Social Choice and Welfare (2021) 57:617–664
https://doi.org/10.1007/s00355-021-01322-1

ORIGINAL PAPER

Identity in public goods contribution

Fuhai Hong[1] · Larry Karp[2] · Tat-How Teh[3]

Received: 11 June 2020 / Accepted: 10 March 2021 / Published online: 12 April 2021

Abstract
Agents' decision whether to join a group, and their subsequent contribution to a public good, depend on the group's ideals. Agents have different preference for this public good, e.g. reductions in greenhouse gas emissions. People who become "climate insiders" obtain identity utility, but suffer disutility if they deviate from the group ideal. That ideal might create a wide but shallow group, having many members but little effect on behavior, or a narrow but deep group. Greater heterogeneity of preferences causes the contribution-maximizing ideal to create narrow but deep groups. The contribution-maximizing ideal maximizes welfare if the population is large.

We thank the editor, an associate editor, two anonymous reviewers, Roland Bénabou, Parkash Chander, In-Koo Cho and conference and seminar participants in the 2016 EAERE Conference (Zurich), the 2017 AERE Conference (Pittsburgh), the 2018 PET Conference (Hue), Arne Ryde Workshop on Identity, Image and Economic Behavior (Lund), NTU Workshop on Environmental Economics, 4th Hanyang-Kobe-NTU Symposium in Economics, Hong Kong Baptist University, Hong Kong University of Science and Technology, Kobe University, National University of Singapore, Peking University HSBC Business School, Renmin University of China, SUFE International Workshop on "Frontiers of Environmental and Resource Economics" (Shanghai), Tsinghua University, University of Groningen, Xi'an Jiao Tong University, and Xiamen University for helpful comments.

✉ Tat-How Teh
tehtathow@gmail.com

Fuhai Hong
fuhaihong@ln.edu.hk

Larry Karp
karp@berkeley.edu

[1] Department of Economics, Lingnan University, Hong Kong, China

[2] Department of Agricultural and Resource Economics, University of California, Berkeley, USA

[3] School of Management and Economics and Shenzhen Finance Institute, The Chinese University of Hong Kong, Shenzhen, China

1 Introduction

Psychologists and management scientists have shown that persuasion strategies or nudges can change behaviour, including behavior involving externalities (Thaler and Sunstein 2008; Schultz et al. 2007; Goldstein et al. 2008; Ito et al. 2018). Costa and Kahn (2013) find that an electricity conservation nudge that provides feedback to households on their own and peers' electricity usage is much more effective with liberals or environmentalists than with conservatives. Electricity conservation has a public good feature because it conserves resources and reduces greenhouse gas emissions. The authors ascribe this asymmetry in behavior to self-identification: the ideologies that people accept, influence their behavior. The nudges appear useful only for those who identify with the ideology embedded in the nudges; the goals or norms provided by one group have little impact on people who do not belong to that group.

As evidence on anthropogenic climate change accumulates, the emerging low-carbon movement has increasingly promoted voluntary lifestyle changes to reduce carbon emissions. Al Gore received the Nobel Peace Prize in 2007 for his advocacy of a green lifestyle to combat climate change. This campaign encourages many people to self-identify as environmentalists and to reduce their carbon footprint, but many people refuse to do so, leading to a significant difference in behavior between environmentalists and non-environmentalists.[1]

The economics literature has incorporated a number of behavioral considerations into the analysis of voluntary public goods contribution. However, the theoretical literature on public goods has given little attention to explaining why the same behavioral nudges have asymmetric effects on different groups of people, why people self-identify as environmentalists or non-environmentalists, and more importantly, how asymmetric effects and self-identification can make the environmental movement more successful in inducing public goods contribution.

We model individuals' self-identification with a group ideal such as the "green lifestyle". We also show how a social planner can adjust the level of the group ideal to increase aggregate contribution to a public good. The model imbeds the economics of group identity *a la* Akerlof and Kranton (2000) into a context of public goods contribution. In this model, the population has a distribution of preferences for a socially beneficial action, such as reducing carbon emissions. Agents self-select into the green or non-green group, by deciding whether to accept the ideal prescribed by the green group. An agent obtains utility from identifying with the green group, but membership can also create costs associated with deviation from the ideal. Members decide how closely to match the group ideal, and nonmembers do not change their

[1] Kotchen and Moore (2008) find that environmentalists are more likely to voluntarily restrain their consumption of goods and services that generate negative externalities. Kahn (2007) finds that those who vote for green policies and register for liberal/environmental political parties live a greener lifestyle, commuting by public transit more often, favoring hybrid vehicles, and consuming less gasoline than non-environmentalists.

behavior. The green group's ideal influences a person's decision whether to join the group, and potentially influences her subsequent behavior.

Agents are individuals or actors such as companies, cities, or states. The green ideal is a recommended level of contribution to the public good, such as a level of abatement. An influential entity can modify this ideal. Opinion-molders, including politicians, authors, and religious leaders, might influence the ideal through public policies, media, books, and church.[2] Nonenforceable international agreements such as the Kyoto Protocol or the 2015 Paris climate agreement can also alter the group ideal. Agents who identify with the ideal can contribute to the public good by reducing their energy consumption below their individually rational level. California's steps to reduce carbon emissions perhaps exceed efforts that would maximize its (narrowly construed) welfare; identity benefits might be an important motivator. The utility from identifying with the green group could be a warm glow or status-related benefit. For individuals, these benefits might be psychological, but for companies or states, there may also be brand-related commercial benefits or political prestige.

In contrast to most of the existing models on social identity, which take the ideal prescribed by the group as fixed,[3] we determine the optimal level of the ideal. The rigor of this ideal influences the self-selection and the behavior of those who identify as green. A demanding ideal might lead to significant changes in behavior, but only among the small group that self-identifies as green: this group is deep but narrow. A relaxed ideal can lead to large green membership but only modest changes in behavior: a wide but shallow group.

We find that the ideal that maximizes aggregate contributions to the public good depends on the distribution of preferences. Under uniformly distributed preferences, width trumps depth when agent heterogeneity is small: the contribution-maximizing ideal makes the agent with weakest green preferences indifferent between joining the green group and remaining outside, and thus all agents join the group. When preference heterogeneity is large, it is too expensive to attract all agents into the group. Only those with high demand for the public good join, and they significantly increase their contribution to the public good; here, depth trumps width. We then characterize the contribution-maximizing ideal under more general preference distributions. Here, the tradeoff between width and depth of the insider group is complicated by the distributional form of preferences. We provide conditions for which width trumps depth.

Our analysis shows that agents' individually rational behavior constrains the possibility of increasing contributions by means of manipulating the group ideal: beyond some level, the greater depth arising from a higher ideal does not make up

[2] For example, authors Peter Nearing and Janet Luhrs advocate sustainable living through vivid illustrations of simple lifestyles; in 2007, the United Nations published "Sustainable Consumption and Production, Promoting Climate-Friendly Household Consumption Patterns," to promote sustainable lifestyles in communities and homes; Nobel Laureate Al Gore further propelled the green lifestyle movement through his movie "*An Inconvenient Truth*".

[3] Examples include Akerlof and Kranton (2005, 2008, 2010)), Benjamin et al. (2010), Benjamin et al. (2016) and Hiller and Verdier (2014). These studies extend Akerlof and Kranton's (2000) framework to analyze behavior in workplaces, schools, churches, and families.

for the resulting loss in width. With a sufficiently large population, we show that this constraint is also binding for the welfare-maximization problem. With a larger number of agents, the non-excludability property of the public good makes individual contributions carry a larger weight (relative to contribution costs) in determining an agent's welfare; thus, the contribution-maximizing ideal also maximizes welfare. However, when the population is not sufficiently large, the contribution costs carry a larger weight; for some distributions of preferences, the welfare-maximizing ideal departs from the level that maximizes contributions, by improving the match between types and levels of contribution.

The characterization of the optimal ideal is analogous to the result in the mechanism design problem with adverse selection (Laffont and Martimort 2001). There, a principal delegates production decisions to agents who have private information about their productivity. The optimal menu of offers induces both types of agents to produce if the difference in productivity is small. In contrast, when the difference in productivity is large, it is not worthwhile attracting the less productive agent; there the optimal menu attracts the more productive agent only. The trade-offs in the two problems are similar: in the mechanism design problem, the principal trades off between the rent extracted from the more productive agent and the participation of the less productive agent; in the identity model, the trade-off is between the contribution by strong-preference individuals and the participation of the weak-preference individuals.

