Image Versus Information:
Changing Societal Norms and Optimal Privacy

S. Nageeb Ali[1] Roland Bénabou[2]

This version: February 2016 [3]

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

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

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

Abstract

We analyze the costs and benefits of using social image to foster virtuous behavior. A Principal seeks to motivate reputation-conscious agents to supply a public good. Each agent chooses how much to contribute based on his own mix of public-spiritedness, private signal about the value of the public good, and reputational concern for appearing prosocial. By making individual behavior more visible to the community the Principal can amplify reputational payoffs, thereby reducing free-riding at low cost. Because societal preferences constantly evolve, however, she knows only imperfectly both the social value of the public good (which matters for choosing her own contribution, matching rate or legal policy) and the importance attached by agents to social esteem or sanctions. Increasing publicity makes it harder for the Principal to learn from what agents do (the "descriptive norm") what they really value (the "prescriptive norm"), thus presenting her with a tradeoff between incentives and information aggregation. We derive the optimal degree of privacy/publicity and the Principal's matching rate, then analyze how they depend on the economy's stochastic and informational structure. 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 generally be greater than in a more static one.

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?

This question is central to institutional design, and of growing policy importance. 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"; and charities publicizing donors' names on buildings and plaques. On the sanctions side, a number of U.S. states and towns 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, etc.); 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, automated licence-plate readers and other tracking technologies, the cost of widely disseminating what someone did, gave, took or even just

[1] On public goods, see, e.g., Ariely, Bracha, and Meier (2009), DellaVigna, List, and Malmendier (2012) or Algan et al. (2013); on voting, Gerber, Green, and Larimer (2008); on blood donors, 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 Madison'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.

said is rapidly falling to 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 (2007)) 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 general attachment to anonymous voting as indispensable to democracy (see, however, Brennan and Pettit (1990)), 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. Governments often expunge criminal records after some time or conceal them from private view (for instance, prohibiting credit bureaus from reporting past arrests), and a major debate is ongoing with search-engine and social-media companies over the *"right to be forgotten"* in the digital sphere. A broad 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 on identity theft and rivals seeking to steal trade secrets. 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.

[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.

1.2 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:

1. *Inefficient variability in the power of social image.* The rewards and sanctions generated by publicizing an individual's actions stem from the reactions that this knowledge 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 (2000)). 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, single-mothers, 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, enlightenment, migration, trade, 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.

Full reversals of societal preferences, where some behavior like overt racism, sexism or domestic violence goes from "normal" to deeply scorned, or on the contrary from intensely stigmatized to widely acceptable, like divorce, cohabitation ("living in sin"), homosexuality or drug use, are relatively common in modern societies and sometimes quite sudden. It is

[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 a shaming that "went too far" are themselves publicly shamed on the same media (Ronson (2015)).

[5] Similar effects of variability 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.

all but certain that some conducts generally seen as abhorrent and shameful today will also become perfectly mundane within a couple of decades, or vice-versa. Uncertainty lies only in which ones it will be and which way the cursor will move. Possible candidates of both types include organ sales, prostitution, extramarital and other parallel relationships, eating meat and wearing animal products, atheism and apostasy, transhumanistic enhancements and many more we cannot even yet conceive of.

Even for behaviors that remain unambiguously bad or good from a social point view (drunk driving, not evading taxes, etc.), moreover, the tradeoff we identify remains. As long as their *relative* importance to social welfare fluctuates over time, a policymaker will need to learn of these evolutions through shifts in aggregate behavior, so as to appropriately redirect her limited financial or enforcement resources.

Formally, we study a Principal interacting with a continuum of agents in a canonical context of public-goods-provision or externalities. Agents have 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, 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 precisely 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 by making aggregate behavior more sensitive to variations in the importance of social esteem. The Principal thus faces a direct tradeoff between using *image as incentives* and gaining better *information* on societal preferences.[6]

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

[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 high-powered incentives, of any kind, can be counterproductive.

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

Social norms and 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"*).

The issue is also relevant to the debate over freedom of speech versus "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 socially desirable behaviors and sanction undesirable ones, using publicity as well as rules or contracts.[7] Here again, the tradeoff is that insufficient individual privacy may prevent the institution-designer from learning what people have really come to think.

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

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

[7] Thus, 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 public buses in the immediate neighborhood of the source (with names and profile pictures blurred, however).

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.

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, and choose 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 (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 (Kreps (1990), Hermalin and Katz (2006), Bolton, Brunnermeier, and Veldkamp (2013)). Our analysis highlights how a leader also faces the dual challenge of using publicity to align agents' goals and values with the of the organization, encourage agents to serve the organization's goals and values, while allowing enough dissent and contrarian behavior for her (and others) to learn how these should adapt 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, and thus ultimately on institutional design.

A first strand focuses on signaling in a public-goods context.[9] Our setting builds on Bénabou and Tirole (2006) who study how incentives, whether material or social, can undermine individual's reputational returns derived from a prosocial activity. We develop this basic framework in two important, more aggregate 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 im-

[8] Lorentzen (2013), 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.

[9] See, e.g., Bernheim (1994), Corneo (1997), Harbaugh (1998), Ellingsen and Johannesson (2008) and Andreoni and Bernheim (2009).

age 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 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 aim to be in line with the Principal's prior (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 then depends on the finer details on game's information structure.[12] On the normative side, which 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 (Fox and Van Weelden (2012), Fehrler and Hughes (2015)). 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, is here one of the key sources of uncertainty.[13]

[10] Excessive constraints on behavior (commitment devices, monitoring with threats of punishment) can also interfere with learning (or self-learning) about agents' individual types, rather than with information aggregation over the state of the world; see, e.g., Bénabou and Tirole (2004), Ichino and Muehlheusser (2008) and Ali (2011).

[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 notable event) is known by the Principal, material incentives such as tax deductibility can be adjusted to offset such reputation-motivated distortions in the level of contributions or their allocation toward more highly visible domains. This is another reason why, in the present model, the fact that the Principal does not know the exact value of image is important.

[12] In particular, it can 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)).

[13] Bénabou and Tirole (2006) study signaling agents with heterogenous (privately known) image-concerns, and Fischer and Verrecchia (2000) and Frankel and Kartik (2014) agents with heterogenous payoffs to misrepresenting their actions; in such settings, greater transparency makes each individual's observed behavior less informative about his true motivations. In none of these papers is there any aggregate uncertainty, nor a Principal who seeks to incentivize agents and learn from their behavior what decision she should make, as is our primary focus here.

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; see Acquisti, Taylor, and Wagman (2016) for a survey. Our focus is instead on what private citizens know about each other's behaviors, on the social value of privacy even when the Principal is benevolent, and more generally on how her learning problem (whatever her preferences might be) affects the 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.[14]

The paper is organized as follows. Section 2 develops the general model. Section 3.1 focuses, for ease of exposition, on the simpler case where agents share a common reputational or social-enforcement concern, thus emphasizing the role of aggregate shocks to societal preferences. It first solves for agents' equilibrium responses to a given level of publicity (observability of one's actions by others), then derives the Principal's optimal choices of this visibility level and of her own contribution or matching rate. Section 4 extends the analysis to allow for heterogeneity in individual's image concerns, and Section 5 characterizes the comparative statics of the optimal policies with respect to the economy's entire informational and stochastic structure. Section 6 outlines further extensions and concludes. All proofs are gathered in the Appendix.

