signal θ_P and set a_P without error, according to (15). For any choice of publicity x, however, the aggregate contribution $\bar{a}(x) = \bar{v} + \theta + x\xi\mu$ will now reflect not only the realized quality of the public good θ, but also variations in μ. Using the distribution of $\bar{a}(x)$ we can derive the Principal's expected payoff from x, denoted $E\tilde{V}(x)$. Relegating that derivation to the appendix (equation (A.4)) we focus here on the corresponding optimality condition, which embodies two opposing effects

$$\frac{dE\tilde{V}(x)}{dx} = \underbrace{(\xi\bar{\mu})\left[(w+\bar{\theta})\varphi - (\bar{v}+\bar{\theta})\lambda(1-\alpha)\right]}_{\text{Incentive Effect}} - \underbrace{\lambda x\xi^2\left(\bar{\mu}^2 + \sigma_\mu^2\right)}_{\text{Variance Effect}}. \tag{18}$$

[27] This is a special case in the proof Proposition 3 below.

The two terms clearly show the tradeoff between leveraging social pressure to promote compliance and the distortions arising from greater publicity: the variability in μ causes inefficient, image-driven variations in aggregate contributions. To the extent (λ) that the Principal internalizes the costs thus borne by the agents, she also loses from this *Variance Effect.*

Proposition 3. ***(Incentive and variance effects)*** *When the Principal faces no ex-post uncertainty about μ (symmetric information), she sets publicity level*

$$x^{SI} = \frac{\bar{\mu}\left[(w+\bar{\theta})\varphi - (\bar{v}+\bar{\theta})\lambda(1-\alpha)\right]}{\lambda\xi\left(\bar{\mu}^2+\sigma_\mu^2\right)} = \frac{x^{FB}}{1+\sigma_\mu^2/\bar{\mu}^2}, \tag{19}$$

where x^{FB} was defined in (16). This optimal x^{SI} is increasing in w, $\bar{\theta}$, α, η and σ_θ^2, decreasing in $\bar{v}$, α, s_v^2 and σ_μ^2, and U-shaped in s_θ^2 and in $1/\bar{\mu}$.

The variance effect makes publicity a blunt instrument of social control, so the Principal naturally wields it more cautiously than under the Pigovian policy: $x^{SI} < x^{FB}$, for all $\lambda > 0$.

3.3 Publicity and Information Distortion

We now turn to the main setting of interest, in which the Principal does not observe the current realization of μ and therefore faces an attribution problem: a (say) high compliance rate $\bar{a}$ could reflect high quality θ, or high image and social-enforcement concerns, μ. Using her *expected* value of μ to invert (11), she now obtains only a noisy (but still unbiased) signal of θ :

$$\hat{\theta} \equiv \frac{1}{\rho}\left[\bar{a} - \bar{v} - x\xi\bar{\mu} - (1-\rho)\,\bar{\theta}\right] = \theta + \left(\frac{x\xi}{\rho}\right)(\mu - \bar{\mu}) \sim \mathcal{N}\left(\theta,\ \frac{x^2\xi^2\sigma_\mu^2}{\rho^2}\right). \tag{20}$$

Greater publicity makes the aggregate contribution less informative (in the Blackwell sense), as it magnifies its sensitivity to variations in image concerns, μ. This *Information-Distortion Effect* will cause the Principal to makes mistakes in setting her contribution a_P (or any other second-stage decision, such as a monetary incentives, laws, etc.). Moderating this informational loss is the fact that she also receives a private signal θ_P, allowing her to update her prior beliefs to an *interim* estimate with mean $\bar{\theta}_P$ and variance $\sigma_{\theta,P}^2$:

$$\bar{\theta}_P = \left(\frac{\sigma_\theta^2}{\sigma_\theta^2 + s_{\theta,P}^2}\right)\theta_P + \left(\frac{s_{\theta,P}^2}{\sigma_\theta^2 + s_{\theta,P}^2}\right)\bar{\theta}, \tag{21}$$

$$\sigma_{\theta,P}^2 = \left(\frac{\sigma_\theta^2}{\sigma_\theta^2 + s_{\theta,P}^2}\right)^2 s_{\theta,P}^2 + \left(\frac{s_{\theta,P}^2}{\sigma_\theta^2 + s_{\theta,P}^2}\right)^2 \sigma_\theta^2. \tag{22}$$

Combining this information with the signal $\hat{\theta}$ inferred from $\bar{a}$, the Principal's posterior expectation of θ is

$$E\left[\theta|\bar{a},\theta_P\right] = \left[1-\gamma(x)\right]\bar{\theta}_P + \gamma(x)\hat{\theta}, \tag{23}$$

where the weight

$$\gamma\left(x\right)\equiv\frac{\rho^{2}\sigma_{\theta,P}^{2}}{\rho^{2}\sigma_{\theta,P}^{2}+x^{2}\xi^{2}\sigma_{\mu}^{2}},\tag{24}$$

which is clearly decreasing in x, measures the relative precision of $\hat{\theta}$, or equivalently the informational content of $\bar{a}$. After observing $\bar{a}$, the Principal optimally sets $a_P = \varphi\left(w + E\left[\theta|\bar{a}\right]\right)/(1-\lambda)k_P$; substituting in (20) and (23) immediately yields the following result.

Proposition 4. *The principal's contribution policy is equivalent to setting a baseline investment* $\underline{a}_P(x,\theta_P) = \varphi(w+\bar{\theta})/(1-\lambda)k_P$ *and a matching rate*

$$m(x)\equiv\frac{\gamma(x)\varphi}{\rho k_P(1-\lambda)}\tag{25}$$

on private contributions $\bar{a}$*. The less informative is* $\bar{a}$ *(in particular, the higher is publicity* x*), the lower is the matching rate.*

Conditioning on the true realizations of θ and μ, (11), (20) and (23) imply that the Principal's forecast error is equal to

$$\Delta\equiv E\left[\theta|\bar{a},\theta_P\right]-\theta=\left[1-\gamma(x)\right]\left(\bar{\theta}_P-\theta\right)+\frac{\gamma(x)x\xi}{\rho}\left(\mu-\bar{\mu}\right).\tag{26}$$

Her ex-ante expected payoff is reduced, relative to the symmetric-information benchmark, by a term proportional to the variance of these forecasting mistakes, which simple derivations (equation (A.7) in the appendix) show to be proportional to her loss of information:

$$EV(x)=E\tilde{V}(x)-\frac{\varphi^{2}\sigma_{\theta,P}^{2}}{2(1-\lambda)k_P}\left[1-\gamma(x)\right].\tag{27}$$

The Principal's first-order condition is now

$$\frac{dEV(x)}{dx}=\underbrace{\frac{dE\tilde{V}(x))}{dx}}_{\text{Incentive and Variance Effects}}-\underbrace{\frac{\varphi^{2}\sigma_{\mu}^{2}\xi^{2}}{\rho^{2}(1-\lambda)k_P}\gamma(x)^{2}}_{\text{Information-Distortion Effect}}x.\tag{28}$$

The first term, previously explicited in (18), embodies the beneficial incentive effect of visibility and its variability cost. The new term is the (marginal) loss from distorting information, which naturally leads to a lower choice of publicity than the optimal Pigovian policy, and even below the symmetric-information benchmark of Section 3.2.2.