1.1 Relation to the literature

An extensive behavioral economics literature studies public goods. A moral imperative, arising from introspection and associated with Kantian absolute laws, can enhance public good provision (Brekke et al. 2003). In the Akerlof and Kranton (2000) framework, the ideal is a social but perhaps not moral norm; people's acceptance of the social norm affects their self-selection into social categories. Fehr and Schmidt (1999) examine the role of inequality aversion in voluntary public good contribution. Andreoni (1990) and Holländer (1990) consider warm glow and social approval as by-products of contributing to a public good. These studies do not consider self-selection into social groups, the focus of our paper. Rege (2004) endogenizes the strength of social approvals, emphasizing interactions among contributors and non-contributors, in a model without an ideal public good contribution.

Aimone et al. (2013) study endogenous group formation in a public good game where players choose reduced rates of return to private investment, and those who chose similar rates are grouped together. This mechanism with seemingly unproductive costs boosts public goods contribution of pro-social players by endogenous sorting (free-riders are not willing to choose reduced rates) and by substitution (reduced return rates of private investment make contribution more appealing).[4]

The empirical and experimental literature shows that group identity has significant effects on interpersonal interactions (e.g. Chen and Li 2009). Burlando and Hey (1997), Benjamin et al. (2016), Solow and Kirkwood (2002), and Croson et al. (2003) estimate the effects of national, religious, social and gender identities on public good contribution.

The theoretical literature of identity and economics emphasizes the following three questions: How does social identity affect behavior? How do people self-select into different social groups? How are group norms/ideals/prescriptions determined? Benjamin et al. (2010) and Benjamin et al. (2016) take group sorting and group prescriptions as given and study the effect of social identity on behavior when an individual derives disutility from deviating from the prescription of his social category. Making a social identity more salient increases the weight of this disutility and thus affects behavior. Almudi and Chóliz (2011) introduce an environmentally friendly identity which depends only on the individual's consumption level; their model has no ideal behavior or social categorization. The seminal studies by Akerlof and Kranton (cited above) mainly investigate the first two questions, while Akerlof and Kranton (2002) discuss the tradeoff for the third question as well. In a school setting, Akerlof and Kranton (2002) consider a tradeoff in choosing the school's ideal: a higher ideal raises the effort of those in the right tail of the distribution but causes other students to reject the school and exert less effort. They also discuss the ideal that maximizes the mean skill acquisition in schools. Our paper studies all three questions in the context of public goods contribution. We solve for the ideals that maximize either contribution or welfare.

Bernard et al. (2016) also study the above three questions concerning norm determination, sorting and behavior. Their model has two types of people, a high type who increases social status of the group, and a low type. A player derives utility from the status of his social group and disutility from the social distance between his own type (behavior) and the group norm. They refer to the low type's incentives to join high-status groups as social free-riding. The norm in their model equals the average type or behavior in the group. In our setting the norm is an ideal behavior promoted by an external influential entity; we find the norm that maximizes either contributions or welfare.[5] Other approaches of modelling identity include oppositional identities (Bisin et al. 2011), and identity investment where a player has incomplete information on her own type (Bénabou and Tirole 2011).

[4] Similarly, Iannaccone (1992) study sacrifice and seemingly inefficient prohibitions as a screening device where individuals sort themselves into different religions. Carvalho (2016) models an identity-based religious organization that sets religious strictness to maximize participation in its activities. In an industrial organization context, Kosfeld and von Siemens (2011) consider how workers with different willingness to cooperate self select into firms having different monetary incentives and level of worker cooperation, leading to heterogeneous corporate cultures.

[5] In a related paper by Shayo (2009), identity utility is also derived from social status while identity cost comes from social distance from other group members rather than an ideal. Shayo (2009) studies the formation of national identities and preferences for redistribution. Costa-i-Font and Cowell (2014) review the related literature on social identity and redistribution.

Akerlof (2017) studies value formation and esteem: Individuals choose values motivated by self-esteem and esteem from others. In both our paper and Akerlof (2017), values or ideals are deliberately chosen. In our model, an external influential entity chooses the ideal behaviour and individuals then decide whether to identify with that ideal. In Akerlof (2017), individuals choose their own values. In this respect, the two papers are also related to Hsiaw (2013) who considers an individual who sets a goal to deal with self-control problems. In our model, individuals incur disutility if their behavior deviates from the ideal they have identified with. In Akerlof (2017), a person obtains self-esteem from achieving more than the group average.

Section 2 describes the model and discusses its assumptions. Sections 3–5 analyze the three stages of the game, contribution to the public good, self-selection, and choice of the group ideal. Section 6 concludes. All proofs are relegated to the Appendix. Proofs of propositions on contribution maximization (Propositions 1, 3 and 4) are in Appendix A.3, while proofs of propositions on welfare maximization (Propositions 2 and 5) are in Appendix A.4.

2 Model setup

The population contains N agents, each of whom makes a voluntary contribution to a public good. In the environmental context, the contribution equals pollution abatement. Agent i contributes a_i, for $i \in \{1, 2, \dots, N\}$, incurring the private cost $\frac{1}{2}a_i^2$. Agent i's constant marginal utility of the public good is the realization of an identically and independently distributed random variable β_i, having continuous probability density function $f(\beta_i)$ and cumulative distribution function $F(\beta_i)$ defined on $\left[\underline{\beta}, \overline{\beta}\right]$, with $\infty \geq \overline{\beta} > \underline{\beta} > 0$. Many public economics models use the linear-quadratic functional assumptions because they lead to a simple equilibrium in dominant strategies (e.g. Barrett 1994; Goeschl and Perino 2017; Ali and Bénabou 2016).

The heterogeneity of β_i may be due to differences in tastes, information, or business opportunities. An individual's preference for air quality may depend on income, which affects their opportunities for adaptation (e.g. air conditioning or filters). A state's preference might depend on population density. Environmentalists and climate skeptics may have different information or beliefs about the consequences of the accumulation of Greenhouse gases (GHG), and therefore about the benefit of abatement.

Agent i's utility associated with the public good (ignoring identity-related utility) is $\beta_i\left(a_i + \sum_{j \neq i} a_j\right) - \frac{1}{2}a_i^2$. The agent takes as given other agents' contributions. Without identity-related utility (our baseline), agent i chooses a_i to maximize $\beta_i a_i - \frac{1}{2}a_i^2$, resulting in the baseline level of abatement

$$a_i^B \equiv a^B(\beta_i) = \beta_i > 0, \tag{1}$$

where the superscript B refers to baseline, while the subscript i denotes agent i; a_i^B is a dominant strategy.

2.1 Ideal, identity and utility

Akerlof and Kranton's (2005) review article on identity economics quotes an argument of Pareto (1920[1980]): much of utility depends not only on what economists normally think of as tastes, but also on norms as to how people think that they and others should behave. In their framework, a person's identity describes gains and losses from behavior that conforms to or departs from the norm of a particular social group in particular situations. In our context of public goods contribution, we suppose that a green group prescribes an ideal level of contribution, which is an injunctive norm (Benjamin et al. 2016). This group ideal level of contribution to the public good, denoted by a^*, affects the sorting of agents and the public good contribution of those who join the group. "Insiders" self-select into the green group, and identify with the ideal; "outsiders" ignore the ideal. We assume that, as is common in the identity literature (e.g. Akerlof and Kranton 2002, 2005), individuals who join the green group obtain utility from being an insider, $V > 0$. Insiders have a sense of belonging and a feeling of pride; the social status of the group also enhances the insider's self image (Akerlof and Kranton 2000, 2008). Companies, cities, or states may obtain commercial or political benefits associated with green membership.[6]

Identity utility depends on the extent to which the insider's behavior matches the ideal behavior prescribed by the social group (Akerlof and Kranton 2000). Following Akerlof and Kranton (2002; 2005; 2008) and Benjamin et al (2010; 2016)), we assume that insiders who deviate from the ideal suffer a utility loss, and that this utility loss depends on the difference between the insider's contribution and the ideal contribution level. We say an insider who contributes not more than the ideal (weakly) under-contributes, and one who contributes more than the ideal over-contributes. Their losses are

$$\text{under-contributing insider's loss (if } a_i \le a^*\text{): } \frac{\theta}{2}\left(a_i - a^*\right)^2$$
$$\text{over-contributing insider's loss (if } a_i > a^*\text{): } \frac{\gamma}{2}\left(a_i - a^*\right)^2.$$

Under-contributors may feel guilt or social pressure about contributing less than the group ideal, or be punished by peers (e.g. Fehr and Gächter 2002). Companies or states that strictly under-contribute may be vulnerable to bad publicity. Over-contributors might also incur disutility from exceeding the ideal. Monin et al. (2008) provide experimental evidence showing that people's positive self-image may be threatened by those who "do the right thing", leading to resentment against them, and a utility loss for over-contributors. Bénabou and Tirole (2011) endogenize the ostracism towards the virtuous "do-gooders". We adopt Akerlof and Kranton's

[6] In our model, V can also be viewed as the difference in identity utility between insiders and outsiders, if the outsiders also enjoy certain identity utility.