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

[14] See, for example, Posner (1998) for such arguments and Bénabou and Tirole (2011) for an analysis of expressive law, including the case of "cruel and unusual punishments. " These may be genuine concerns 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.

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.[15] 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. Here and throughout the paper, we use the following mnemonics: aggregate variabilities are denoted as $\sigma_\cdot^2$, cross-sectional dispersions as $s_\cdot^2$.

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.[16] The importance of maintaining a good reputation typically varies across people, communities and time periods, as well as with institutional choices of how much visibility and recognition to accord individual actions. Signaling prosociality is thus more important for people engaged in long-run relationships and in settings where most transactions rely on trust than where exchange occurs primarily through impersonal markets and 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 hard to predict. We denote the strength of agent i's reputational concerns as μ_i (specifying below how it affects his payoffs) and allow it to be distributed cross-sectionally as $N(\mu, s_\mu^2)$ around the group average μ, which itself varies as $N(\bar{\mu}, \sigma_\mu^2)$ around a common prior $\bar{\mu}$ held by agents and Principal alike. We assume that $\bar{\mu}$ is large enough that, with very high probability, the fraction of agents who desire a positive reputation is very close to 1.

Formally, an agent i's complete type is a triplet (v_i, θ_i, μ_i); for tractability, we take the three components to be independent of each other.[17] Another individual j observing his contribution a_i does not know to what extent it was motivated intrinsically (high v_i), by a high signal realization about the value of the public good (high θ_i) or a strong image motive (high μ_i), but he can use his own signal θ_j and reputational concern μ_j (since (θ_i, θ_j) and (μ_i, μ_j) are correlated),

[15] 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.

[16] 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.

[17] In particular, if μ_i was correlated with v_i or θ_i the inference problems of agents and Principal would no longer have a linear-normal structure.

as well as the realized average contribution $\bar{a}$, to form his assessment $E[v_i|a_i, \bar{a}, \theta_j, \mu_j]$ of player i. Thinking ahead, Agent i uses his ex-ante information 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

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

We assume that from a social image $R(a, \theta_i, \mu_i)$, agent i obtains a net payoff of $\mu_i x[R(a, \theta_i, \mu_i) - \bar{v}]$, where μ_i reflects his baseline concern for social esteem 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, agent i chooses a_i to solve

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

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 "joy of giving" utility, relative to their material payoffs. As to image gains and losses, by Bayes' rule (or the law of iterated expectations) they sum to zero across agents ($\int_0^1 R(a_i, \theta_i) \, di = \bar{v}$), so whether or not the Principal internalizes this part of aggregate utility 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 private contributions (or norm compliance). To cut down on the number of cases we shall focus the exposition on the case where $\eta > 0$, which in turn implies that $\varphi > 0$ and $\partial\omega/\partial\bar{\theta} = \lambda\alpha + (1-\lambda)\eta > 0$, meaning that "higher quality" is indeed something that the Principal values positively. Her preferences over the quality of the public good are thus congruent with those of the agents, even though her preferences over the level and sharing of its supply may be quite different.[18]

Our framework includes as special cases:

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

(b) 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, e.g., due to a shadow price of public funds.)

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

In order to set her own provision a_P efficiently, the Principal must learn about θ. A key piece of data she observes is the aggregate contribution or compliance rate $\bar{a}$, which embodies information about both aggregate shocks, θ and μ, generating a signal-extraction problem. The Principal shares agents' prior $\theta \sim N\left(\bar{\theta}, \sigma_\theta^2\right)$ about the quality of the public good and may also obtain an independent signal $\theta_P \equiv \theta + \varepsilon_P$, with error distributed as $N(0, s_{\theta,P}^2)$. Her prior for the importance of image is $N(\bar{\mu}, \sigma_\mu^2)$. 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.[19]

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 how important social esteem is to him, μ_i, 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.

[18] The model and all analytical results also allow for $\eta < 0$ (even potentially $\varphi < 0$, $\omega < 0$ and $\partial\omega/\partial\bar{\theta} < 0$), however. This corresponds to Principal who intrinsically dislikes the activity that agents consider socially valuable –political opposition, cultural resistance, etc.

[19] This information is typically limited: polling is costly (see Auriol and Gary-Bobo (2012) on the optimal sample size or number of representatives) and invites strategic responses from agents who would like the Principal to contribute more (Morgan and Stocken (2008), Hummel, Morgan, and Stocken (2013)). Allowing the Principal to obtain an independent, noisy signal of μ would also not affect our analysis.

We focus, for tractability, on Perfect Bayesian Equilibria in which an agent's contribution is linear in his type, (v_i, θ_i, μ_i). This will be shown to imply that an equivalent formulation of the Principal's decision problem is:

(a) Given any x, optimally set a baseline investment level she will provide (based on her own signal) 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. Agent's contributions become driven in larger part by variations in their social-image concerns, rather than by their signals concerning the social value of the public good.

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 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 high intrinsic motivation v_i, favorable information θ_i and strong image concerns μ_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 solutions. The basic tradeoff between incentives and information identified here would, however, apply even with complementarity between these dimensions.

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 from that 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.

[20] Papers studying an informed-principal problem in related contexts include Bénabou and Tirole (2003), Sliwka (2008), Van der Weele (2013) and Bénabou and Tirole (2011).

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 is negligible in 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 matching.[21] For simplicity, and to focus squarely on the effects of publicity, we abstract from the use of price incentives to motivate agents. More generally, monetary incentives entail both direct and indirect costs that limit the extent to which they can be used to induce compliance.[22]

3 Equilibrium Behavior and Optimal Publicity

In this section we focus on the case in which *all agents share the same value for social image:* $\mu_i = \mu$ for every i, or equivalently, $s_\mu^2 = 0$. This simplifying assumption, almost universal in the literature on signaling, will most clearly highlight the role of *aggregate* variability in reputational concerns, which is key to the tradeoff between incentives and learning faced by the Principal.

We proceed in two steps. First, we 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. Then, 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:

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

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, $\partial R(a, \theta_i, \mu)/\partial a$ is constant, leading to a unique outcome.

[21] There is no "right answer" on what these preferences should be: the limited evidence on this question. 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.

[22] We can thus also define ω as the wedge left after the Principal has already used any standard incentives at her disposal. Note also that it will always be optimal to use some positive level of publicity as an additional incentive: the gain is initially first-order, whereas the induced distortions are second-order.

Proposition 1. *In the unique linear equilibrium, an agent of type* (v_i, θ_i) *chooses:*

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

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

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. The reputational return ξ corresponds to the *signal-to-noise ratio* faced by an *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).