Proposition 5. *When the Principal is uncertain about the importance of social image, the optimal degree of publicity $x^* \in \left(0, x^{SI}\right)$ solves the implicit equation*

$$x = \left(\frac{\bar{\mu}}{\xi}\right)\left[\frac{(w+\bar{\theta})\varphi - (\bar{v}+\bar{\theta})\lambda(1-\alpha)}{\lambda(\bar{\mu}^2+\sigma_\mu^2) + \frac{1}{(1-\lambda)k_P}\left(\frac{\varphi\sigma_\mu\gamma(x)}{\rho}\right)^2}\right]. \tag{29}$$

In general, (29) could have multiple solutions, because the cost of information distortion is not globally convex: the marginal loss, proportional to $\gamma(x)^2 x$, is hump-shaped in x.[28] While there may thus be multiple local optima, *all are below* x^{SI} (the optimum absent information-distortion issues), and therefore so is the global optimum x^*. All also share the same comparative-statics properties, to which we now turn.

4 Comparative Statics

Let us now examine how the Principal's choice of *publicity* x^* and *matching rate* $m^* = \gamma(x^*)/[\rho k_P(1-\lambda)]$ depend on key features of the environment.

A. Basic results. It is easily verified, from the first-order condition (28), that EV has positive cross-partials in (x, k_P), (x, w), $(x, -\alpha)$ and $(x, -\bar{v})$, whereas $\partial^2 EV/\partial x \partial\bar{\theta}$ is proportional to $\partial EV/\partial\bar{\theta}$, which has the sign of $\psi \equiv \lambda\alpha + (1-\lambda)\eta = \partial\omega/\partial\bar{\theta}$. To cut down on the number of cases we shall assume in what follows that $\psi > 0$, meaning that "higher quality" is indeed something that the Principal values positively. Put differently, her preferences over the quality of the public good are congruent with those of the agents, even though her preferences over the level and sharing of its supply may be quite different.[29] Hence a first set of results, summarized in Table I below.

		Optimal publicity x^*	Optimal matching rate m^*
Principal's relative cost	k_P	$\nearrow$	$\searrow$
Baseline externality	w	$\nearrow$	$\searrow$
Prior on quality	$\bar{\theta}$	$\nearrow$	$\searrow$
Weight on agents' warm-glow	α	$\nearrow$	$\searrow$
Average intrinsic motivation	$\bar{v}$	$\searrow$	$\nearrow$

Table I: Comparative-Static Effects of First-Moment Parameters

These properties are quite intuitive. For instance, a principal who faces a higher costs of own

[28] By (24), it equals $x/(1+Ax^2)^2$, where $A \equiv \xi^2\sigma_\mu^2/\rho^2\sigma_{\theta,P}^2$. Simple derivations show this function to be increasing up to $x = 1/\sqrt{3A}$, then decreasing.

[29] Clearly, $\eta \geq 0$ meaning that the Principal derives private benefits from the aggregate supply of the public good, is sufficient to ensure $\psi > 0$, as well as $\varphi > 0$ in (5). The model and all analytical results also allow for $\eta < 0$, however, corresponding to a Principal who intrinsically dislikes the activity that agents consider socially valuable –political opposition, cultural resistance, etc.

funds, or who internalizes agents' warm-glow utility, wants to encourage private contributions. She therefore makes behavior more observable and, as they become less informative, also reduces her matching rate.

We next turn to the dependence of the optimal policies on *second-moment* parameters of cross-sectional heterogeneity and aggregate variability.

B. Heterogeneity in intrinsic motivation. An increase in s_v^2 directly raises the variability of individual contributions (last term in (A.4)), and this has both costs and benefits for the Principal. To the extent that she weighs agents' warm glow positively (coefficient $\lambda\alpha$) she appreciates variability, but on other hand she suffers from internalizing its effect on their total contribution cost (coefficient λ).[30]

In addition to these direct effects, a rise in s_v^2 has indirect ones, as it increases the marginal impact of contributions on image ξ and therefore the reputational incentive to contribute, $x\xi$. For a fixed publicity x, this affects all three components of the Principal's tradeoff: it raises average contributions but further increases their sensitivity to μ, and consequently also worsens the information loss (γ declines). When publicity is optimally chosen, however, these three effects balance out exactly: because ξ and x enter each term in EV only through the product $x\xi$, we can think of the Principal as *directly optimizing over* $x\xi$; see (24) and (A.4). Variations in ξ thus have zero effect on her payoff at the first-order, *leaving only the direct impact* of s_v^2 on EV. Using the same property, we also show that the Principal responds at the margin only to the direct (variance) effect of an increase in s_v^2 : she reduces x to partially offset it, so as to keep $x\xi$ constant. Since s_v^2 influences γ and the matching rate only through the value of $x\xi$, both remain unchanged.

Proposition 6. *The Principal's optimal publicity x^* choice is decreasing in s_v^2, the variance of intrinsic motivation in the population, while the optimal matching rate m^* is independent of it. The Principal's expected payoff (at the optimal x^*) changes with s_v^2 proportionately to* $\lambda(\alpha - 1/2)$.

C. Variability in societal preferences. Comparative statics with respect to σ_θ^2 are less straightforward, it as matters through two very different channels: it represents the Principal's *ex-ante uncertainty* about θ, but also the extent to which agents disregard their signal and *follow the common prior*.

To neutralize the latter effect and highlight the Principal's key tradeoff between raising $\bar{a}$ and learning about θ, let us focus here on the limiting case in which agents' private signals are (nearly) perfect, so that σ_θ^2 plays no role in their inferences and reputational calculus. As $s_\theta^2 \to 0$, both ρ and ξ approach 1, so $\gamma(x)$ simplifies to $\sigma_\theta^2/(\sigma_\theta^2 + x^2\sigma_\mu^2)$ and the Principal's first-order condition becomes

$$\frac{dEV(x)}{dx} = \bar{\mu}\left[\left(\bar{\theta} + w\right)\varphi - \lambda(1-\alpha)\left(\bar{v} + \bar{\theta}\right)\right] - \lambda x\left(\bar{\mu}^2 + \sigma_\mu^2\right) - \frac{\varphi^2\sigma_\mu^2\gamma(x)^2 x}{(1-\lambda)k_P}. \tag{30}$$

[30] Since in equilibrium each a_i is increasing in v_i, a mean-preserving spread in v_i increases the benefit term $\alpha\int_0^1 v_i a_i d_i$ in (4), but it also magnifies the cost term $(-1/2)\int_0^1 a_i^2 d_i$.

Note that σ_θ^2 enters this expression only through $\gamma(x)$, which increases with it, while s_v^2 does not appear anywhere. Hence the following, intuitive results.

Proposition 7. *When agents have (near) perfect signals about the quality of the public good ($s_\theta^2 \to 0$), the optimal degree of visibility x^* is decreasing in σ_θ^2, the variability of this quality, while the optimal matching rate γ^* is decreasing in it. Both are independent of s_v^2, the heterogeneity in agents' valuations.*

D. Variability in the importance of social image or social enforcement. An increase in σ_μ^2 does not affect ρ or ξ and therefore leaves the incentive effect of visibility unchanged. For fixed publicity x it naturally makes $\bar{a}$ less informative about θ, so $\gamma(x)$ declines. It also leads to a higher variance effect, so for both reasons the Principal is worse off. The effects of σ_μ^2 on the optimal level of publicity and matching rate, on the other hand, are generally ambiguous: by (28), the marginal information cost is proportional to $\sigma_\mu^2 x \gamma^2(x)$, which can be seen from (24) to be hump-shaped in σ_μ^2.

Somewhat surprisingly, is may thus be that the Principal uses *more publicity* when σ_μ^2 increases. Such a "paradoxical" possibility (confirmed by simulations) only arises for intermediate values of σ_μ^2 (where the marginal information cost is near its minimum), however. When σ_μ^2 is sufficiently low or high, on the contrary, the information effect goes in the same direction as the variance effect, leading the Principal to reduce publicity, the more unpredictable is agents' sensitivity to it –as one would expect.

Another (more straightforward) case in which the result is unambiguous is when k_P is large enough: since the Principal will not contribute much anyway, information is not very valuable to her, so as σ_μ^2 rises her main concern is the variance effect.