(2002) assumption that the utility loss is a quadratic function of the gap between the insider's action and the ideal level, but we relax their assumption of symmetric loss in our main analysis. With $\theta > 0$ and $\gamma > 0$, the model includes both the symmetric loss case ($\gamma = \theta$) and the case where there is (almost) no loss from over-contribution (γ converges to zero).

We then assume that, in the community or the society, there exists an influential person or entity who can influence or nudge the ideal of contribution to the public good, a^*, prior to the membership decisions. Influence-molders such as Al Gore or James Hansen, or those who support them, may be able to use the media, schools, and churches to alter the green ideal. Political entities may be able to use non-binding contracts or international agreements to adjust the green ideal.[7] We will identify the choice of a^* that maximizes provision of the public good, and show when this level also maximizes aggregate welfare.

The game's timeline is:

- At stage 0, an influential entity chooses (or adjusts) the ideal a^*.
- At stage 1, agents learn their preference type β_i and individually decide whether to identify with the ideal a^* (and become an insider) or remain an outsider.
- At stage 2, agents individually choose their public good contribution.

3 Public good contributions (stage 2)

Consider an agent with β_i. If the agent is an outsider, she contributes a_i^B, because the ideal does not affect the outsider's preference. If she is an insider, the agent solves:

$$\max_{a_i} U_i\left(a_i|a^*\right) = \begin{cases} \beta_i\left(a_i + \sum_{j\neq i} a_j\right) - \frac{1}{2}a_i^2 - \frac{\theta}{2}\left(a_i - a^*\right)^2 + V \text{ if } a_i \leq a^* \\ \beta_i\left(a_i + \sum_{j\neq i} a_j\right) - \frac{1}{2}a_i^2 - \frac{\gamma}{2}\left(a_i - a^*\right)^2 + V \text{ if } a_i > a^* \end{cases}, \quad (2)$$

where $\sum_{j\neq i} a_j$ indicates the total contributions of players other than i. Adopting the Nash assumption, player i treats $\sum_{j\neq i} a_j$ as exogenous. The following lemma shows the optimal contribution of insiders:

Lemma 1 *An insider (superscript I) with β_i contributes*

[7] The literature on identity and economics addresses the determination of group stereotypes or prescriptions ("ideals" in our model) in varied ways. Benjamin et al. (2010; 2016) treat group prescriptions as exogenous and given. In Shayo (2009) and Bernard et al. (2016), the group stereotype is determined by the average type or behavior of the group members. In Akerlof (2017), individuals choose their own values. Akerlof and Kranton (2002) discusses the choice of a group ideal in a school setting. In this approach, an ideal setter (the school) interacts with a continuum of agents (students) who decide whether to identify with the ideal. We follow this approach in the context of public goods contribution.

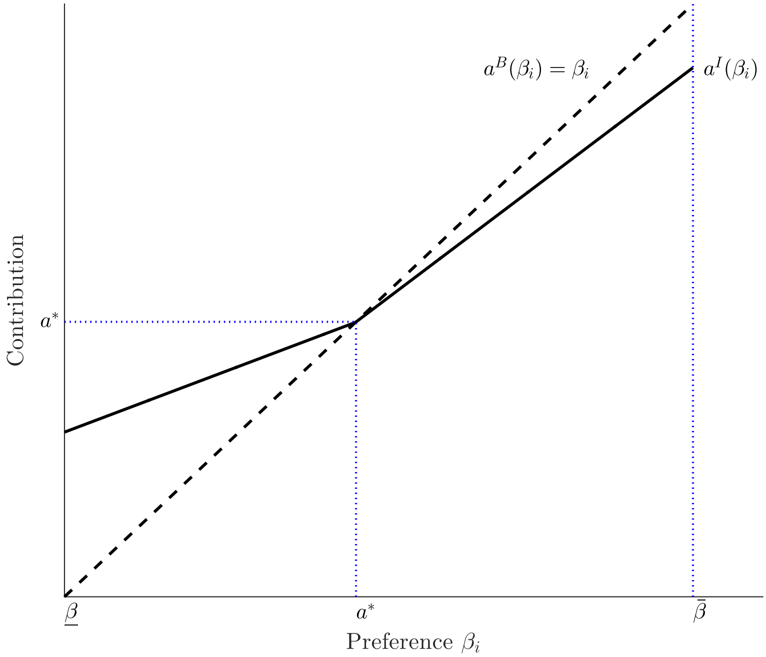

Fig. 1 The solid line shows a_i^I, and the dashed line shows $a_i^B = \beta_i$. $\left[\underline{\beta}, \bar{\beta}\right] = [0, 9/5]$, $a^* = 7/8$, $\theta = 3/2$, $\gamma = 1/4$.

$$a_i^I \equiv a^I\left(\beta_i\right) = \begin{cases} \frac{\theta\left(a^*-\beta_i\right)}{1+\theta} + \beta_i \text{ if } \beta_i \leq a^* \\ \frac{\gamma\left(a^*-\beta_i\right)}{1+\gamma} + \beta_i \text{ if } \beta_i > a^*. \end{cases} \tag{3}$$

Define $\Delta\left(\beta_i\right) = a_i^I - a_i^B$, the change of public good contribution due to membership in the group:

$$\Delta\left(\beta_i\right) = \begin{cases} \frac{\theta\left(a^*-\beta_i\right)}{1+\theta} \geq 0 \text{ if } \beta_i \leq a^* \\ \frac{\gamma\left(a^*-\beta_i\right)}{1+\gamma} < 0 \text{ if } \beta_i > a^*. \end{cases} \tag{4}$$

The expression in (4) implies:

Remark 1 An insider with β_i under-contributes if $\beta_i \leq a^*$; she is an over-contributor otherwise. Membership increases an under-contributing insider's contribution and decreases an over-contributing insider's contribution. In each case, the effect is proportional to the gap $a^* - \beta_i$.

Figure 1 illustrates the relation between a^*, a_i^B and a_i^I, when $a^* \in \left(\underline{\beta}, \overline{\beta}\right)$. A larger utility loss from deviating from the ideal (larger θ or γ) induces the insider to move toward the ideal ($\frac{da_i^I}{d\theta} \geq 0$ for $\beta_i \leq a^*$and $\frac{da_i^I}{d\gamma} < 0$ for $\beta_i > a^*$).