Benchmarking. 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. Therefore, they judge a given individual identically: $E[v_i \mid a_i, \bar{a}, \theta_j]$ is 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:

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

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

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

where ξ is given by (10). Consequently $\partial E[v_i \,|\, a_i, \bar{a}]\,/\partial a_i = \xi$ measures the marginal improvement in social image that an additional unit of contribution will buy.[23] Rewriting this image return as

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

yields the following result.

Proposition 2. *Let* $\mu > 0$. *Reputational incentives and equilibrium contributions are increasing in the dispersion of agents' preferences* s_v^2, *decreasing in aggregate their variability* σ_θ^2 *and U-shaped in the quality of their information* s_θ^2.

The restriction $\mu > 0$ means that agents want to perceived as prosocial, rather than antiso-

[23] Equation (9) also shows that visibility leads all agents to raise their contributions by the same amount, which comes from the payoff of image being linear and (in this section only) independent of type.

cial; since $\bar{\mu}$ is taken to be very large, this case occurs with probability very close to 1.

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 more novel and subtle: the U-shape in s_θ^2 reflects the idea that reputational effects are strongest when agents have the same *interim* belief about the 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: $E[v_i \mid a_i] = v_i$. 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 ensure that total public good provision is efficient (from her standpoint). 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 $x\mu R(a, \theta_i)$. To focus on how the cost of publicity –or conversely, the *social value of privacy*– arises endogenously, 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.[24]

We uncover three distinct motivations for the Principal to grant agents some degree of privacy, and to isolate each effect, we consider in turn:

(a) A simple benchmark without any variability in image motives, $\sigma_\mu^2 = 0$.

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

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

These three nested cases provide insights into how the Principal would set publicity if she could fine-tune its impact $x\mu$ perfectly, the "variance effect" that emerges when she cannot

[24] This cost advantage is one of the main arguments put forward by proponents of publicity and shame (e.g., Kahan (1996), Brennan and Pettit (2004), (Jacquet (2015),) 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 rather than increasing it.

do so but observes μ *ex post* and, finally, the "information-distortion effect" that arises when publicizing behavior generates a signal-extraction problem.

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

Consider first the simple case 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 contribution $\bar{a}$, the Principal perfectly infers 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

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

where the superscripts stands for "First Best" and the wedge $\omega > 0$ was defined in (6).[25]

Image-based Pigovian policy. Consider in particular 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, $\sigma_\mu^2 > 0$, 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

[25] This result a special case in the proof of Proposition 3 below.

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 (A.2), we focus here on the corresponding optimality condition, which embodies two opposing effects:

$$\frac{dE\tilde{V}(x)}{dx} = \underbrace{(\xi\bar{\mu})\,\omega}_{\text{Incentive Effect}} - \underbrace{\lambda x\xi^2\left(\bar{\mu}^2 + \sigma_\mu^2\right)}_{\text{Variance Effect}}. \tag{18}$$

The two terms clearly show the tradeoff between leveraging social pressure to promote compliance and the inefficient, image-driven variations in aggregate contributions that arise from greater publicity. 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*

$$x^{SI} = \frac{\bar{\mu}\omega}{\lambda\xi\left(\bar{\mu}^2 + \sigma_\mu^2\right)} = \frac{x^{FB}}{1 + \sigma_\mu^2/\bar{\mu}^2}, \tag{19}$$

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 $\bar{v}$, 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, as emphasized by Posner (2000), so the Principal naturally wields it more cautiously than under the Pigovian policy: $x^{SI} < x^{FB}$, for all $\lambda > 0$.

3.2.3 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: the overall contribution or compliance rate $\bar{a}$ reflects both public-good quality θ and social-enforcement concerns, μ. Using her *expected* value of μ to invert (11), she now obtains 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 compliance $\bar{a}$. After observing $\bar{a}$, the Principal optimally sets $a_P = \varphi\left(w + E\left[\theta|\bar{a}, a_P\right]\right)/(1-\lambda)k_P$; substituting in (20) and (23) yields:

Proposition 4. *The Principal's contribution policy is equivalent to setting a baseline investment $\underline{a}_P(x, \theta_P)$ (given in the Appendix) 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 in the Appendix (A.6) 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

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

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*

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

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.[26] 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, which we shall analyze in Section 5 for the more general model where agents may differ in how they value reputation.

4 Allowing for Heterogeneous Image Concerns

4.1 Agents' Behavior and Inference

We have so far assumed, for expositional simplicity, that all agents have identical reputational concerns ($s_\mu^2 = 0$). In reality some people care more about their social image than others, and this introduces another source of uncertainty about what accounts for someone's contribution: to what extent was he intrinsically motivated to do good or just seeking to improve his image? This additional *overjustification effect* reduces the reputational return to contributing, thus dampening the direct effect of publicity as an incentive to contribute to the public good. It is therefore far from obvious *a priori* how heterogeneity in image concerns will affect the Principal and her optimal policies.[27]

Recall that in the general setting with $s_\mu^2 > 0$, the values for reputation μ_i and μ_j of any two agents i and j share a common component, μ. Neither can observe this common component, but each knows that it is drawn from $N(\bar{\mu}, \sigma_\mu^2)$, and that his own μ_i or μ_j is drawn from $N(\mu, s_\mu^2)$. An agent i's information set (or type) when choosing his contribution a_i thus comprises his warm-glow motivation v_i, his signal θ_i about the quality of the public good, and his reputational concern μ_i. When evaluating the type of some other individual he also uses their observed contribution level a_j as well as the aggregate $\bar{a}$, which will again serve as an informative "standard."

[26] 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.

[27] Bénabou and Tirole (2006) analyze this overjustification effect in a setting without aggregate shocks to θ orμ, nor any learning and optimizing Principal.

Proposition 6. *For every $x \geq 0$, there exists a unique linear equilibrium in which an agent of type (v_i, θ_i, μ_i) chooses*

$$a_i(v_i, \theta_i, \mu_i) = v_i + \rho\theta_i + (1-\rho)\bar{\theta} + \mu_i x\tilde{\xi}(x), \tag{30}$$

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

$$\tilde{\xi}(x) = \frac{s_v^2}{x^2\tilde{\xi}(x)^2 s_\mu^2 + s_v^2 + \rho^2 s_\theta^2}. \tag{31}$$

The resulting aggregate contribution is

$$\bar{a}(\theta; \mu) = \bar{v} + \rho\theta + (1-\rho)\bar{\theta} + \mu x\tilde{\xi}(x). \tag{32}$$

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

The interpretation of $\tilde{\xi}(x)$ is identical to that of ξ in Proposition 1: it measures the marginal impact that an additional unit of contribution has on one's image, given the equilibrium decision rule (30). In anonymous settings, $\tilde{\xi}(0) = \xi$, but as soon as there is some visibility $x > 0$, $\tilde{\xi}(x) < \xi$. This reflects the overjustification effect from heterogeneity in publicity-seeking motives, which gets amplified when actions become more visible, resulting in a *partial crowding out:* $\beta(x) \equiv x\xi(x)$ increases *less than one for one* with x. For the same reason, and in contrast to the case of a common $\mu_i = \mu$, the reputational return $\xi(x)$ is now determined as a fixed point of equation (31), which depends on x.