Proposition 8. *Variability in the importance of social image, σ_μ^2, has the following effects on the Principal's payoffs and decisions:*

1. *The Principal's payoff is decreasing in σ_μ^2.*
2. *If $k_P \geq \bar{k}_P \equiv \varphi^2 / \left[27\lambda(1-\lambda)\rho^2\right]$, the optimal level of publicity x^* also decreases with σ_μ^2. Otherwise, there exist $\underline{\sigma}$ and $\bar{\sigma}$ such that x^* is decreasing in σ_μ^2 if either $\sigma_\mu < \underline{\sigma}$ or $\sigma_\mu > \bar{\sigma}$.*
3. *As σ_μ tends to 0, x^* tends to the first-best level x^{FB}, while as σ_μ tends to $+\infty$, x^* tends to 0 (full privacy).*

E. Precision of private signals

1. Principal's signal. When the variance $s_{\theta,P}^2$ of her independent signal increases, the Principal is naturally worse off from having less information. To see how she responds, note from (24) and (28) that $s_{\theta,P}^2$ appears only in the information-distortion effect, through γ. Therefore

$$\frac{\partial^2 EV(x)}{\partial x \partial s_{\theta,P}^2} = -\frac{2\varphi^2\sigma_\mu^2\xi^2\gamma(x)x}{\rho^2(1-\lambda)k_P}\left(\frac{\partial \gamma}{\partial s_{\theta,P}^2}\right) < 0.$$

This is again intuitive: as the Principal becomes less well-informed about agents' preferences, she reduces publicity so as to learn more from their behavior. Since γ increases with both x and $s_{\theta,P}$, it then follows that so does the optimal matching rate: a Principal with access to less independent information relies more on agents' behavior as a guide for her own actions.

Proposition 9. *The Principal's payoff and optimal publicity choice x^* decrease with the variance of her information, $s^2_{\theta,P}$,whereas her optimal matching rate m^* increases with it.*

2. Agents' signals. The quality of agents' private information has more ambiguous effects. At a given level of x, greater idiosyncratic noise s^2_θ reduce everyone's responsiveness to their private signal, and thereby also the informativeness of aggregate contributions. At the same time, recall from Proposition 2 that the reputational return ξ is U-shaped in s^2_θ : the level, variance and informativeness of agents' contributions are thus non-monotonic in s^2_θ, and therefore so are the Principal's optimal level of publicity and matching rate.

We can say more in the limiting case in which agents' common prior is completely uninformative (improper uniform prior), so that an individual's signal accounts for most of his belief about θ. As $\sigma^2_\theta \to +\infty$, ρ approaches 1 and the the reputational return ξ simplifies to $\bar{\xi} \equiv s^2_v / \left(s^2_v + s^2_\theta\right)$, which is strictly decreasing in s^2_θ. The Principal's first-order condition (28) then again involves x and ξ only through their product $x\xi$, with an optimal value independent of s^2_θ. Since $\bar{\xi}$ is decreasing (rather than U-shaped) in s^2_θ, we have:

Proposition 10. *When agents' prior is (nearly) uninformative ($\sigma^2_\theta \to +\infty$), the Principal's payoff is decreasing in the variance of their signals, s^2_θ. Her optimal choice of publicity is increasing in s^2_θ, and her optimal matching rate independent of it.*

The table below summarizes the results from the preceding five propositions.

	Optimal publicity x^*	Optimal matching rate m^*
s^2_v	Decreasing	Invariant
σ^2_θ	Decreasing, for s_θ small enough	Increasing, for s_θ small enough
σ^2_μ	Decreasing outside $[\underline{\sigma}, \bar{\sigma}]$, or if $k_P \geq \bar{k}_P$	Increasing outside $[\underline{\sigma}, \bar{\sigma}]$, or if $k_P \geq \bar{k}_P$
$s^2_{\theta,P}$	Decreasing	Increasing
s^2_θ	Increasing, for σ_θ large enough	Invariant

Table II: Comparative-Statics Effects of Second-Moment Parameters

5 Conclusion

We studied the tradeoff between social-incentive benefits of publicizing individual behaviors that constitute public-goods (or bads) and the costs this imposes on society (or any other Principal) when the exact social value of these actions, as well as the strength of agents' reputational concerns, are imperfectly known. Among other results, we showed that when social attitudes (what behaviors agents regard as socially desirable or undesirable) or monitoring and norms-enforcement technologies (means of communication and coordination, e.g., social media) are subject to significant change, a higher degree of privacy is optimal: this allows policy-makers to better learn, by observing overall compliance, how taxes and subsidies, the law or institutions should be adapted. When societal preferences over public goods and reputation remain relatively stable, on the other hand, the visibility of individual actions should be raised.

There are several directions in which the analysis could be further developed. A first one is an overlapping-generations environment in which the value of the public good θ, and possibly also the strength of reputational concerns μ, evolve stochastically over time. Compared to our current setup, such an explicitly dynamic analysis will introduce interesting lifecyle effects: older agents are less responsive to publicity (and more to fundamental information) since their past record is already indicative of their type, whereas younger agents are more keen to signal their motivation through their actions.

A second extension would be to examine mechanisms by which principals may overcome or alleviate thee informational problem we identify. This could for instance involve a two-stage procedure, in which agents first get to choose their participation levels anonymously—thereby revealing the state –and then, in a second stage, are asked to contribute. Dynamic procedures of this form may lead to efficiency gains because: (a) information is better or fully revealed in the first stage; (b) in the second stage, image is even more responsive to contributions than before, as the informational overjustification effect (rationalizing a low contribution as possibly reflecting a low private signal) is eliminated. Of course, such a mechanism may not be feasible in all contexts.

A third direction would be to incorporate and analyze what social psychologists refer to as *pluralistic ignorance,* namely the fact that agents themselves must often try to parse out how much of the prevailing mode of behavior around them is driven by deep preferences versus image motivations. Formally, this would correspond to an extension of the model in which image motivations are heterogeneous (as in Appendix B) and each agent knows only his own μ_i but not the average μ due again to aggregate shocks to technology, visibility or coordination in social enforcement.

6 Appendix A: Main Proofs

Proof of Proposition 1 on p. 13

Let $r(a_i,\theta_i) = \frac{dR(a,\theta_i)}{da}|_{a=a_i}$. In a differentiable reputation equilibrium, when type (v_i,θ_i) maximizes the utility represented by (3) on p. 9), we have the following first-order condition:

$$a_i - x\mu r(a_i,\theta_i) = v_i + \rho\theta_i + (1-\rho)\bar{\theta}. \tag{A.1}$$

Let us suppose that $\bar{a}$ is strictly increasing or strictly decreasing in θ so that agents can identify θ from $\bar{a}$ and knowing μ. By standard results for Normal distributions, it follows that

$$R(a_i,\theta_i)|_{\bar{a}} = E[v_i|a_i,\bar{a}] = \bar{v} + \xi\left(a_i - x\mu r(a_i,\theta_i) - \bar{v} - \rho\theta - (1-\rho)\bar{\theta}\right), \tag{A.2}$$

Taking the derivative with respect to a_i and taking expectations with respect to θ_i implies that

$$r(a_i,\theta_i) = \xi(1 - x\mu r_a(a_i,\theta_i)). \tag{A.3}$$

The generic solution to this differential equation is $r(a,\theta_i) = \xi + \zeta e^{-a/\xi x\mu}$, in which ζ is a constant of integration. Notice that if $\zeta \neq 0$, then each agent's problem is not globally concave, and may be maximized at $\pm\infty$. Therefore, at a well-defined equilibrium, $\zeta = 0$ leading to the solution specified in Proposition 1. ■

Proof of Proposition 3 on p. 16

For each agent i, $a_i = x\xi\mu + v_i + \rho\theta_i + (1-\rho)\,\bar{\theta}$, and therefore $\bar{a}(\theta,\mu) \equiv x\xi\mu + \bar{v} + \bar{\theta} + \rho\left(\theta - \bar{\theta}\right)$. Let $\bar{\bar{a}} \equiv x\xi\bar{\mu} + \bar{v} + \bar{\theta}$ represent the expected aggregate contribution.