Remark 2 For any $\beta_j \neq \beta_i$,

- (i) we have $\left|a^I\left(\beta_j\right) - a^I\left(\beta_i\right)\right| < \left|a^B\left(\beta_j\right) - a^B\left(\beta_i\right)\right|$;
- (ii) when $\beta_i < a^* < \beta_j$ or $\beta_j < a^* < \beta_i$, we have

$$\left|a^B\left(\beta_j\right) - a^B\left(\beta_i\right)\right| > \left|a^B\left(\beta_j\right) - a^I\left(\beta_i\right)\right|;$$

when $\beta_j < \beta_i < a^*$ or $\beta_j > \beta_i > a^*$, we have

$$\left|a^B\left(\beta_j\right) - a^B\left(\beta_i\right)\right| < \left|a^B\left(\beta_j\right) - a^I\left(\beta_i\right)\right|.$$

The first part of the remark shows that membership in the green group decreases the difference between the contributions of insiders with different preferences. This can be observed from Fig. 1. The second part of the remark shows that membership in the green group might increase or decrease the difference in contributions between insiders and outsiders. Consider, without loss of generality, an agent with β_i who becomes an insider and an agent with β_j who remains an outsider. With either $\beta_i < a^* < \beta_j$ or $\beta_j < a^* < \beta_i$, membership promotes convergence in contribution between the insider and the outsider: $\left|a^B\left(\beta_j\right) - a^B\left(\beta_i\right)\right| > \left|a^B\left(\beta_j\right) - a^I\left(\beta_i\right)\right|$. The convergence happens either when the insider has lower baseline contribution than the outsider but is motivated by the ideal to contribute more, or when the insider has higher baseline contribution than the outsider but is discouraged by the ideal and thus contributes less than her baseline. In contrast, with either $\beta_j < \beta_i < a^*$ or $\beta_j > \beta_i > a^*$, membership induces divergence in contribution: $\left|a^B\left(\beta_j\right) - a^B\left(\beta_i\right)\right| < \left|a^B\left(\beta_j\right) - a^I\left(\beta_i\right)\right|$. The divergence happens either when the insider has higher baseline contribution than the outsider and is motivated by the ideal to contribute even more, or when the insider has lower baseline contribution than the outsider but is discouraged by the ideal and thus contributes even less than her baseline.

4 Self-selection (stage 1)

At stage 1, agents compare their utility as insiders and outsiders and decide whether to join the green group. Strategies are dominant, so the agent's choice does not depend on other agents' action. Using (1) and suppressing the part of the payoff due to other agents' actions, agent i's utility of remaining an outsider equals

$$\beta_i^2 - \frac{1}{2}\beta_i^2 = \frac{1}{2}\beta_i^2. \tag{5}$$

Using (3) in (2) (and again suppressing the part of the payoff due to other agents' actions), the insider's utility equals

$$\begin{matrix} \frac{1}{2}\frac{\beta_i(\beta_i+2\theta a^*)-\theta a^{*2}}{1+\theta} + V \text{ if } \beta_i \leq a^* \\ \frac{1}{2}\frac{\beta_i(\beta_i+2\gamma a^*)-\gamma a^{*2}}{1+\gamma} + V \text{ if } \beta_i > a^*. \end{matrix} \tag{6}$$

Define

$$B \equiv \left(\frac{2V(1+\theta)}{\theta}\right)^{1/2} \text{ and } D \equiv \left(\frac{2V(1+\gamma)}{\gamma}\right)^{1/2}.$$

The functions B and D increase with identity utility (V) and decrease with the insiders' cost of departing from the ideal (θ and γ). B and D thus provide measures of membership appeal to under-contributors and over-contributors. Under the tie-breaking assumption that an agent who is indifferent between the choices decides to join the group, we have:

Lemma 2 *Agents with $\beta_i \in [a^* - B, a^* + D]$ join the group, and other agents remain outsiders.*

By Remark 1, agents with $\beta_i \in [a^* - B, a^*]$ are under-contributing insiders, while agents with $\beta_i \in (a^*, a^* + D]$ are over-contributing insiders.

5 Choice of the group ideal (stage 0)

We now move to Stage 0 when a social leader chooses the group ideal. The social leader's objective depends on his or her ideology. Environmentalists, who put the environmental quality at the center of their value system, may want to maximize contributions to the public good.[8] A materialist social leader wants to maximize the (expected) material welfare. If identity (dis-)utilities are purely psychological, then the material welfare is measured by the direct benefits of the public good net of its provision costs. A utilitarian social leader also takes into account the identity (dis-) utilities. In this paper, we will not take a stance on what objective function the social leader should adopt. We will first determine the ideal that maximizes the expected contribution to the public good, and then find the conditions for which this ideal also maximizes expected welfare.

[8] For example, Aidt (1998) and Conconi (2003) assume that environmentalists organize a green lobby group that only cares about the environment. List and Sturm's (2006) model also assumes that environmentalists' payoff depends only on which environmental policy is undertaken.

For each given a^*, we know from (1) that each agent contributes β_i if she is an outsider, and adjusts her contribution by (4) if she becomes an insider. By taking integral over agent type space $\left[\underline{\beta}, \overline{\beta}\right]$, we can express the Stage-0 expected contribution of an agent as $E\left(\beta_i\right) + g(a^*)$. Here, $E(.)$ is the expectation operator while $g(a^*)$ denotes the expected change in contribution due to the ideal:

$$g(a^*) = \frac{\theta}{1+\theta} \int_{[a^*-B, a^*]\cap[\underline{\beta}, \overline{\beta}]} \left(a^* - \beta_i\right) dF(\beta_i) + \frac{\gamma}{1+\gamma} \int_{[a^*, a^*+D]\cap[\underline{\beta}, \overline{\beta}]} \left(a^* - \beta_i\right) dF(\beta_i), \tag{7}$$

where we have used Lemma 2 that determines the set of insiders. The first term in (7) is positive, representing the increase in expected contribution by under-contributing insiders. The second term in (7) is negative, representing the decrease in expected contribution by over-contributing insiders. Since $g(a^*)$ represents the expected change in contribution relative to the baseline, an ideal that maximizes $g(a^*)$ must be contribution-maximizing, and so we focus on $g(a^*)$ in what follows.

Expression (7) shows that a higher a^* increases insiders' contributions, leading to a "deeper" group. The ideal also affects the ranges of the under- and over-contributing insiders, altering the group's "width". The contribution-maximizing a^* typically involves a trade-off between depth and width.

5.1 Uniformly distributed preferences

Here we assume that β_i is uniformly distributed over $\left[\underline{\beta}, \overline{\beta}\right]$, as in Tabarrok (1998), Barbieri and Malueg (2008), and Kotchen (2009). We first consider the relation between the ideal and expected contributions, and then turn to welfare effects. We let $\bar{a}$ denote the ideal that maximizes the expected contribution.

Proposition 1 *Under uniformly distributed β_i, the contribution-maximizing ideal $\bar{a} = \max\left\{\underline{\beta} + B, \overline{\beta}\right\}$.*

The contribution-maximizing ideal exhibits a few notable properties. First, when preference heterogeneity is small (i.e. $\overline{\beta} - \underline{\beta} \leq B$), the contribution-maximizing ideal $\bar{a} = \underline{\beta} + B$, being (weakly) higher than $\overline{\beta}$, attracts all agents to become under-contributing insiders and motivate them to contribute more. Intuitively, a marginal decrease in the ideal starting from $\underline{\beta} + B$ does not change the group's "width" but reduces its "depth". A marginal increase in the ideal starting from $\underline{\beta} + B$ reduces the group's "width" as agents with the lowest type drop out as outsiders; this loss dominates the gain from the group's increased "depth", with uniformly distributed types.

Second, when preference heterogeneity is large (i.e. $\overline{\beta} - \underline{\beta} > B$), the contribution-maximizing ideal $\bar{a} = \bar{\beta}$ attracts agents with preference $\beta_i \in \left[\overline{\beta} - B, \overline{\beta}\right]$ as under-contributing insiders, while lower-type agents stay out. To see the intuition, first note that at ideal $\overline{\beta}$ there are no overcontributing insiders. A marginal decrease in the ideal causes the highest type agents to become over-contributing insiders and thus discourages their contribution; moreover, with uniformly distributed types, the marginal decrease in the ideal does not affect the "depth" or "width" of the group of under-contributing insiders. Starting from $\overline{\beta}$, a marginal increase in the ideal reflects the same trade-off as in the last paragraph: it causes low-type members to drop out, the loss of which dominates the gain in the "depth" of the remaining insiders' contribution.

In both cases, all insiders are under-contributing insiders; the contribution-maximizing ideal is high enough to encourage all insiders to weakly increase their contribution to the public good.

5.1.1 Implications

When preference heterogeneity is small, Proposition 1 implies that the contribution-maximizing ideal attracts all agents to become insiders, and weakly increases their public good contribution. The resulting group is wide but shallow. Here, the agent with the lowest preference for the public good ($\beta_i = \underline{\beta}$) is indifferent between becoming an insider and staying out. This conclusion is consistent with Knack and Keefer's (1997) and Hardin's (2005) empirical finding that in more homogeneous societies, there is typically a higher degree of acceptance to social norms.