A sufficient-statistic result. Most remarkable is the *simplicity* of the social-image computations that emerge from this complex setting: the expected reputational return $\tilde{\xi}(x)$ is the *same for all agents,* even though they have different signals (θ_i, μ_i) that are predictive of the average θ and μ, hence also of the θ_j's and μ_j's which observers will have at their disposal to extract v_i from a_i, using (30). The reason for this surprising result is *a form of benchmarking:* an observer j does not need to separately estimate and filter out the contributions of θ_i and μ_i to a_i, but only that of the linear combination $\rho\theta_i + \mu_i x\tilde{\xi}(x)$, and for this purpose $\rho\theta + \mu x\tilde{\xi}(x)$, hence also $\bar{a}$, is a *sufficient statistic.*[28]

Put differently, whereas $E[\theta_i | a, \bar{a}, \theta_j, \mu_j]$ and $E[\mu_i | a, \bar{a}, \theta_j, \mu_j]$ both depend on j's private type, $E[a_i - \rho\theta_i - \mu_i x\tilde{\xi}(x) | a_i, \bar{a}, \theta_j, \mu_j]$ does not, so all observers of agent i again share the *same beliefs* about his motivation: $E[v_i | a, \bar{a}, \theta_j, \mu_j] = E[v_i | a, \bar{a}]$. Agent i's reputation will thus be a linear function of $a_i - \bar{a}$ only, implying in turn that his own (θ_i, μ_i), while critical to forecast $\bar{a}$ itself, will not affect the marginal return: $\partial R(a, \theta_i, \mu_i)/\partial a = \tilde{\xi}(x)$.

[28] Such would no longer be the case if $\bar{a}$ itself was observed with noise, or subject to small-sample variations with a finite number of agents.

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 all remain the same as in (24), (25) and (27) respectively, except that $x\xi$ is replaced everywhere by $x\tilde{\xi}(x)=\beta(x)$.*

4.2 Optimal Publicity

Relegating derivations to the Appendix, the marginal effect of publicity on the Principal's payoff now takes the form

$$\frac{1}{\beta'(x)}\frac{dEV(x)}{dx}=\omega\bar{\mu}-\lambda\beta(x)\left(\bar{\mu}^2+\sigma_\mu^2+s_\mu^2\right)-\frac{\varphi^2\sigma_\mu^2}{\rho^2(1-\lambda)k_P}\beta(x)\gamma(x)^2. \tag{33}$$

Setting it to equal 0 and noting that $\beta(x)=x\xi(x)$ yields the following results.

Proposition 7. *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*

$$x^*=\left(\frac{\bar{\mu}}{\xi(x^*)}\right)\left[\frac{\omega}{\lambda(\bar{\mu}^2+\sigma_\mu^2+s_\mu^2)+\frac{(\varphi\sigma_\mu\gamma(x^*)/\rho)^2}{(1-\lambda)k_P}}\right], \tag{34}$$

where $\xi(x)$ is given by (31) and $\gamma(x)$ remains given by (24). The solution is thus identical to that in Proposition 5, except, that σ_μ^2 is replaced by $\sigma_\mu^2+s_\mu^2$ and ξ by $\xi(x)$ everywhere.

As before, (34) could have multiple solutions but all stable ones, including the global optimum x^*, are below x^{SI} (the level of visibility in the "symmetric uncertainty" benchmark) and share the same comparative-statics properties, to which we now turn.

5 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. We use the first-order condition (33) to derive cross-partials. Observe that $\partial^2 EV/\partial x\partial\omega>0$, and therefore, EV has positive cross-partials in (x,w), $(x,\bar{\theta})$, (x,α), and $(x,-\bar{v})$. Analogously, $\partial^2 EV/\partial x\partial k_P>0$, leading to the results summarized in Table I below. These properties are quite intuitive. For instance, a principal who faces a higher costs of own funds, or who internalizes agents' warm-glow utility, wants to encourage private contributions. She therefore makes behavior more observable and, as it becomes less informative, also reduces her matching rate.

		Optimal publicity x^*	Optimal matching rate m^*
Baseline externality	w	Increasing	Decreasing
Ex ante expected quality	$\bar{\theta}$	Increasing	Decreasing
Weight on agents' warm-glow	α	Increasing	Decreasing
Average intrinsic motivation	$\overline{v}$	Decreasing	Increasing
Principal's relative cost	k_P	Increasing	Decreasing

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

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, and this has both costs and benefits for the Principal. To the extent that she weighs agents' warm glow positively she appreciates variability, but on the other hand suffers from internalizing its effect on their total contribution cost.[29]

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 $\xi(x)$ and therefore the reputational incentive to contribute, $\beta(x) = x\xi(x)$. For a fixed publicity x, this affects all three components of the Principal's trade-off: 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 $\xi(x)$ and x enter each term in EV only through the product $x\xi(x)$ we can think of the Principal as *directly optimizing over* the value of $\beta(x)$. Changing s_v^2 therefore only has a direct effect on her payoff. For the same reason, 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 $\beta(x)$ constant. Since s_v^2 influences γ and m only through the value of $\beta(x)$, both remain unchanged.

Proposition 8. *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 second effect and highlight the Principal's tradeoff between raising $\bar{a}$ and learning about θ, let us focus here on the limiting case in which agents' private signals are perfect, or more generally far more informative than their prior, so that $s_\theta/\sigma_\theta \approx 0$ or, equivalently, $\rho \approx 1$. In this case, $\tilde{\xi}(x)$ becomes independent of σ_θ^2, which then enters (33) only by raising $\gamma(x)$, through its effect on $s_{\theta,P}^2$; see (31), (21) and (24). Therefore:

[29] 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 $(-1/2)\int_0^1 a_i^2 d_i$.

Proposition 9. *When agents' private signals about the quality of the public good are sufficiently more precise than their prior over it ($s_\theta^2/\sigma_\theta^2$ small enough), the optimal degree of visibility x^* is decreasing in σ_θ^2, the variability of this quality, while the optimal matching rate γ^* is increasing in it.*

D. Variability in the importance of social image or social enforcement.

1. Average social image concern. An increase in σ_μ^2 does not affect ρ or $\xi(x)$ 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 $\sigma_\mu^2\beta(x)\gamma^2(x)$, which can be seen from (24) to be hump-shaped in σ_μ^2 for a fixed x.

Somewhat surprisingly, the Principal may thus use *more publicity* when the source of "noise" in her learning problem 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. In what follows we shall denote $\bar{k}_P \equiv \varphi^2/\left[27\lambda(1-\lambda)\rho^2\right]$.

Proposition 10. *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 $k_P \geq \bar{k}_P$, 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 $+\infty$, x^* tends to 0 (full privacy), while as σ_μ tends to 0, x^* approaches the first-best level x^{FB} that solve $\beta(x) = \bar{\mu}\omega/\left[\lambda\left(\bar{\mu}^2 + s_\mu^2\right)\right]$.*

2. Heterogeneity in image concerns. An increase in s_μ^2 magnifies the variance effect (second term in (33)) but, for given x, influences the image incentive $\beta(x)$ in complex ways (see (31)). As a result, its net effect on the optimal degree of publicity and matching rate is ambiguous. Due to an envelope-theorem result, however, the impact on the Principal's payoff is always negative.