Since the Principal observes μ, she an infer θ perfectly from $\bar{a}$ and so will set $a_P = (w + \theta)\varphi/(1-\lambda)k_P$, independently of x (recall that $\varphi \equiv \lambda + \eta(1-\lambda)$). Let us define $\bar{a}_P \equiv (w + \bar{\theta})\varphi/(1-\lambda)k_P$ as the expected Principal's contribution.

Integrating over θ and μ, we obtain from (4):

$$\begin{aligned}
E\tilde{V}(x) =& \lambda\left[\alpha\left(s_v^2 + \rho\sigma_\theta^2 + (\bar{v}+\bar{\theta})(\bar{\bar{a}})\right) + (w+\bar{\theta})(\bar{\bar{a}} + \bar{a}_P + \rho\sigma_\theta^2) + \frac{\sigma_\theta^2\varphi}{(1-\lambda)k_P}\right] \\
&+ (1-\lambda)\eta\left[(w+\bar{\theta})(\bar{\bar{a}} + \bar{a}_P + \rho\sigma_\theta^2) + \frac{\sigma_\theta^2\varphi\eta}{(1-\lambda)k_P}\right] \\
&- \frac{\lambda}{2}[\bar{\bar{a}}^2 + \rho^2(\sigma_\theta^2 + s_\theta^2) + s_v^2 + x^2\xi^2\sigma_\mu^2] \\
&- \frac{(1-\lambda)k_P}{2}\left[\bar{a}_P^2 + \sigma_\theta^2\left(\frac{\varphi}{(1-\lambda)k_P}\right)^2\right].
\end{aligned} \tag{A.4}$$

Differentiating yields:

$$\frac{dE\tilde{V}(x)}{dx} = \{\lambda\left[\alpha(\bar{v}+\bar{\theta}) + (w+\bar{\theta})\right] + (1-\lambda)\eta(w+\bar{\theta})\}\xi\bar{\mu} - \lambda\left[\xi\bar{\mu}\left(x\xi\bar{\mu} + \bar{v} + \bar{\theta}\right) + x\xi^2\sigma_\mu^2\right]. \tag{A.5}$$

For all $\lambda > 0$, the expression is strictly concave in x, therefore the first-order condition described in (18) characterizes the unique optimum. Equating the right-hand-side to zero yields (19), which simplifies to (16) when $\sigma^2_\mu = 0$. ■

Proof of Proposition 5 on p. 19

For every θ, were the Principal to observe θ or the realization of μ, recall that she would choose a contribution level of $(w+\theta)\varphi/(1-\lambda)k_P$. When she is unable to observe θ or μ, she sets $a_P = (w + E[\theta|\bar{a}, \theta_P])\varphi)/(1-\lambda)k_P$, which makes clear how the forecast error Δ derived in (26) generates inefficient deviations from full-information optimality. Note that

$$E\left[\Delta^2\right] = (1-\gamma)^2\,\sigma^2_{\theta,P} + (\gamma\xi x/\rho)^2\,\sigma^2_\mu = \sigma^2_{\theta,P}\left[(1-\gamma)^2 + \gamma^2\,(1/\gamma - 1)\right] = \sigma^2_{\theta,P}\,(1-\gamma)\,, \quad \text{(A.6)}$$

where we abbreviated $\gamma(x)$ as γ and used the fact that $x^2\xi^2\sigma^2_\mu/\rho^2 = \sigma^2_\theta\,(1-\gamma)\,/\gamma$.

Therefore, in a state θ, the distribution of the Principal's contribution is

$$N\left(\frac{(w+\theta)\varphi}{(1-\lambda)k_P}, \left(\frac{\varphi}{(1-\lambda)k_P}\right)^2 \sigma^2_{\theta,P}\,(1-\gamma)\right),$$

and its variance effectively increases the Principal's expected cost by

$$\frac{(1-\lambda)k_P}{2}\left[\left(\frac{\varphi}{(1-\lambda)k_P}\right)^2 \sigma^2_{\theta,P}\,(1-\gamma)\right] = \frac{\varphi^2\sigma^2_{\theta,P}}{2(1-\lambda)k_P}(1-\gamma). \quad \text{(A.7)}$$

For given x and for every realization of θ, note that $E[\theta\Delta|\theta] = 0$. Therefore, it follows by inspection that all the other terms in the Principal's payoff (26) remain unchanged from the case where she knows μ, and therefore, (27) characterizes the change in payoffs. Note also that

$$\begin{aligned}\frac{\sigma^2_{\theta,P}}{2}\frac{d\gamma}{dx} &= -\frac{\sigma^2_{\theta,P}}{2}\left(\frac{2\rho^2\sigma^2_{\theta,P}\xi^2\sigma^2_\mu}{\left(\rho^2\sigma^2_{\theta,P} + x^2\xi^2\sigma^2_\mu\right)^2}x\right) = \frac{\sigma^2_{\theta,P}\gamma\,(1-\gamma)}{x} \\ &= -\sigma^2_{\theta,P}\left(\frac{\gamma^2\xi^2\sigma^2_\mu}{\rho^2\sigma^2_{\theta,P}}x\right) = -\frac{\sigma^2_\mu\gamma^2\xi^2 x}{\rho^2}.\end{aligned}$$

Therefore

$$\begin{aligned}\frac{\partial EV}{\partial x} &= \frac{\partial E\tilde{V}}{\partial x} - \frac{\varphi^2}{(1-\lambda)k_P}\left(\frac{\sigma^2_\mu\gamma^2\xi^2 x}{\rho^2}\right) \\ &= (\xi\bar{\mu})\left[(w+\bar{\theta})\varphi - (\bar{v}+\bar{\theta})(1-\alpha)\lambda\right] - \lambda x\xi^2\left(\bar{\mu}^2 + \sigma^2_\mu\right) - \frac{\varphi^2}{(1-\lambda)k_P}\left(\frac{\sigma^2_\mu\gamma^2\xi^2 x}{\rho^2}\right), \quad \text{(A.8)}\end{aligned}$$

which corresponds to (29). Recall now that $E\tilde{V}(x)$ is strictly concave in x and maximized at $\tilde{x} > 0$. Therefore, $\partial EV/\partial x < \partial E\tilde{V}/\partial x < 0$ for all $x \geq \tilde{x}$, and at $x = 0$, $\partial EV/\partial x = \partial E\tilde{V}/\partial x > 0$. Consequently, the global maximum of EV on $\mathbb{R}$ is reached at some $x^* \in (0, \tilde{x})$

where $\partial EV/\partial x = 0$. ■

Proof of Proposition 6 on p. 20 Denote $x\xi$ by z and note using A.4 and ((27)) that $EV(x)$ can be reformulated as

$$\begin{aligned}\mathcal{V}(z) =& s_v^2\left(\lambda\alpha - \frac{\lambda}{2}\right) + z\bar{\mu}\left((\bar{w}+\bar{\theta})\varphi - \lambda(1-\alpha)(\bar{v}+\bar{\theta})\right)\\ &- \frac{\lambda}{2}z^2(\bar{\mu}^2+\sigma_\mu^2) - \frac{\varphi^2\sigma_{\theta,P}^2}{2(1-\lambda)k_P}\left[1-\tilde{\gamma}(z)\right] + C,\end{aligned} \tag{A.9}$$

in which

$$\tilde{\gamma}(z) \equiv \frac{\rho^2\sigma_{\theta,P}^2}{\rho^2\sigma_{\theta,P}^2 + z^2\sigma_\mu^2}, \tag{A.10}$$

and C is a constant that is independent of both s_v^2, x, and ξ. Therefore, the optimal z solves the first-order condition

$$\bar{\mu}[(\bar{w}+\bar{\theta})\varphi - \lambda(1-\alpha)(\bar{v}+\bar{\theta})] - \lambda z(\bar{\mu}^2+\sigma_\mu^2) + \frac{\varphi^2\sigma_{\theta,P}^2}{2(1-\lambda)k_P}\tilde{\gamma}'(z) = 0. \tag{A.11}$$