When preference heterogeneity is large, Proposition 1 implies that the contribution-maximizing ideal equals the baseline contribution of the agent with the highest demand for the public good $\bar{a} = \overline{\beta} = a^B\left(\overline{\beta}\right)$, leading to a narrow but deep group. Agents with high demand for the public good become under-contributing insiders and contribute more than their baseline, and others with low demand remain outsiders and do not change their contribution. The ideal enlarges the gap in contribution between the high-type agents and low type agents. This case appears to fit some observations of the climate change issue. People have diverse and even contradictory views about climate change, leading to considerable public disagreement about the value of GHG abatement: there is large preference heterogeneity for this public good. From an online experiment with thousands of Germans, Diederich and Goeschl (2014) find that only 15% of the respondents are willing to give up a significant cash prize for a verified one ton reduction in CO_2 emissions; members of that small group, however, are willing to sacrifice as much as 100 Euro to reduce CO_2 emissions by one ton. This example illustrates that the environmental ideal is not accepted by everyone and that the gap in contribution between environmentalists and nonenvironmentalists is large.

We then substitute $\bar{a}$ in Proposition 1 to (7) and use $B \equiv \left(\frac{2V(1+\theta)}{\theta}\right)^{1/2}$ to identify the increase in expected contributions under the contribution-maximizing ideal:

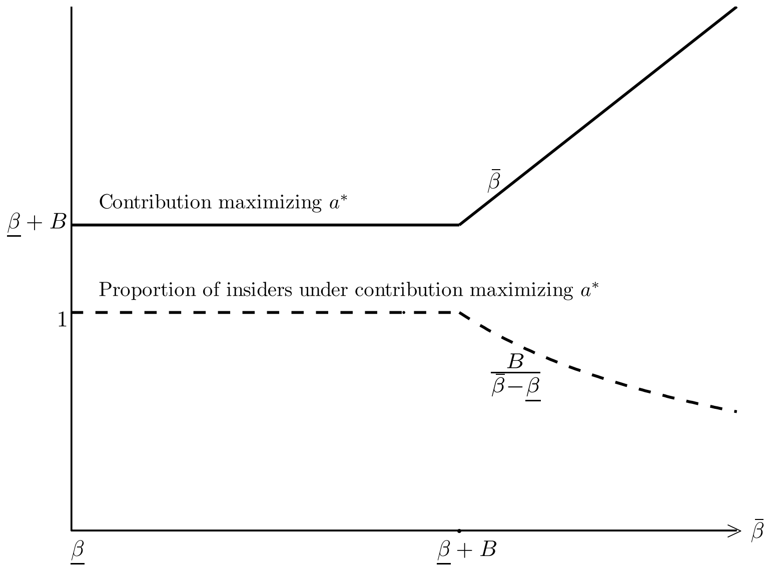

Fig. 2 The horizontal line represents $\overline{\beta}$, with the origin set at $\overline{\beta} = \underline{\beta}$. The solid curve shows the contribution-maximizing ideal while the dashed curve shows the proportion of insiders under this ideal, assuming $\underline{\beta} + B > 1$

$$g\left(\underline{\beta} + B\right) = \frac{\theta}{1+\theta}\left(\left(\frac{2V(1+\theta)}{\theta}\right)^{1/2} - \frac{\overline{\beta} - \underline{\beta}}{2}\right) \text{if } \overline{\beta} - \underline{\beta} \leq B$$

and

$$g(\overline{\beta}) = \frac{V}{\overline{\beta} - \underline{\beta}} \text{ if } \overline{\beta} - \underline{\beta} > B.$$

The following remark provides comparative statics results.

Remark 3 Under uniform distribution of β_i:

- (i) If $\overline{\beta} - \underline{\beta} < B$, $\bar{a} = \underline{\beta} + B$ is increasing in V and decreasing in θ.
- (ii) If $\overline{\beta} - \underline{\beta} \geq B$, $\bar{a} = \overline{\beta}$ attracts agents with $\beta_i \geq \overline{\beta} - B$ as insiders; the expected proportion of insiders is $\frac{B}{\overline{\beta}-\underline{\beta}} = \frac{1}{\overline{\beta}-\underline{\beta}}\left(\frac{2V(1+\theta)}{\theta}\right)^{1/2}$, which increases in V, and decreases in both θ and $\overline{\beta} - \underline{\beta}$.
- (iii) The expected increase in public good contribution, $g(\bar{a})$, increases in V, weakly increases in θ, and decreases in $\overline{\beta} - \underline{\beta}$.
- (iv) The contribution-maximizing ideal, its associated proportion of insiders, and $g(\bar{a})$ are independent of γ.

We first discuss the effect of V, the identity utility, and θ, which determines under-contributors' cost of departing from the ideal. A higher V or a lower θ makes it more attractive to become an insider. When preference heterogeneity is small, $\bar{a} = \underline{\beta} + B$ makes the lowest-type agent indifferent between becoming an insider and staying out. A higher V or a lower θ leads to a higher ideal that satisfies this indifference condition. When preference heterogeneity is large, only high-type agents join the group as insiders. A higher V or a lower θ increases the attractiveness of the ideal, thereby increasing the proportion of insiders. Next, $g(\bar{a})$ (weakly) increases with V and θ. This is because a higher V expands the "width" of the insider group for any ideal, while a higher θ induces under-contributors to raise their contribution in order to better conform to the group's ideal. Part (iv) of the remark follows naturally from the fact that there are no over-contributors under the contribution-maximizing ideal.

Next, preference heterogeneity, $\overline{\beta} - \underline{\beta}$, affects the contribution-maximizing ideal, the resulting size of the insider group, and the efficacy of identity. In Fig. 2 (for $\underline{\beta} + B > 1$), the solid curve shows the contribution-maximizing ideal and the dashed curve shows the proportion of insiders under this ideal. Fixing $\underline{\beta}$, when $\overline{\beta} - \underline{\beta} \leq B$, the contribution maximizing ideal equals $\underline{\beta} + B$, where all agents join the group. When $\overline{\beta} - \underline{\beta} > B$, the contribution-maximizing ideal increases in (and is equal to) $\overline{\beta}$, and the proportion of insiders decreases in $\overline{\beta}$. Social identity is less effective in enhancing public good contribution in a more heterogenous population (Remark 3).

Remark 2 notes that identity may promote either convergence or divergence in contributions across insiders and outsiders. We revisit this issue under the contribution-maximizing ideal, $\bar{a}$. When $\overline{\beta} - \underline{\beta} < B$, where all agents are insiders, Remark 2 implies that identity promotes convergence in contribution among any two agents. When $\overline{\beta} - \underline{\beta} \geq B$, with $a^* = \bar{a}$, the gap in public good contribution between insiders and outsiders increases; here, agents with $\beta_i \geq \overline{\beta} - B$ become insiders and increase their contribution, while outsiders do not change their contribution. This result is consistent with the asymmetric effects, across groups, of energy conservation nudges (Costa and Kahn 2013). The nudges influence insiders (political liberals/environmentalists) but not outsiders (political conservatives). The outsiders do not identify with the ideology embedded in the nudges, so the nudges widen the gap in energy use between the two groups of people.

The above analysis assumes linear benefit functions and quadratic cost functions. In Appendix B, we consider more general functional forms, e.g. replacing the linear benefit function with a concave function. With that change, agents' actions are strategic substitutes instead of being dominant. The analysis in the appendix suggests that the main insights from Proposition 1 continue to hold.

5.1.2 Welfare

Since agents are ex ante identical, for welfare analysis it suffices to consider the expected welfare of a representative agent i.[9] We can express the expected welfare of agent i as a weighted sum of a standard payoff component $M(a^*)$ and an identity payoff component $P(a^*)$:

$$W(a^*) = M(a^*) + \lambda P(a^*), \tag{8}$$

where λ indicates the weight of the identity payoff component relative to that of the standard payoff component.