Proposition 11. *The Principal's expected payoff is strictly decreasing in s_μ^2.*

E. Precision of private signals

1. Principal's signal. When the variance $s^2_{\theta,P}$ 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 $s^2_{\theta,P}$ appears only in the information-distortion effect, through γ. Therefore

$$\frac{\partial^2 EV(x)}{\partial x \partial s^2_{\theta,P}} = -\frac{2\varphi^2\sigma^2_\mu \beta'(x)\beta(x)\gamma(x)}{\rho^2(1-\lambda)k_P}\left(\frac{\partial \gamma}{\partial s^2_{\theta,P}}\right) < 0.$$

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 $s_{\theta,P}$ and decreases with x, 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. This argument establishes the following result.

Proposition 12. *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 reduces 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 again say more when agents' common prior over θ is far less informative than their private signal: as $s_\theta/\sigma_\theta \to 0$, ρ approaches 1 and the Principal's optimality condition (33) involves x and ξ only through their product $\beta(x) = x\tilde{\xi}(x)$, while s^2_θ enters it only through $\tilde{\xi}(x)$. It follows that the optimal value of β is independent of s_θ, while the associated x must rise with it so as to maintain that constancy. Therefore, we have:

Proposition 13. *When agents' private signals about the quality of the public good are far more precise than their prior over it ($s^2_\theta/\sigma^2_\theta$ sufficiently small), 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 is independent of it.*

Table II below summarizes the results from the preceding four propositions.

	Optimal publicity x^*	Optimal matching rate m^*
s^2_v	Decreasing	Invariant
σ^2_θ	Decreasing, for s_θ/σ_θ sufficiently small	Increasing, for s_θ/σ_θ sufficiently small
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 s_θ/σ_θ sufficiently small	Invariant

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

6 Conclusion

We studied the tradeoff between the incentive benefit of publicizing individual behaviors that constitute public-goods (or bads) and the costs which reduced privacy imposes on society (or any other Principal) when the overall distribution of preferences is subject to unpredictable shifts and evolutions.

First, such imperfect knowledge renders publicity hard to fine-tune, generating inefficient variations in both individual and aggregate behavior. Second, leveraging social-image concerns makes it even harder to infer from prevailing norms the true social value of the public good or conduct in question. Among other results, we thus showed that where societal attitudes (what behaviors agents regard as socially desirable or undesirable) and/or technologies for monitoring and norms enforcement (means of communication and coordination, e.g., social media) are prone to significant change, a higher degree of privacy is optimal: policy-makers can then better learn, by observing overall compliance, how taxes and subsidies, the law or other institutions should be adapted. When preferences over public goods and reputation remain or have become relatively stable, conversely, visibility 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, and possibly also the strength of reputational concerns, 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 alleviate the informational problem we identify. This could for instance involve a two-stage procedure, in which agents first choose their participation levels anonymously –thereby revealing the state– then, in a second stage, are asked to contribute. Such dynamic procedures may lead to efficiency gains because: (a) information is better 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 mechanisms may not be feasible in all contexts.

7 Appendix

Proof of Proposition 1 on p. 13

The result is a special case (for $s_\mu^2 = 0$) of Proposition 7, the proof of which is given further below. ■

Proof of Proposition 3 on p. 17

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):

$$E\tilde{V}(x) = \lambda\left[\alpha\left(s_v^2 + \rho\sigma_\theta^2 + (\bar{v}+\bar{\theta})(\bar{\bar{a}})\right) + (w+\bar{\theta})(\bar{\bar{a}} + \bar{a}_P) + \rho\sigma_\theta^2 + \frac{\sigma_\theta^2\varphi}{(1-\lambda)k_P}\right]$$

$$+ (1-\lambda)\eta\left[(w+\bar{\theta})(\bar{\bar{a}} + \bar{a}_P) + \rho\sigma_\theta^2 + \frac{\sigma_\theta^2\varphi}{(1-\lambda)k_P}\right] \tag{A.1}$$

$$- \frac{\lambda}{2}[\bar{\bar{a}}^2 + \rho^2(\sigma_\theta^2 + s_\theta^2) + s_v^2 + x^2\xi^2\sigma_\mu^2] - \frac{(1-\lambda)k_P}{2}\left[\bar{a}_P^2 + \sigma_\theta^2\left(\frac{\varphi}{(1-\lambda)k_P}\right)^2\right]. \tag{A.2}$$

Differentiating yields:

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

$$= \omega\xi\bar{\mu} - \lambda x\xi^2(\bar{\mu}^2 + \sigma_\mu^2). \tag{A.3}$$

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 $\sigma_\mu^2 = 0$. ■

Proof of Proposition 4 on p. 18

The formula for $m(x)$ follows directly from the reasoning in the text. As to the baseline investment,

$$\underline{a}_P(x,\theta_P) = \frac{\varphi(w + (1-\gamma(x))E[\theta|\theta_P] - \frac{\gamma(x)}{\rho}(\bar{v} + x\xi\bar{\mu} + (1-\rho)\bar{\theta}))}{(1-\lambda)k_P}, \tag{A.4}$$

it follows from the same equations, together with (21). ■

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 $\Delta \equiv E\left[\theta|\bar{a},\theta_P\right] - \theta$,

derived in (26), generates inefficient deviations from full-information optimality. Note that

$$E\left[\Delta^2\right] = (1-\gamma)^2\sigma_{\theta,P}^2 + (\gamma\xi x/\rho)^2\sigma_\mu^2 = \sigma_{\theta,P}^2\left[(1-\gamma)^2 + \gamma^2\left(1/\gamma - 1\right)\right] = \sigma_{\theta,P}^2\left(1-\gamma\right), \quad \text{(A.5)}$$

where we abbreviated $\gamma(x)$ as γ and used the fact that $x^2\xi^2\sigma_\mu^2/\rho^2 = \sigma_{\theta,P}^2\left(1-\gamma\right)/\gamma$.

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

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

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

$$\frac{(1-\lambda)k_P}{2}\left[\left(\frac{\varphi}{(1-\lambda)k_P}\right)^2\sigma_{\theta,P}^2\left(1-\gamma\right)\right] = \frac{\varphi^2\sigma_{\theta,P}^2}{2(1-\lambda)k_P}(1-\gamma). \quad \text{(A.6)}$$

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 (A.2) remain unchanged from the case where she knows μ, and (27) thus characterizes the change in payoffs. Note also that

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

Therefore

$$\begin{aligned}\frac{\partial EV}{\partial x} &= \frac{\partial E\tilde{V}}{\partial x} - \frac{\varphi^2}{(1-\lambda)k_P}\left(\frac{\sigma_\mu^2\gamma^2\xi^2 x}{\rho^2}\right) \\ &= \left(\xi\bar{\mu}\right)\left[(w+\bar{\theta})\varphi - (\bar{v}+\bar{\theta})(1-\alpha)\lambda\right] - \lambda x\xi^2\left(\bar{\mu}^2+\sigma_\mu^2\right) - \frac{\varphi^2}{(1-\lambda)k_P}\left(\frac{\sigma_\mu^2\gamma^2\xi^2 x}{\rho^2}\right), \quad \text{(A.7)}\end{aligned}$$

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.