Notice that none of these terms depend on s_v^2, and so the optimal z is independent of s_v^2. Therefore, for each s_v, the optimal $x^*(s_v)\xi(s_v)$ is constant. This fact automatically implies that in equilibrium, changes in s_v^2 do not influence γ or the matching rate. Since $\xi(s_v)$ is increasing in s_v, it follows that $x^*(s_v)$ must be decreasing in s_v. Finally, it follows from (A.9) that $d[EV(x^*(s_v^2); s_v^2)]/ds_v^2 = \lambda\left(\alpha - 1/2\right)$. ■

Proof of Proposition 7 on p. 21

In (30), $\partial^2 EV/\partial x\partial\sigma_\theta < 0$, so $\partial x^*/\partial\sigma_\theta < 0$ since $\partial^2 EV/\partial x^2 < 0$ at the optimum. As σ_θ rises, the third term on the right-hand side therefore increases $dEV(x)/dx$, implying that $x^*\gamma(x^*;\sigma_\theta)^2$ must also increase to maintain optimality. Therefore $\gamma(x^*;\sigma_\theta)$ rises with σ_θ, and hence so does $m^* = \gamma(\lambda + (1-\lambda)\eta)/((1-\lambda)k_P)$. ■

Proof of Proposition 8 on p. 21 The negative impact of increasing σ_μ^2 on payoffs is clear: for every θ and x, changes in σ_μ^2 have no effect on $\bar{a}$ but increase the variance of aggregate contributions and the information cost. To consider their impact on optimal publicity, observe from (28) that

$$\frac{\partial^2 EV}{\partial x\partial\sigma_\mu^2} = -\lambda x\xi^2 - \frac{\varphi^2\xi^2 x}{\rho^2(1-\lambda)k_P}\left(\gamma^2 + 2\gamma\sigma_\mu^2\frac{d\gamma}{d\sigma_\mu^2}\right), \tag{A.12}$$

in which

$$\frac{\partial\gamma}{\partial\sigma_\mu^2} = -\frac{\rho^2\sigma_\theta^2 x^2\xi^2}{\left(\rho^2\sigma_\theta^2 + x^2\xi^2\sigma_\mu^2\right)^2} = -\frac{\gamma x^2\xi^2}{\rho^2\sigma_\theta^2 + x^2\xi^2\sigma_\mu^2} = -\frac{\gamma\left(1-\gamma\right)}{\sigma_\mu^2}. \tag{A.13}$$

Thus,

$$\frac{\partial^2 EV}{\partial x\partial\sigma_\mu^2} = -\lambda x\xi^2 - \frac{\varphi^2\xi^2 x}{\rho^2(1-\lambda)k_P}\left(\gamma^2 - 2\gamma^2(1-\gamma)\right) = -\lambda x\xi^2 - \frac{\varphi^2\xi^2\gamma^2 x}{\rho^2(1-\lambda)k_P}\left(2\gamma - 1\right). \tag{A.14}$$

This expression is non-positive if and only if

$$\frac{\lambda(1-\lambda)\rho^2 k_P}{\varphi^2} \geq \gamma^2(1-2\gamma). \tag{A.15}$$

Because $\gamma^2\,(1-2\gamma)$ takes on a maximum value of $1/27$, a sufficient condition is that the left-hand side of the equation above exceeds $1/27$. In this case, $\partial x/\partial\sigma_\mu^2 < 0$ for all values of σ_μ. Intuitively, when k_P is large enough the value of information for the Principal is small (she does not have much of a decision to make), so whether a higher σ_μ^2 improves or worsens the information effect, it is dominated by its worsening of the variance effect.

If the condition is not satisfied, then monotonicity generally does not hold everywhere, but:

(a) As σ_μ^2 tends to 0, $\gamma(x^*(\sigma_\mu^2);\sigma_\mu^2)$ approaches 1, because by Proposition 5, $x^*(\sigma_\mu^2)$ remains bounded above: $x^*(\sigma_\mu^2) < \bar{x}$. Therefore, (A.15) holds for σ_μ small enough.

(b) As σ_μ^2 tends to ∞, $x^*(\sigma_\mu^2)$ must tend to 0 fast enough that the product $\sigma_\mu^2 x^*(\sigma_\mu^2)$ remains bounded above. Otherwise, equation (28) shows that the first-order condition $\partial EV/\partial x = 0$ cannot hold, as the marginal variance effect and the marginal information-distortion effects both become arbitrarily large. It then follows that that $\sigma_\mu^2\left[x^*(\sigma_\mu^2)\right]^2$ tends to 0, and therefore $\gamma(x^*(\sigma_\mu^2);\sigma_\mu^2)$ tends to 1. Thus, for σ_μ^2 large enough (A.15) holds, and $x^*(\sigma_\mu^2)$ decreases toward 0. ■

Proof of Proposition 10 on p. 22

Taking limit as $\sigma_\theta \to \infty$, ρ converges to 1 and therefore, ξ converges to $\bar{\xi}$. By inspection, all terms in (27) in which x and ξ enter do so through their product. Therefore, in order to study how the optimal $x^*(s_\theta)$ and the Principal's welfare depend on $x^*(s_\theta^2)$, we can write $EV\,(x;s_\theta) = \mathcal{V}(x\bar{\xi}(s_\theta)) - \lambda s_\theta^2/2$ for some appropriately defined function $\mathcal{V}$. The same reasoning as in the proof of Proposition 6 then shows that $d\left[EV(x^*(s_\theta);s_\theta)\right]/ds_\theta = -\lambda s_\theta^2/2 < 0$. Because the Principal keeps $x^*(s_\theta)\bar{\xi}(s_\theta)$ constant as s_θ increases, it must be that increases in $x^*(s_\theta)$ compensate for how $\bar{\xi}(s_\theta)$ decreases in s_θ. ■

7 Appendix B: Heterogenous Image Concerns

7.1 Extension of the Model and Results

We have so far assumed that all agents have similar reputational concerns, but in practice, some naturally care more about their social image than others. As shown by Bénabou and Tirole (2006), such heterogeneity introduces another source of doubt about what accounts for an individual's contribution –was he more motivated by the common good, or improving his image? This additional *overjustification effect* reduces the reputational return to contributing and therefore has a detrimental effect on the overall provision of the public good. In the present context, on the other hand, it also implies that agents become less responsive to fluctuations in the value of social image, so that their aggregate behavior $\bar{a}$ is more informative about the quality of the public good. For this reason, is not obvious *a priori* how heterogeneity in image concerns will affect the Principal and her optimal policy.

To study this issue, let each agent's image concern be the sum of a common factor and some idiosyncratic component: Agent i's overall preferences are now given by

$$U(a_i, \bar{a}, \theta) + \mu_i R\left(a_i, \theta_i\right), \tag{B.1}$$

where U and R are defined as before and μ_i is distributed in the population as $N\left(\mu, s_\mu^2\right)$, with $\mu \sim N\left(\mu, \sigma_\mu^2\right)$ as before.