The standard payoff component $M(a^*)$ captures the direct benefits and costs arising from actions:

$$M(a^*) = E_{\beta_1,\ldots,\beta_N}\left[\beta_i a_i - \frac{1}{2}a_i^2 + \beta_i \sum_{j\neq i} a_j\right], \tag{9}$$

where the first two terms in the brackets capture the benefit and cost associated with agent i's own contribution, while the last term captures the external benefit that agent i derives from the contribution of other agents $j \neq i$. In the right hand side of (9), individual contributions, denoted by a_i and a_j, are equilibrium contributions (and subject to the agents' participation decisions after observing their types and a^*). Given the i.i.d. assumption and the fixed group size N, (9) can be simplified as

$$M(a^*) = E\left[\beta_i a_i - \frac{1}{2}a_i^2\right] + (N-1)E\left(\beta_i\right)E\left(a_j\right). \tag{10}$$

The identity payoff component $P(a^*)$ captures the benefits and costs associated with the identity. It can be expressed as $P(a^*) = E_{\beta_i}\left[p_i(\beta_i, a^*)\right]$, with

$$p_i(\beta_i, a^*) = \begin{cases} V - \frac{\theta}{2}\left(a_i - a^*\right)^2 \text{ if } i \text{ is an undercontributing insider} \\ V - \frac{\gamma}{2}\left(a_i - a^*\right)^2 \text{ if } i \text{ is an overcontributing insider} \\ 0 \text{ if } i \text{ is an outsider} \end{cases},$$

where a_i denotes agent i's equilibrium contribution subject to her participation decision after observing her type and a^*.

If the identity (dis-)utilities are purely psychological, the standard payoff component $M(a^*)$ represents the material welfare. In this case, a materialistic social leader wants to maximize $W(a^*)$ with $\lambda = 0$, while a utilitatrian social leader may want to maximize $W(a^*)$ with $\lambda = 1$.

Welfare assessments for behavioral models can be controversial. Bernheim and Rangel (2007) note that welfare assessment (normative analysis) of behavioural models often diverges from preferences that generate individual decisions.

[9] This is an agent behind the veil of ignorance under stage 0 and the social leader has not learnt the agent's type yet. With a fixed group size N, maximizing a representative agent's welfare is equivalent to maximizing the total welfare.

Examples include models on time inconsistent preferences, self-control and addiction, and various models of bounded rationality. See Bernheim and Rangel (2007) for more detailed discussions. Similarly, Akerlof and Kranton (2008, p.23) simply view the utility function in their identity model as a description of motivation and do not relate maximization of utility to maximization of welfare.

For this reason, in our welfare analysis we focus on the case where $\lambda = 0$, i.e. we focus on $M(a^*)$, the standard welfare measure in the neo-classical economics. The key advantage of this approach is that it permits an apples-to-apples comparison between the cases with and without identity. To see this, notice that the identity-related utility, V, might be manipulated by opinion-makers promoting the ideal or it might include commercial or political advantages that have offsetting costs. For example, a stronger green brand for insiders amounts to a relatively weaker brand for outsiders. By ignoring identity benefits in the welfare criterion, we avoid taking a stand on the extent to which they are manipulated or impose offsetting costs.

Let $\hat{a}$ denote the ideal that maximizes $M(a^*)$, subject to the contribution constraints in (3) and the participation decisions stated in Lemma 2. Then:

Proposition 2 *Suppose*

$$NE(\beta_i) > E(\beta_i) + \left(\frac{2V\theta}{1+\theta}\right)^{1/2}. \tag{11}$$

The contribution-maximizing ideal also maximizes the standard welfare measure $M(a^*)$, *i.e.* $\hat{a} = \bar{a}$.

To understand condition (11), we first note that the left hand side of (11) is an agent's first-best level of expected contribution, while the right hand side is an agent's baseline expected contribution, $E(\beta_i)$, plus the maximum ideal-induced increase in an insider's contribution.[10] Hence, condition (11) can be understood as ruling out the possibility of an ideal inducing more contribution than the first-best level. This is sufficient but not necessary; it holds if the population (N) is large.

Proposition 2 implies that the contribution maximizing ideal maximizes the standard welfare measure as long as the population is large enough. The intuition is the following. Identity increases agents' expected contribution. Each agent benefits from other agents' higher contribution, but insiders incur a cost due to deviating from their baseline, β_i. The contribution-maximizing ideal, $\bar{a}$, is independent of N, so the insiders' expected cost due to contributing more than their baseline level is also independent of N. However, the benefit due to other agents' higher contribution is increasing in N. For sufficiently large N, and for all contributions that can be supported by a group ideal, the agent's benefit from other agents' increased contribution exceeds the cost from his own increased contribution.

[10] By Eq. (4) and Lemma 2, the largest increase in contribution implemented by an ideal a^* is $\frac{\theta}{1+\theta}[a^* - (a^* - B)] = \frac{\theta B}{1+\theta} \equiv \left(\frac{2V\theta}{1+\theta}\right)^{1/2}$.

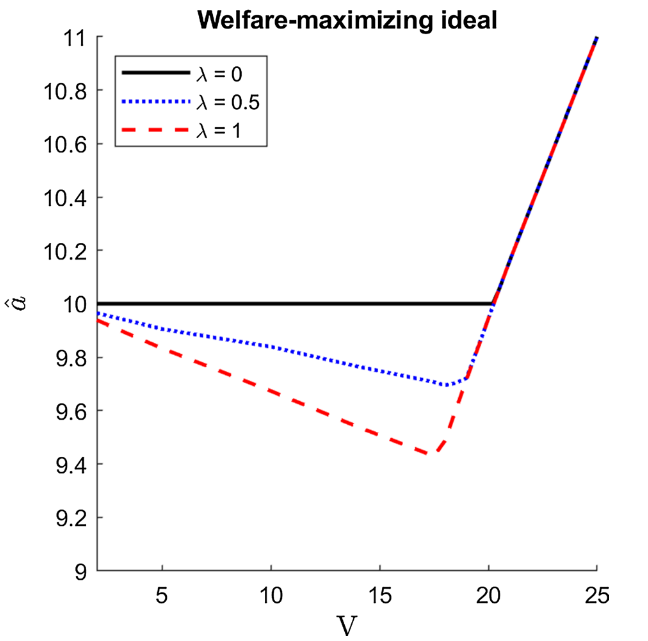

Fig. 3 .

Before ending this section, we consider what happens to the welfare assessment when the welfare function (8) has $\lambda > 0$. Figure 3 plots the welfare-maximizing ideal for three values of λ, $\lambda = 0$, $\lambda = 0.5$ and $\lambda = 1$, with V at the horizontal axis, where we assume $N = 10$, $\underline{\beta} = 1$, $\overline{\beta} = 10$, $\gamma = \theta = 1$, and $V \in [2, 25]$ so that condition (11) holds.

The black solid line plots the welfare-maximizing ideal for $\lambda = 0$, $\hat{a}$, which equals the contribution-maximizing ideal, $\max\left\{\overline{\beta}, \underline{\beta} + B\right\}$ (by Proposition 2). When V is small, $\underline{\beta} + B \equiv \underline{\beta} + \left(\frac{2V(1+\theta)}{\theta}\right)^{1/2} < \overline{\beta}$, the welfare-maximizing ideal is $\overline{\beta}$, a constant, which attracts only high-type agents; when V is large enough, the welfare-maximizing ideal is $\underline{\beta} + B$, which increases in V and attracts all agents.

The blue dotted line and the red dashed line plot the welfare-maximizing ideal for $\lambda = 0.5$ and $\lambda = 1$, respectively, both of which are (weakly) lower than the welfare-maximizing ideal with $\lambda = 0$. This is because once we consider the identity payoff component, the welfare-maximizing ideal shifts towards a lower level to induce more agents to become insiders so as to enjoy the identity utility V. As indicated by the downward-sloping part of the two lines, this shift towards shallower ideals is more pronounced when V is larger. However, when V is sufficiently large, the welfare-maximizing ideal (for $\lambda > 0$) becomes low enough to attract all agents, where $a^* = \underline{\beta} + B$. At this point, there is no need to further lower the ideal. From then on, the welfare-maximizing ideal is equal to $\underline{\beta} + B$, which is increasing in V. This corresponds to the increasing part of the blue dotted line and of the red dashed line. The welfare-maximizing ideal with $\lambda > 0$ eventually converges to the welfare-maximizing ideal with $\lambda = 0$ when the latter also takes the value $\underline{\beta} + B$. As V increases, the welfare-maximizing ideal with $\lambda > 0$ takes the value of $\underline{\beta} + B$ (to attract all agents) earlier than the welfare-maximizing ideal with $\lambda = 0$ because of the consideration of identity utility in welfare assessment.