Consider linear strategies of the form $a_i = A\mu_i + Bv_i + C\theta_i + D$, implying that $\bar{a} = A\mu + B\bar{v} + C\theta + D$. We first establish the following result.

Claim 1 (Benchmarking). *The expectation $E[v_i|\theta_j, \mu_j, \bar{a}, a_i]$ is independent of (θ_j, μ_j) and equal*

to:

$$E[v_i|\theta_j,\mu_j,\bar{a},a_i]=\bar{v}+\frac{Bs_v^2}{B^2s_v^2+C^2s_\theta^2+A^2s_\mu^2}(a_i-\bar{a}). \tag{A.8}$$

Proof. Subtracting $\bar{a}$ from a_i, and re-arranging, we obtain $Bv_i = B\bar{v}+(a_i-\bar{a})-(C\varepsilon_i^\theta+A\varepsilon_i^\mu)$, where ε_i^θ and let ε_i^μ denote $\theta_i-\theta$ and $\mu_i-\mu$ respectively. Observe that $(Bv_i, a_i-\bar{a}, \bar{a}, \theta_j, \mu_j, C\varepsilon_i^\theta+A\varepsilon_i^\mu)$ is jointly normally distributed: every linear combination of these components is a linear combination of a set of independent normal random variables, and therefore has a univariate normal distribution. Because $\bar{a}$, θ_j, and μ_j are uncorrelated to both $C\varepsilon_i^\theta+A\varepsilon_i^\mu$ and $a_i-\bar{a}$, and these variables are jointly normally distributed, it follows from independence that

$$E[C\varepsilon_i^\theta+A\varepsilon_i^\mu|a_i,\bar{a},\theta_j,\mu_j]=E[C\varepsilon_i^\theta+A\varepsilon_i^\mu|a_i-\bar{a}].$$

Observe that

$$\begin{pmatrix} v_i \\ a_i-\bar{a}\end{pmatrix}\sim N\left(\begin{pmatrix}\bar{v}\\0\end{pmatrix},\begin{pmatrix}s_v^2 & Bs_v^2\\ Bs_v^2 & B^2s_v^2+B^2s_\theta^2+A^2s_\mu^2\end{pmatrix}\right),$$

and therefore, $E[v_i|\theta_j,\mu_j,\bar{a}-a_i]$ equals the expression in (A.8). ■

From Claim 1 it follows that

$$\begin{aligned}R(a_i,\theta_i,\mu_i) &= E\left[E\left[v_i|a_i,\bar{a}\right]|\theta_i,\mu_i\right]\\ &= E\left[\left(\bar{v}+\frac{Bs_v^2}{A^2s_\mu^2+B^2s_v^2+C^2s_\theta^2}(a_i-\bar{a})\right)|\theta_i,\mu_i\right]\\ &= \bar{v}+\frac{Bs_v^2}{A^2s_\mu^2+B^2s_v^2+C^2s_\theta^2}[a_i-A\{\nu\mu_i+(1-\nu)\bar{\mu}\}-B\bar{v}-C\{\rho\theta_i+(1-\rho)\bar{\theta}\}-D],\end{aligned}$$

where $\nu\equiv\sigma_\mu^2/\left(\sigma_\mu^2+s_\mu^2\right)$. Utility maximization then yields the first-order condition:

$$a_i=v_i+\rho\theta_i+(1-\rho)\bar{\theta}+x\mu_i\left(\frac{Bs_v^2}{A^2s_\mu^2+B^2s_v^2+C^2s_\theta^2}\right). \tag{A.9}$$

Therefore, $B=1$, $C=\rho$, $D=(1-\rho)\bar{\theta}$, and $A=xs_v^2/\left(A^2s_\mu^2+s_v^2+\rho^2s_\theta^2\right)$. Substituting $A=x\tilde{\xi}(x)$ yields

$$\tilde{\xi}(x)=\frac{s_v^2}{x^2\tilde{\xi}(x)^2s_\mu^2+s_v^2+\rho^2s_\theta^2}. \tag{A.10}$$

It remains to show that for each choice of x, $\tilde{\xi}(x)$ is unique. Given x, $\tilde{\xi}(x)$ solves the equation

$$\tilde{\xi}=\frac{s_v^2}{x^2\tilde{\xi}^2s_\mu^2+s_v^2+\rho^2s_\theta^2}.$$

The right-hand side is continuous and decreasing in $\tilde{\xi}$, clearly cutting the diagonal at a unique solution $\tilde{\xi}(x)$. Furthermore, $\tilde{\xi}(x)$ must be strictly decreasing in x, strictly increasing in s_v^2, strictly decreasing in s_μ^2 and in σ_θ^2 and U-shaped in s_θ (noting that $\rho s_\theta=\sigma_\theta^2/[s_\theta+\sigma_\theta^2/s_\theta]$).

To derive comparative statics, note that $\beta(x) = x\tilde{\xi}(x)$ solves the implicit equation

$$x = \beta[\beta^2(s_\mu^2/s_v^2) + \rho^2 s_\theta^2/s_v^2 + 1],$$

which makes clear that $\beta(x)$ is strictly increasing in x, with $\lim_{x\to\infty}\beta(x) = +\infty$.■

Proof of Proposition 7 on p. 21

We proceed again in two stages, starting with the benchmark of "symmetric uncertainty" where the Principal, like the agents, learns the realization of (the average) μ after x has been set. Then, we incorporate the information-distortion effect.

Claim 2. *When the Principal faces no ex-post uncertainty about μ and observes it perfectly, she sets a publicity level $\tilde{x}^{SI}$ given by the unique solution to*

$$\tilde{x}^{SI} = \frac{\bar{\mu}\omega}{\lambda\tilde{\xi}\left(x^{SI}\right)\left(\bar{\mu}^2 + \sigma_\mu^2 + s_\mu^2\right)}, \tag{A.11}$$

which is lower than x^{FB} and strictly decreasing in s_μ^2. The aggregate contribution is

$$\bar{a}\left(\theta;\mu\right) = \bar{v} + \rho\theta + (1-\rho)\,\bar{\theta} + \frac{\mu}{\bar{\mu}}\left[\frac{\omega/\lambda}{1 + \left(\sigma_\mu^2 + s_\mu^2\right)/\bar{\mu}^2}\right], \tag{A.12}$$

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

Proof. Proposition 6 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 $\tilde{\xi}\left(x\right)$, 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 $a_i^0 \equiv v_i + \rho\theta_i + (1-\rho)\,\bar{\theta} + \mu x\tilde{\xi}\left(x\right)$ the value of a_i corresponding to the mean value of $\mu_i = \mu$, or equivalently the value of a_i in the original (homogeneous-μ) 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.2) 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.3) for $dEV(x)/dx$:

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

In the Principal's actual loss function (4), however, the heterogeneity in agents' μ_i's generates an additional loss due 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 observes the realization of μ, the optimal (symmetric-

information) value of x is given by the first-order condition

$$\frac{d\tilde{EV}}{dx} = \omega\bar{\mu}\beta'(x) - \lambda\beta'(x)\beta(x)\left(\bar{\mu}^2 + \sigma_\mu^2 + s_\mu^2\right) = 0, \tag{A.13}$$

or

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

which is equivalent to (A.11). Furthermore, the right-hand side is strictly decreasing in s_μ^2 and $\beta(x)$ was shown to be strictly increasing in x, so $\tilde{x}^{SI}$ must be decreasing in s_μ^2. Since $\bar{a}(\theta,\mu)$ is strictly increasing in $\beta(x_*)$ as long as $\mu > 0$, $\bar{a}$ is then decreasing in s_μ^2 for every θ and $\mu > 0$. ■

We now extend the results to the case where the Principal does not know the mean image concern μ when setting her contribution. Corollary 1 allows us to simply combine (A.13) and (27) to obtain the relevant version of her 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

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

which is equivalent to (34). ■

Proof of Proposition 8 on p. 22

Denote $x\xi(x)$ by z and note that $EV(x)$ can be reformulated as

$$\mathcal{V}(z) = s_v^2\left(\lambda\alpha - \frac{\lambda}{2}\right) + z\bar{\mu}\omega - \frac{\lambda}{2}z^2(\bar{\mu}^2 + \sigma_\mu^2 + s_\mu^2) - \frac{\varphi^2\sigma_{\theta,P}^2}{2(1-\lambda)k_P}\left[1 - \tilde{\gamma}(z)\right] + C, \tag{A.16}$$

in which

$$\tilde{\gamma}(z) \equiv \frac{\rho^2\sigma_{\theta,P}^2}{\rho^2\sigma_{\theta,P}^2 + z^2\sigma_\mu^2}, \tag{A.17}$$

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

$$\bar{\mu}\omega - \lambda z(\bar{\mu}^2 + \sigma_\mu^2 + s_\mu^2) + \frac{\varphi^2\sigma_{\theta,P}^2}{2(1-\lambda)k_P}\tilde{\gamma}'(z) = 0. \tag{A.18}$$

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 $x^*(s_v)\xi(x^*(s_v), s_v)$ is constant. This fact automatically implies that in equilibrium, changes in s_v^2 do not influence γ or the matching rate.

Because the Principal maintains constancy of $x^*(s_v)\xi(x^*(s_v), s_v)$, it follows from (A.10) that increases in s_v must strictly increase $\xi(x^*(s_v), s_v)$. Therefore, $x^*(s_v)\xi(x^*(s_v), s_v)$ remains unchanged only if $x^*(s_v)$ is decreasing in s_v. Finally, it follows from (A.16) that $d[EV(x^*(s_v^2); s_v^2)]/ds_v^2 = \lambda(\alpha - 1/2)$. ■

Proof of Proposition 9 on p. 23

Setting $\rho = 1$ in (33), $\partial^2 EV/\partial x \partial \sigma_\theta < 0$ implies that x^* is decreasing in σ_θ^2. Decreasing x decreases $\beta(x)$ (recall that in this limiting case, $\beta(x)$ is independent of σ_θ^2), so if $\gamma(x^*; \sigma_\theta)$ did not increase with σ_θ the right-hand-side of (33) could not remain equal to zero. Thus, at the optimal x^*, $\gamma(x^*; \sigma_\theta)$ must increase with σ_θ. ■

Proof of Proposition 10 on p. 23

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 (33) that

$$\frac{\partial^2 EV}{\partial x \partial \sigma_\mu^2} = -\lambda \beta'(x)\beta(x) - \frac{\varphi^2 \beta'(x)\beta(x)}{\rho^2(1-\lambda)k_P}\left(\gamma^2 + 2\gamma\sigma_\mu^2 \frac{d\gamma}{d\sigma_\mu^2}\right), \tag{A.19}$$

in which

$$\frac{\partial \gamma}{\partial \sigma_\mu^2} = -\frac{\rho^2 \sigma_\theta^2 \beta(x)^2}{\left(\rho^2\sigma_\theta^2 + \beta(x)^2\sigma_\mu^2\right)^2} = -\frac{\gamma\beta(x)^2}{\rho^2\sigma_\theta^2 + \beta(x)^2\sigma_\mu^2} = -\frac{\gamma(1-\gamma)}{\sigma_\mu^2}. \tag{A.20}$$

Thus,

$$\frac{\partial^2 EV}{\partial x \partial \sigma_\mu^2} = -\beta'(x)\beta(x)\left[\lambda - \frac{\varphi^2\gamma^2}{\rho^2(1-\lambda)k_P}(2\gamma - 1)\right]. \tag{A.21}$$

This expression is non-positive if and only if

$$\frac{\lambda(1-\lambda)\rho^2 k_P}{\varphi^2} \geq \gamma^2(1-2\gamma). \tag{A.22}$$

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.22) 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.22) holds, and $x^*(\sigma_\mu^2)$ decreases to 0. ■

Proof of Proposition 11 on p. 23

Since x enters EV only through $\beta(x) = x\tilde{\xi}(x)$, the Principal's problem is again equivalent to optimizing over the value of β, so the indirect effects of s_μ^2 on the optimized objective function $EV(x^*(s_\mu^2), s_\mu^2)$ cancel out at the first order, leaving only the direct (variance) effect $(-\lambda/2)\beta(x)^2 < 0$. ■

Proof of Proposition 13 on p. 24

As $\sigma_\theta \to \infty$, ρ converges to 1 and therefore, $\xi(x, s_\theta)$ converges to a solution to the equation

$$\xi = \frac{s_v^2}{x^2\xi^2 s_\mu^2 + s_v^2 + s_\theta^2}, \tag{A.23}$$

for each x. Note that this must be strictly decreasing in s_θ^2. By inspection, x and $\xi(x)$ enter all terms in (33) only through their product $\beta(x)$. Therefore, to study how the optimal $x^*(s_\theta)$ and the Principal's welfare depend on s_θ^2, we can follow the same steps as in he proof of Proposition 8, leading to $d\left[EV(x^*(s_\theta); s_\theta)\right]/ds_\theta = -\lambda s_\theta^2/2 < 0$. Finally, since the Principal keeps $x^*(s_\theta)\xi(x^*(s_\theta), s_\theta)$ constant as s_θ increases, it follows from (A.23) that $\xi(x^*(s_\theta), s_\theta)$ must decrease in s_θ. To compensate, $x^*(s_\theta)$ must then be increasing in s_θ. ■

References

Acquisti, Alessandro, Curis Taylor, and Liad Wagman. 2016. "The Economics of Privacy." *Journal of Economic Literature, forthcoming* .

Algan, Yann, Yochai Benkler, Mayo Fuster Morell, and Jérôme Hergueux. 2013. "Cooperation in a Peer Production Economy: Experimental Evidence from Wikipedia." In *Workshop on Information Systems and Economics*. 1–31.