7.1.1 Agents' behavior

Proposition 11. *1. For any $x \geq 0$, there exists a unique linear equilibrium in contributions. An agent of type v_i with signal θ_i and image concern μ_i chooses*

$$a_i\left(v_i, \theta_i; \mu_i\right) = v_i + \rho\theta_i + (1-\rho)\bar{\theta} + \mu_i x\tilde{\xi}\left(x\right), \tag{B.2}$$

where ρ is still given by (9) and $\tilde{\xi}\left(x\right)$ is the unique solution to

$$\tilde{\xi}\left(x\right) = \frac{s_v^2}{x^2\tilde{\xi}\left(x\right)^2 s_\mu^2 + s_v^2 + \rho^2 s_\theta^2}. \tag{B.3}$$

The resulting aggregate contribution is

$$\bar{a}\left(\theta; \mu\right) = \bar{v} + \rho\theta + (1-\rho)\bar{\theta} + \mu x\tilde{\xi}\left(x\right). \tag{B.4}$$

2. The signal-to-noise ratio $\tilde{\xi}\left(x\right)$ is strictly decreasing in x, s_μ^2, σ_θ^2, strictly increasing in s_v^2 and inverse-U shaped in and s_θ^2. The impact of visibility on contributions, $\beta(x) \equiv x\tilde{\xi}\left(x\right)$, is strictly increasing in x, with $\lim_{x\to\infty} x\tilde{\xi}\left(x\right) = +\infty$, and shares the properties of $\tilde{\xi}\left(x\right)$ with respect to variance parameters.

The interpretation of $\tilde{\xi}\left(x\right)$ is identical to that of ξ in Section 2: it measures the marginal impact that contributions have on image, given the equilibrium decision rule (B.2). Note that in anonymous settings, $\tilde{\xi}\left(0\right) = \xi$, but as soon as there is some visibility $x > 0$, $\tilde{\xi}\left(x\right) < \xi$. This reflects the overjustification effect from heterogeneity in publicity-seeking motives, which gets amplified when actions become more visible are. This weakens the direct effect of publicity on the reputational incentive to contribute (a form of partial crowding out): $\beta(x) = x\xi(x)$ increases less than one for one with x. For the same reason, in contrast to the case of a common $\mu_i = \mu$, the reputational return $\xi(x)$ is determined as a fixed point that depends on x; see (B.3).

Note, finally, that idiosyncratic differences in μ_i's wash out in the aggregate contribution $\bar{a}$, implying:

Corollary 1. *At any given level of x, the informational content $\gamma(x)$ of aggregate compliance*

$\bar{a}$, the Principal's optimal matching rate $m(x)$ and her informational loss $EV(x) - \tilde{E}V(x)$ from not observing the aggregate realization μ remain the same as in (24), (25) and (27) respectively, except that $x\xi$ is everywhere replaced by $x\tilde{\xi}(x) = \beta(x)$.

7.1.2 Optimal Publicity

1. Symmetric information. We again first consider the case in which the Principal, like the agents, learns the realization of (the average) μ after x has been set (or together with observing $\bar{a}$). The following results thus generalize Proposition 3.

Proposition 12. *When the Principal faces no ex-post uncertainty about μ, she sets a publicity level x_* given by the unique solution to*

$$x^{SI} = \frac{\bar{\mu}\omega}{\lambda\tilde{\xi}\left(x^{SI}\right)\left(\bar{\mu}^2 + \sigma_\mu^2 + s_\mu^2\right)}, \tag{B.5}$$

,which is lower than x^{FB} and strictly decreasing in s_μ^2. The resulting aggregate contributions are

$$\bar{a}\left(\theta;\mu\right) = \bar{v} + \rho\theta + \left(1-\rho\right)\bar{\theta} + \frac{\mu}{\bar{\mu}}\left(\frac{\omega/\lambda}{1 + \left(\sigma_\mu^2 + s_\mu^2\right)/\bar{\mu}^2}\right), \tag{B.6}$$

which decreases with s_μ^2 for all $\mu > 0$. The Principal's utility is decreasing in s_μ^2.

We saw that as s_μ^2 increases the "overjustification effect" worsens ($\tilde{\xi}(x)$ decreases), leading to a lower average contribution (for $\mu > 0$, the probability of which can be made arbitrarily large by taking $\bar{\mu} >> 0$; otherwise everything just flips). Increasing publicity would boost contributions, but this would further aggravate the loss from inefficient cost variance, which directly rises with s_μ^2. The comparative-statics result $\partial x^{SI}/\partial s_\mu^2 < 0$ shows that the latter concern always dominates.

2. Asymmetric information. We next incorporate the information-distortion concern into the Principal's problem, recalling from the above Corollary that it is unchanged from the case of common μ, except that $\gamma(x)$ embodies the new signal-to-noise ratio $\xi(x)$.

Proposition 13. *When the Principal is uncertain about the importance of social image, the optimal degree of publicity $x^* \in \left(0, x^{SI}\right)$ solves the implicit equation*

$$x^* = \left(\frac{\bar{\mu}}{\xi(x^*)}\right)\left[\frac{\omega}{\lambda(\bar{\mu}^2 + \sigma_\mu^2 + s_\mu^2) + \frac{1}{(1-\lambda)k_P}\left(\frac{\varphi\sigma_\mu\gamma(x^*)}{\rho}\right)^2}\right], \tag{B.7}$$

where $\xi(x)$ is given by (B.3) and $\gamma(x)$ remains given by (24). The solution is thus identical to that in (5), except, that σ_μ^2 is replaced by $\sigma_\mu^2 + s_\mu^2$ and ξ by $\xi(x)$ everywhere.

As before, (B.7) could have multiple solutions but , including therefore the global optimum x^*, are below x^{SI} and share the same comparative-statics properties.

It is also easy to verify that, because x enters EV only through the increasing function $\beta(x) = x\tilde{\xi}(x)$,the indirect effects of s_μ^2 on the Principal's optimized objective function $EV(x^*)$ cancel out at the first order, leaving only the direct (variance) effect $(-\lambda/2)\beta(x)^2 < 0$. Therefore:

Proposition 14. *The Principal's expected payoff is strictly decreasing in* s_μ^2.

7.2 Proofs

Proof of Proposition 11 on p. 28. As in the proof of Proposition 1, suppose that the linear contribution function is given by

$$a\left(\mu_i, v_i, \theta_i\right) = Ax\mu_i + Bv_i + C\theta_i + D,$$

which implies that $\bar{a} = A\mu + B\bar{v} + C\theta + D$. Since agents know μ, they can solve for $\theta = \left[\bar{a} - Ax\mu - B\bar{v} - D\right]/C$. Therefore, regardless of their signal realization θ_j,

$$E\left[v_i \mid a_i, \bar{a}, \theta_j\right] = E\left[v_i \;\middle|\; a_i,\; \theta = \frac{\bar{a} - Ax\mu - B\bar{v} - D}{C}\right] = \bar{v} + \frac{\tilde{\xi}}{B}\left(a_i - \bar{a}\right), \tag{B.8}$$

$$\text{where } \tilde{\xi} = \frac{s_v^2}{\left(Ax/B\right)^2 s_\mu^2 + s_v^2 + \frac{C^2}{B^2}s_\theta^2}. \tag{B.9}$$

This simplifies the (normalized) reputational payoffs to:

$$\begin{aligned} R\left(a_i, \theta_i\right) = E\left[E\left[v_i | a_i, \bar{a}\right] | \theta_i\right] &= E\left[\left(\bar{\nu} + \frac{\tilde{\xi}}{B}\left(a_i - \bar{a}\right)\right) | \theta_i\right] \\ &= \left(\bar{v} + \frac{\tilde{\xi}}{B}\left(a_i - Ax\mu - B\bar{v} - C\rho\theta_i - C\left(1-\rho\right)\bar{\theta} - D\right)\right). \end{aligned} \tag{B.10}$$