Overall, when $\lambda > 0$, the welfare-maximizing ideal is (weakly) lower than the one indicated in Proposition 2 because of the consideration of identity utility. The difference is negligible when (i) λ is sufficiently small; or (ii) V is either sufficiently small or sufficiently large.

5.2 General distribution of preferences

This section extends the analysis of Section 5.1 to a more general class of distribution functions. Throughout this section, we assume:

Assumption 1 F is continuously differentiable and has a weakly increasing hazard rate, i.e. $\frac{f(\beta_i)}{1-F(\beta_i)}$ is weakly increasing in β_i.

Many commonly used distributions such as uniform, normal, exponential, and extreme value distributions satisfy this property.[11] We first consider contribution maximization in the case of small preference heterogeneity.

Proposition 3 (Small preference heterogeneity) *Suppose* $\bar{\beta} - \underline{\beta} \leq B$. *If* $\frac{1}{f(\underline{\beta})} \leq B$, *then* $\bar{a} = \underline{\beta} + B$; *otherwise,* $\bar{a} > \underline{\beta} + B$ *and it is the unique solution of*

$$1 - F(\bar{a} - B) = Bf(\bar{a} - B). \tag{12}$$

With uniform distribution, Proposition 1 shows that, with small preference heterogeneity, the contribution maximizing ideal $\bar{a} = \underline{\beta} + B$ attracts all agents as insiders. With general distribution as in Assumption 1, increasing the ideal beyond $\underline{\beta} + B$ is sometimes optimal. Intuitively, for all $\underline{\beta} + B \leq a^* < \bar{\beta} + B$, a marginal increase in the ideal causes the lowest-preference members, with density $f(a^* - B)$, to drop out.[12] By Remark 1 and Eq. (4), the increase in insider i's contribution is proportional to $a^* - \beta_i$. Thus, membership would have increased each of these agents' contribution by an amount proportional to B, so their defection leads to a marginal loss in contribution, represented by $Bf(a^* - B)$. At the same time, the higher ideal causes a proportional marginal increase in each of the existing insiders' contribution; their measure is $1 - F(a^* - B)$. When $\frac{1}{f(\underline{\beta})} \leq B$, the first effect dominates the second for all $\underline{\beta} + B \leq a^* < \bar{\beta} + B$, and so $\bar{a} = \underline{\beta} + B$. Otherwise, we have $\bar{a} > \underline{\beta} + B$ that equates the marginal loss and benefit, as stated in the first-order condition (12).

[11] See Bagnoli and Bergstrom (2005) for more examples and applications.

[12] By Lemma 2, any ideal $a^* > \bar{\beta} + B$ does not attract any insider.

Note that Proposition 3 nests Proposition 1 (when $\overline{\beta} - \underline{\beta} \leq B$) as a special case. If F is uniform, then $\frac{1}{f(\underline{\beta})} = \overline{\beta} - \underline{\beta} \leq B$, so $\bar{a} = \underline{\beta} + B$. More generally, the condition $\frac{1}{f(\underline{\beta})} \leq B$ depends on the skewness of the distribution density. For example, if $f' \leq 0$ for all β_i ("extremely left-skewed distribution"), then $\frac{1}{f(\underline{\beta})} \leq B$; if instead $f' > 0$ for all β_i("extremely right-skewed distribution"), then it is possible to have $\frac{1}{f(\underline{\beta})} > B$.

The analysis for the case of large preference heterogeneity is more complicated and requires stronger distributional assumptions:[13]

Proposition 4 (Large preference heterogeneity) *Suppose* $\overline{\beta} - \underline{\beta} > B$.

(i) *Suppose* $f' \geq 0$ *for all* β_i. *If* $\frac{1-F(\overline{\beta}-B)}{f(\overline{\beta}-B)} \leq B$, *then* $\bar{a} = \overline{\beta}$; *otherwise,* $\bar{a} > \overline{\beta}$ *and it is the unique solution of* (12).
(ii) *Suppose* $f' < 0$ *for all* β_i, *then* $\bar{a} = \underline{\beta} + B$ *when* γ *is small enough.*

Note that there are over-contributing insiders, whose contributions are discouraged, when $a^* < \overline{\beta}$, and there are no over-contributing insiders when $a^* = \overline{\beta}$. Therefore, $a^* = \overline{\beta}$ dominates all $a^* < \overline{\beta}$ as long as $a^* = \overline{\beta}$ leads to a larger increase in contribution by undercontributing insiders, i.e.

$$\frac{\theta}{1+\theta} \int_{a^*-B}^{a^*} \left(a^* - \beta_i\right) f\left(\beta_i\right) d\beta_i. \tag{13}$$

The derivative of (13) is proportional to $F(a^*) - F(a^* - B) - Bf(a^* - B)$, which is positive whenever F is convex (i.e. $f' \geq 0$). Therefore, (13) is higher under $a^* = \overline{\beta}$ than under all $a^* < \overline{\beta}$. Next, starting from $\overline{\beta}$, any marginal increase in the ideal involves a trade-off similar to that in Proposition 3. Here, when $\frac{1-F(\overline{\beta}-B)}{f(\overline{\beta}-B)} \leq B$ (a sufficient condition is F being uniform), we have a boundary solution $\bar{a} = \overline{\beta}$. Otherwise, $\bar{a}$ is determined by the first-order condition (12).

When $f' < 0$, F is concave, and so the derivative of (13) becomes negative for $a^* \in \left[\underline{\beta} + B, \bar{\beta} + B\right]$. This means that setting ideal $a^* < \bar{\beta}$ yields higher contribution increase by under-contributing insiders than $a^* = \bar{\beta}$. However, any such ideal involves a positive mass of over-contributing insiders, so the contribution-maximizing ideal has to balance the ideal's effects on both types of insiders. Nonetheless, if γ is sufficiently small, the ideal has a small enough effect on over-contributing insiders, in which case we have $\bar{a} = \underline{\beta} + B$, attracting the most densely populated agents with $\beta_i = \underline{\beta}$ as insiders, as stated in the last part of Proposition 4.

[13] In Appendix A.3.2, we consider non-monotonic f, and characterize the contribution-maximizing ideal when γ converges to 0.

5.2.1 Comparison with uniformly distributed preferences

If we focus on the class of distributions with monotonic density functions ($f' \leq 0$ or $f' \geq 0$ for all β_i), then Propositions 3–4 imply the following results:

- With small preference heterogeneity ($\overline{\beta} - \underline{\beta} \leq B$), $\bar{a}$ attracts all agents as insiders if $f' \leq 0$; otherwise $\bar{a}$ may attract only agents with sufficiently high β_i.[14]
- With large preference heterogeneity ($\overline{\beta} - \underline{\beta} > B$), $\bar{a}$ attracts all agents as insiders if $f' < 0$ and γ is small enough; $\bar{a}$ attracts only agents with sufficiently high β_i if $f' \geq 0$.

Compared to uniformly distributed preferences, there are two additional factors that determine whether "width" trumps "depth": (i) the shape of preference distribution; and (ii) γ, which determines the extent to which over-contributing insiders reduce their contribution after joining the group. In particular, even if preference heterogeneity is small, $\bar{a}$ may not attract all agents if $f' \geq 0$, where the mass of high-preference agents is larger than the mass of low-preference agents. Likewise, even if preference heterogeneity is large, $\bar{a}$ may still attract all agents if γ is small and $f' < 0$.

5.2.2 Welfare

Proposition 5 below considers the relation between the contribution-maximizing and the welfare-maximizing ideal, extending Proposition 2.[15]

Proposition 5 (Welfare maximization) *Suppose* (11) *holds.*

- *As N converges to ∞, the welfare maximizing ideal converges to the contribution maximizing one, i.e. $\hat{a}$ converges to $\bar{a}$.*
- *Suppose $f' \leq 0$ for all β_i. Then $\hat{a} = \bar{a} = \underline{\beta} + B$ if* (i) *$\overline{\beta} - \underline{\beta} \leq B$, or if* (ii) *$\gamma$ converges to 0 and $f' < 0$ for all β_i.*
- *Suppose $f' \geq 0$ for all β_i. Then $\hat{a} \geq \bar{a}$.*

The contribution-maximizing ideal increases other agents' contributions, but also suboptimally (from the viewpoint of $M(a^*)$) causes the agent's own contribution to exceed the baseline level. The increase in $M(a^*)$ arising from the first effect is increasing in N, and the second effect does not depend on N. When N is sufficiently large, the first effect dominates, so the contribution-maximizing ideal also maximizes welfare (the first part of Proposition 5).