Ali, S Nageeb. 2011. "Learning Self-Control." *Quarterly Journal of Economics* 126 (2):857–893.

Andreoni, James. 2006. "Leadership Giving in Charitable Fund-Raising." *Journal of Public Economic Theory* 8 (1):1–22.

Andreoni, James and B. Douglas Bernheim. 2009. "Social Image and the 50–50 Norm: A Theoretical and Experimental Analysis of Audience Effects." *Econometrica* 77 (5):1607–1636.

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

Auriol, Emmanuelle and Robert J. Gary-Bobo. 2012. "On the Optimal Number of Representatives." *Public Choice* 153 (3-4):419–445.

Bar-Isaac, Heski. 2012. "Transparency, Career Concerns, and Incentives for Acquiring Expertise." *BE Journal of Theoretical Economics* 12 (1):1–15.

Bénabou, Roland and Jean Tirole. 2003. "Intrinsic and Extrinsic Motivation." *Review of Economic Studies* 70 (3):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. Douglas. 1994. "A Theory of Conformity." *Journal of Political Economy* 102 (5):841–877.

Bolton, Patrick, Markus K Brunnermeier, and Laura Veldkamp. 2013. "Leadership, Coordination, and Corporate Culture." *Review of Economic Studies* 80 (2):512–537.

Brennan, Geoffrey and Philip Pettit. 1990. "Unveiling the Vote." *British Journal of Political Science* 20 (3):311–333.

———. 2004. *The Economy of Esteem*. Oxford University Press New York.

Cooter, Robert D. 2003. "The Donation Registry." *Fordham Law Review* 72 (5):1981–1989.

Corneo, Giacomo G. 1997. "The Theory of the Open Shop Trade Union Reconsidered." *Labour Economics* 4 (1):71–84.

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

Del Carpio, Lucia. 2014. "Are The Neighbors Cheating? Evidence from a Social Norm Experiment on Property Taxes in Peru." INSEAD.

DellaVigna, Stefano, John List, and Ulrike Malmendier. 2012. "Testing for Altruism and Social Pressure in Charitable Giving." *Quarterly Journal of Economics* 127 (1):1–56.

Ellingsen, Tore and Magnus Johannesson. 2008. "Pride and Prejudice: The Human Side of Incentive Theory." *American Economic Review* 98 (3):990–1008.

Fehrler, Sebastian and Niall Hughes. 2015. "How Transparency Kills Information Aggregation." IZA Discussion Paper No. 9027.

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

Fox, Justin and Richard Van Weelden. 2012. "Costly Transparency." *Journal of Public Economics* 96 (1):142–150.

Frankel, Alex and Navin Kartik. 2014. "Muddled Information." Columbia University.

Frey, Bruno S. 2007. "Awards As Compensation." *European Management Review* 4 (1):6–14.

Gerber, Alan S., Donald P. Green, and Christopher W. Larimer. 2008. "Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment." *American Political Science Review* 102 (1):33–48.

Harbaugh, William T. 1998. "What do Donations Buy? A Model of Philanthropy Based on Prestige and Warm Glow." *Journal of Public Economics* 67 (2):269–284.

Harbaugh, William T., Ulrich Mayr, and Daniel R. Burghart. 2007. "Neural Responses to Taxation and Voluntary Giving Reveal Motives for Charitable Donations." *Science* 316 (5831):1622–1625.

Hermalin, Benjamin E. and Michael L. Katz. 2006. "Privacy, Property Rights and Efficiency: The Economics of Privacy as Secrecy." *Quantitative Marketing and Economics* 4 (3):209–239.

Hummel, Patrick, John Morgan, and Phillip C. Stocken. 2013. "A Model of Flops." *RAND Journal of Economics* 44 (4):585–609.

Ichino, Andrea and Gerd Muehlheusser. 2008. "How Often Should you Open the Door?: Optimal Monitoring to Screen Heterogeneous Agents." *Journal of Economic Behavior & Organization* 67 (3):820 – 831.

Jacquet, Jennifer. 2015. *Is Shame Necessary? New Uses for an Old Tool.* Pantheon Books, Random House.

Kahan, Dan M. 1996. "Between Economics and Sociology: The New Path of Deterrence." *Michigan Law Review* 95 (5):2477–2497.

Kahan, Dan M. and Eric A. Posner. 1999. "Shaming White-Collar Criminals: A Proposal for Reform of the Federal Sentencing Guidelines." *Journal of Law and Economics* 42 (1):365–392.

Kreps, David M. 1990. "Corporate Culture and Economic Theory." In *Perspectives on Positive Political Economy*, edited by James E. Alt and Kenneth A. Shelpsle. Cambridge Univ Press, 90–143.

Kuran, Timur. 1997. *Private Truths, Public Lies: The Social Consequences of Preference Falsification.* Harvard University Press.

Lacetera, Nicola and Mario 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, Ian. 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." *American Economic Review* 97 (1):150–168.

Lohmann, Susanne. 1994. "Information Aggregation through Costly Political Action." *American Economic Review* 84 (3):518–30.

Lorentzen, Peter L. 2013. "Regularizing Rioting: Permitting Public Protest in an Authoritarian Regime." *Quarterly Journal of Political Science* 8 (2):127–158.

Loury, Glenn C. 1994. "Self-Censorship in Public Discourse: A Theory of "Political Correctness" and Related Phenomena." *Rationality and Society* 6 (4):428–461.

Morgan, John and Phillip C. Stocken. 2008. "Information Aggregation in Polls." *American Economic Review* 98 (3):864–896.

Morris, Stephen. 2001. "Political Correctness." *Journal of Political Economy* 109 (2):231–265.

Ottaviani, Marco and Peter Sørensen. 2001. "Information Aggregation in Debate: Who Should Speak First?" *Journal of Public Economics* 81 (3):393–421.

Posner, Eric A. 1998. "Symbols, Signals, and Social Norms in Politics and the Law." *Journal of Legal Studies* 27 (2):765–797.

———. 2000. *Law and Social Norms*. Harvard University Press.

Prat, Andrea. 2005. "The Wrong Kind of Transparency." *American Economic Review* 95 (3):862–877.

Prendergast, Canice. 1993. "A Theory of "Yes Men"." *American Economic Review* 83 (4):757–70.

Prendergast, Canice and Lars Stole. 1996. "Impetuous Youngsters and Jaded Old-Timers: Acquiring a Reputation for Learning." *Journal of Political Economy* 104 (6):1105–34.

Reeves, Richard V. 2013. "Shame is Not a Four-Letter Word." *New York Times* .

Ronson, Jon. 2015. "How One Stupid Tweet Blew Up Justine Sacco's Life." *New York Times* .

Segal, David. 2013. "Mugged by a Mug Shot Online." *New York Times* .

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

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

Vesterlund, Lise. 2003. "The informational Value of Sequential Fundraising." *Journal of Public Economics* 87 (3):627–657.

Visser, Bauke and Otto H. Swank. 2007. "On Committees of Experts." *Quarterly Journal of Economics* 122 (1):337–372.