Utility maximization yields the first-order condition:

$$a_i = v_i + E\left[\theta | \theta_i\right] + \frac{x\mu_i\tilde{\xi}}{B} = \frac{x\tilde{\xi}}{B}\mu_i + v_i + \rho\theta_i + \left(1-\rho\right)\bar{\theta}. \tag{B.11}$$

Therefore, $B = 1$, $C = \rho$, $D = 1 - \rho$, and $A = x\tilde{\xi}/B = x\tilde{\xi}$. It remains to be shown that for each choice of x, $\tilde{\xi}\left(x\right)$ is unique. Let $\beta\left(x\right) \equiv x\tilde{\xi}\left(x\right)$, and observe from (B.8) that $\beta\left(x\right)$ solves the implicit equation

$$\frac{x}{\beta} = \beta^2\left(\frac{s_\mu^2}{s_v^2}\right) + \frac{\rho^2 s_\theta^2}{s_v^2} + 1. \tag{B.12}$$

Since the left-hand side is decreasing in β and the right-hand side increasing in β, there exists a unique solution for each choice of x. From (B.3), $\tilde{\xi}\left(x\right)$ must be strictly decreasing in x, s_μ^2, σ_θ^2, strictly increasing in s_v^2 and inverse-U shaped in s_θ, since ρ is U-shaped. From Equation (B.12), $\beta\left(x\right)$ must be strictly increasing in x and decreasing in s_μ^2. Finally, observe that $\lim_{x\to\infty}\beta\left(x\right) = \infty$, otherwise the left-hand side of (B.12) would be unbounded while the right-hand side would remain bounded, generating a contradiction. ■

Proof of Proposition 12 on p. 29. Proposition 11 shows that, given any x, the equilibrium among agents is the same as in the case where $s_\mu^2 = 0$, except that ξ is replaced everywhere by $\tilde{\xi}\,(x)$, or equivalently $x\xi$ by $\beta(x) = x\tilde{\xi}(x)$ in all type-independent expressions (first and second moments), while at the individual level $\mu x\xi$ is replaced by $\mu_i\beta(x)$.

Let us denote by $a_i^0 \equiv v_i + \rho\theta_i + (1-\rho)\,\bar{\theta} + \mu x\tilde{\xi}\,(x)$ the value of a_i corresponding to the mean value of $\mu_i = \mu$, or equivalently the value of a_i in the original (homogeneousμ) model where we simply replace ξ by $\tilde{\xi}(x)$. Similarly, let $\tilde{V}^0(x)$ (respectively, $V^0(x)$) be the utility level the Principal would achieve if agents behaved according to a_i^0 and she observes (respectively, does not observe) the realization of the average μ.

We can obtain $E\tilde{V}^0(x)$ directly by replacing ξ with $\tilde{\xi}(x)$ in the expression (A.4) giving $EV(x)$, and similarly $dEV^0(x)/dx$ by replacing $x\xi$ with $\beta(x)$ and ξ with $\beta'(x)$ in the expression (A.5) for $dEV(x)/dx$:

$$\frac{dE\tilde{V}^0(x)}{dx} = \omega\bar{\mu}\beta'(x) - \lambda\beta'(x)\left[\bar{\mu}\left(\beta(x)+\bar{v}+\bar{\theta}\right)+\beta(x)\sigma_\mu^2\right] = 0.$$

In the Principal's actual loss function (4), however, the heterogeneity in agents' μ_i's generates an additional loss dues to inefficient cost variations, equal to $(\lambda/2)E[(a_i)^2 - (a_i^0)^2] = (\lambda/2)\beta(x)^2 s_\mu^2$. Therefore, when the Principal observe the realization of μ, the optimal (symmetric-information) value of x is given by the first-order condition

$$\frac{dE\tilde{V}}{dx} = \bar{\mu}\beta'\,(x)\left[\left(w+\bar{\theta}\right) - \alpha\left(\bar{v}+\bar{\theta}\right)\right] - \alpha\beta'\,(x)\,\beta\,(x)\left(\bar{\mu}^2+\sigma_\mu^2+s_\mu^2\right) = 0, \tag{B.13}$$

hence

$$\beta\left(x^{SI}\right) = \frac{\bar{\mu}\omega}{\lambda\left(\bar{\mu}^2+\sigma_\mu^2+s_\mu^2\right)}, \tag{B.14}$$

which is equivalent to (B.5). Furthermore, since the right-hand side is strictly decreasing in s_μ^2 and $\beta\,(x)$ was shown to be strictly increasing in x, x^{SI} must be decreasing in s_μ^2. Since $\bar{a}\,(\theta,\mu)$ is strictly increasing in $\beta\,(x_*)$ as long as $\mu > 0$, finally $\bar{a}$ is then decreasing in s_μ^2 for every θ and $\mu > 0$. ■

Proof of Proposition 13 on p. 29. Combining (B.13) and (27) yields for the Principal's first-order condition:

$$\frac{dEV}{dx} = \bar{\mu}\beta'\,(x)\,\omega - \lambda\beta'\,(x)\,\beta\,(x)\left(\bar{\mu}^2+\sigma_\mu^2+s_\mu^2\right) + \frac{\varphi^2\sigma_{\theta,P}^2}{2(1-\lambda)k_P}\gamma'(x) = 0.$$

Recalling that

$$\gamma\,(x) \equiv \frac{\rho^2\sigma_{\theta,P}^2}{\rho^2\sigma_{\theta,P}^2+\beta(x)^2\sigma_\mu^2} \quad\Rightarrow\quad \gamma'\,(x) = -\frac{2\sigma_\mu^2}{\rho^2\sigma_{\theta,P}^2}\beta(x)\beta'(x)\gamma(x)^2,$$

this yields

$$\beta(x) = \frac{\bar{\mu}\omega}{\lambda(\bar{\mu}^2 + \sigma_\mu^2 + s_\mu^2) + \frac{1}{(1-\lambda)k_P}\left(\frac{\varphi\sigma_\mu\gamma(x)}{\rho}\right)^2}, \tag{B.15}$$

which is equivalent to (B.7). ■

References

Acquisti, A., C. Taylor, and L. Wagman (2016): "The Economics of Privacy," *Journal of Economic Literature, forthcoming.*

Algan, Y., Y. Benkler, M. F. Morell, and J. Hergueux (2013): "Cooperation in a Peer Production Economy Experimental Evidence from Wikipedia," .

Andreoni, J. (2006): "Leadership giving in charitable fund-raising," *Journal of Public Economic Theory*, 8(1), 1–22.

Andreoni, J., and B. D. Bernheim (2009): "Social image and the 50–50 norm: a theoretical and experimental analysis of audience effects," *Econometrica*, 77(5), 1607–1636.

Ariely, D., A. Bracha, and S. Meier (2009): "Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially," *The American Economic Review*, 99(1), 544–555.

Ashraf, N., O. Bandiera, and K. Jack (2012): "No margin, No mission," *A Field Experiment on Incentives for Pro-Social Tasks, CEPR Discussion Papers*, 8834.

Auriol, E., and R. J. Gary-Bobo (2012): "On the optimal number of representatives," *Public Choice*, 153(3-4), 419–445.

Bagwell, L., and B. D. Bernheim (1996): "Veblen effects in a theory of conspicuous consumption," *The American Economic Review*, 86, 349–373.

Bar-Isaac, H. (2012): "Transparency, career concerns, and incentives for acquiring expertise," *The BE Journal of Theoretical Economics*, 12(1).

Battaglini, M., and R. Bénabou (2003): "Trust, coordination, and the industrial organization of political activism," *Journal of the European Economic Association*, 1(4), 851–889.

Bénabou, R., and J. Tirole (2003): "Intrinsic and extrinsic motivation," *Review of Economic Studies*, pp. 489–520.

——— (2004): "Willpower and personal rules," *Journal of Political Economy*, 112(4), 848–886.

——— (2006): "Incentives and prosocial behavior," *American Economic Review*, 96(5), 1652–1678.

——— (2011): "Laws and Norms," *NBER Working Paper 17579.*

Bernheim, B. D. (1994): "A Theory of Conformity," *Journal of Political Economy*, 102, 841–877.

Bolton, P., M. K. Brunnermeier, and L. Veldkamp (2013): "Leadership, coordination, and corporate culture," *The Review of Economic Studies*, 80(2), 512–537.

Brandenburger, A., and B. Polak (1996): "When managers cover their posteriors: Making the decisions the market wants to see," *RAND Journal of Economics*, 27(3), 523–541.

Brennan, G., and P. Pettit (2004): *The economy of esteem.* Oxford University Press New York.

Cooter, R. D. (2003): "Donation Registry, The," *Fordham L. Rev.*, 72, 1981–1989.

Corneo, G., and O. Jeanne (1997): "Conspicuous consumption, snobbism and conformism," *Journal of public economics*, 66(1), 55–71.

Corneo, G. G. (1997): "The theory of the open shop trade union reconsidered," *Labour Economics*, 4(1), 71–84.

Croson, R., and M. Marks (1998): "Identifiability of individual contributions in a threshold public goods experiment," *Journal of Mathematical Psychology*, 42(2), 167–190.

Daughety, A., and J. Reinganum (2010): "Public Goods, Social Pressure, and the Choice Between Privacy and Publicity," *American Economic Journal: Microeconomics*, 2(2), 191–221.

Del Carpio, L. (2014): "Are The Neighbors Cheating? Evidence from a Social Norm Experiment on Property Taxes in Peru," *Princeton University mimeo,*.

Della Vigna, S., J. List, and U. Lalmendier (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," *American Economic Review*, 98(3), 990–1008.

Fehrler, S., and N. Hughes (2014): "How Transparency Kills Information Aggregation," .

Fischer, P. E., and R. E. Verrecchia (2000): "Reporting Bias," *The Accounting Review*, 75(2), 229–245.

Fox, J., and R. V. Weelden (2012): "Costly transparency," *Journal of Public Economics*, 96(1), 142–150.

Frankel, A., and N. Kartik (2014): "Muddled Information," .

Frey, B. S., and S. Neckermann (2013): "Prizes and Awards," in *Handbook on the Economics of Reciprocity and Social Enterprise*, ed. by L. Bruni, and S. Zamagni, pp. 271–276. Edward Elgar, Cheltenham, UK.

Gerber, A. S., D. Green, and C. Larimer (2008): "Social pressure and voter turnout: Evidence from a large-scale field experiment," *American Political Science Review*, 102(01), 33–48.

Glazer, A., and K. Konrad (1996a): "A signaling explanation for charity," *The American Economic Review*, 86, 1019–1028.

——— (1996b): "A signaling explanation for charity," *The American Economic Review*, pp. 1019–1028.

Harbaugh, W. (1998): "What do donations buy? A model of philanthropy based on prestige and warm glow," *Journal of Public Economics*, 67(2), 269–284.

Harbaugh, W. T., U. Mayr, and D. R. Burghart (2007): “Neural responses to taxation and voluntary giving reveal motives for charitable donations,” *Science*, 316(5831), 1622–1625.

Hermalin, B., and M. Katz (2006): “Privacy, property rights and efficiency: The economics of privacy as secrecy,” *Quantitative Marketing and Economics*, 4(3), 209–239.

Holmström, B. (1999): “Managerial incentive problems: A dynamic perspective,” *Review of Economic Studies*, pp. 169–182.

Hummel, P., J. Morgan, and P. C. Stocken (2013): “A model of flops,” *The RAND Journal of Economics*, 44(4), 585–609.

Ichino, A., and G. Muehlheusser (2008): “How often should you open the door?: Optimal monitoring to screen heterogeneous agents,” *Journal of Economic Behavior & Organization*, 67(3âĂŞ4), 820 – 831.

Jacquet, J. (2015): *Is Shame Necessary? New Uses for an Old Tool.* Pantheon Books, Random House.

Kahan, D. M. (1996): “Between economics and sociology: The new path of deterrence,” *Michigan Law Review*, 95, 2477–2497.

Kahan, D. M., and E. A. Posner (1999): “Shaming White-Collar Criminals: A Proposal for Reform of the Federal Sentencing Guidelines*,” *The Journal of Law and Economics*, 42(S1), 365–392.

Kreps, D. M. (1990): “Corporate Culture and Economic Theory,” in *Perspectives on Positive Political Economy*, ed. by J. E. Alt, and K. A. Shelpsle, pp. 90–143. Cambridge Univ Press.

Kuran, T. (1997): *Private truths, public lies: The social consequences of preference falsification.* Harvard University Press.

Lacetera, N., and M. Macis (2010): “Social image concerns and prosocial behavior: Field evidence from a nonlinear incentive scheme,” *Journal of Economic Behavior & Organization*, 76(2), 225–237.

Larkin, I. (2011): “Paying 30K for a Gold Star: An Empirical Investigation into the Value of Peer Recognition to Software Salespeople,” Harvard Business School.

Levy, G. (2005): “Careerist judges and the appeals process,” *Rand Journal of Economics*, 36(2), 275–297.

——— (2007): “Decision Making in Committees: Transparency, Reputation, and Voting rules,” *The American Economic Review*, 97(1), 150–168.

Lohmann, S. (1994): “Information aggregation through costly political action,” *The American Economic Review*, pp. 518–530.

Lorentzen, P. (2008): “Regularized Rioting: Strategic Toleration of Popular Protest in China,” University of California, Berkeley.

Loury, G. (1994): “Self-Censorship in Public Discourse: A Theory of ŚPolitical CorrectnessŠ and Related Phenomena,” *Rationality and Society*, 6(4), 428–61.

Morris, S. (2001): “Political correctness,” *Journal of Political Economy*, 109(2), 231–265.

Ottaviani, M., and P. Sørensen (2001): "Information aggregation in debate: who should speak first?," *Journal of Public Economics*, 81(3), 393–421.

Posner, E. A. (1998): "Symbols, signals, and social norms in politics and the law," *The Journal of Legal Studies*, 27(S2), 765–797.

——— (2009): *Law and social norms*. Harvard university press.

Prat, A. (2005): "The wrong kind of transparency," *American Economic Review*, pp. 862–877.

Prendergast, C. (1993): "A theory of" yes men"," *The American Economic Review*, pp. 757–770.

Prendergast, C., and L. Stole (1996): "Impetuous youngsters and jaded old-timers: Acquiring a reputation for learning," *Journal of Political Economy*, pp. 1105–1134.

Reeves, R. (2013): "Shame is not a four-letter word," *New York Times*, p. A21.

Ronson, J. (2015): "How One Stupid Tweet Blew Up Justine SaccoŠs Life," *The New York Times*.

Segal, D. (2013): "Mugged by a mug shot online," *New York Times*.

Sliwka, D. (2008): "Trust as a Signal of a Social Norm and the Hidden Costs of Incentives Schemes," *American Economic Review*, 97(3), 999–1012.

Vesterlund, L. (2003): "The Informational Value of Sequential Fundraising," *Journal of Public Economics*, 87(3-4), 627–657.

Visser, B., and O. Swank (2007): "On Committees of Experts," *The Quarterly Journal of Economics*, 122(1), 337–372.

Weele, van der, J. (2013): "The Signalling Power of Sanctions in Social Dilemmas," *Journal of Law, Economics and Organization*, 28(1), 103–25.