Agents' individual rationality constrains the ability of the ideal to raise contributions. When the density is uniform or decreasing (and under certain conditions), this constraint binds in the welfare maximization problem: agents would have higher

[14] Suppose $f' \leq 0$ for all β_i, then $f\left(\underline{\beta}\right) \geq \frac{1}{\bar{\beta}-\underline{\beta}}$, implying $\frac{1}{f\left(\underline{\beta}\right)} \leq \bar{\beta} - \underline{\beta} \leq B$ under small preference heterogeneity. By Proposition 3, $\bar{a} = \underline{\beta} + B$.

[15] In this subsection, we focus on the standard welfare measure $M(a^*)$.

welfare if it were possible to increase the contribution, but that increase is not feasible (Proposition 2 and the second part of Proposition 5). In contrast, for increasing densities, the welfare-maximizing ideal exceeds the contribution-maximizing level (the last part of Proposition 5).

6 Conclusion

Free-riding typically leads to under-provision of a public good. The global and persistent nature of GHG pollution exacerbates the under-provision of climate services. National sovereignty and differing views about the severity of climate change make it difficult to reach an effective international agreement on GHG regulation. Voluntary contribution to the public good of emission abatement might nevertheless be important in curbing climate change. We adopt a behavioral perspective, showing how identity-related benefits can influence voluntary public good contributions. The agents in our setting might be individuals, in which case the identity-related benefits are primarily psychological. Agents might also be companies or cities or states, in which case the identity-related benefits may include both status and commercial or political benefits. A member whose public good contribution differs from the group ideal has a loss in identity benefit.

We examine the effect of the ideal, without attempting to explain the mechanism that determines it. Opinion-molders, including politicians, educators, and religious leaders, might influence the ideal through public policies, media, school, and church. Nonenforceable international agreements such as the Kyoto Protocol or the 2015 Paris climate agreement can also alter the group ideal. Manipulation of the group ideal can help alleviate the free-riding problem associated with voluntary public good contributions. A change in the ideal can alter both the depth and the width of the group associated with it. A higher ideal encourages insiders to contribute more, but reduces the range of the agents who become insiders. For uniformly distributed preferences, homogeneity tends to make the contribution-maximizing group wide but shallow, and heterogeneity tends to make that group narrow but deep. With more general classes of distributions, two additional factors affect the trade-off between the depth and width of the insider group: the shape of preference distribution and the disutility intensity for over-contributing group members. Moreover, we show that the contribution-maximizing ideal also maximizes a standard welfare measure if the number of agents is large.

It would be interesting to consider two extensions of our model. First, we assume that a single ideal divides individuals into insider and outsider groups. More generally, there might be multiple groups, each promoting an ideal. Second, the ideal in our model is one-dimensional, but in reality public good issues are multi-dimensional. Climate change, adherence to the rules of international trade, and transfers from rich to poor countries might be related. In that case, ideals could also be multidimensional. It would be interesting to explore whether the insights from our model also hold in these richer settings.

Appendix A: Proofs

A.1: Lemmas

Proof (**Lemma 1**). The insider's optimization problem is equivalent to

$$\max_{a_i} U_i\left(a_i|a^*\right) = \begin{cases} \beta_i a_i - \frac{1}{2}a_i^2 - \frac{\theta}{2}\left(a_i - a^*\right)^2 + V \text{ if } a_i \leq a^* \\ \beta_i a_i - \frac{1}{2}a_i^2 - \frac{\gamma}{2}\left(a_i - a^*\right)^2 + V \text{ if } a_i > a^* \end{cases}. \tag{14}$$

Because both alternative forms of $U_i\left(a_i|a^*\right)$ in the right side of (14) are concave, by first order conditions, we have

$$\arg\max_{a_i} \beta_i a_i - \frac{1}{2}a_i^2 - \frac{\theta}{2}\left(a_i - a^*\right)^2 + V = \beta_i + \frac{\theta\left(a^* - \beta_i\right)}{1+\theta}, \tag{15}$$

and

$$\arg\max_{a_i} \beta_i a_i - \frac{1}{2}a_i^2 - \frac{\gamma}{2}\left(a_i - a^*\right)^2 + V = \beta_i + \frac{\gamma\left(a^* - \beta_i\right)}{1+\gamma}. \tag{16}$$

If $\theta = \gamma$, then trivially the solution is given by either of these two expressions, which are the same. Now suppose $\theta \neq \gamma$. Observe that

$$\beta_i + \frac{\eta\left(a^* - \beta_i\right)}{1+\eta} \leq a^* \iff \beta_i \leq a^* \tag{17}$$

for both $\eta \in \{\theta, \gamma\}$. Suppose $\beta_i \leq a^*$. Then the right side of (15) is no higher than a^* and thus optimal given $a_i \leq a^*$. Note that when $a_i = a^*$, the two functional forms in the right side of (14) are the same, so the utility function is continuous at $a_i = a^*$. We then claim that $\frac{dU_i\left(a_i|a^*\right)}{da_i} \leq 0$ for all $a_i \geq a^*$: Concavity of $U_i\left(a_i|a_i \geq a^*\right)$ implies that $\frac{dU_i\left(a_i|a^*\right)}{da_i} \leq 0$ for all $a_i \geq \beta_i + \frac{\gamma\left(\beta_i - a^*\right)}{1+\gamma}$ by (16), and meanwhile $a_i \geq a^*$ and $\beta_i \leq a^*$ imply $a_i \geq a^* \geq \beta_i + \frac{\gamma\left(\beta_i - a^*\right)}{1+\gamma}$ by (17) . Therefore, $\frac{dU_i\left(a_i|a^*\right)}{da_i} \leq 0$ for all $a_i \geq a^*$. Consequently the right side of (15) is optimal if $\beta_i \leq a^*$.

Suppose $\beta_i > a^*$, then the right side of (16) is no lower than a^* and thus optimal for $a_i \in [a^*, \infty)$. A logic similar to the above shows that $\frac{dU_i\left(a_i|a^*\right)}{da_i} \geq 0$ for all $a_i \leq a^*$ and consequently the right side of (16) is optimal if $\beta_i > a^*$. □

Proof (**Lemma 2**). Comparing (5) and (6) while using the definitions of B and D, we have the following.

$$\begin{aligned} &\frac{1}{2}\frac{\beta_i\left(\beta_i+2\theta a^*\right)-\theta a^{*2}}{1+\theta}+V\geq\frac{\beta_i^2}{2}\\ &\quad\Leftrightarrow 2V(1+\theta)\geq\theta\beta_i^2-2\theta a^*\beta_i+\theta a^{*2}\\ &\quad\Leftrightarrow\frac{2V(1+\theta)}{\theta}\geq\left(\beta_i-a^*\right)^2\\ &\quad\Leftrightarrow a^*-B\leq\beta_i\leq a^*+B. \end{aligned} \tag{18}$$

Similarly,

$$\begin{aligned} &\frac{1}{2}\frac{\beta_i\left(\beta_i+2\gamma a^*\right)-\gamma a^{*2}}{1+\gamma}+V\geq\frac{1}{2}\beta_i^2\\ &\quad\Leftrightarrow a^*-D\leq\beta_i\leq a^*+D. \end{aligned}$$

If $\beta_i \leq a^*$, the agent will be an under-contributor as an insider, and if Inequality (18) is satisfied, the utility of being an (under-contributing) insider will be not lower than that of being an outsider. Combining $\beta_i \leq a^*$ and Inequality (18), an agent will identify with the ideal and become an under-contributing insider if and only if $a^* - B \leq \beta_i \leq a^*$. Similar logic with $\beta_i \geq a^*$ shows that an agent will identify with the ideal and become an over-contributing insider if and only if $a^* < \beta_i \leq a^* + D$. □

A.2: Remarks

Proof (**Remark** 1). This follows directly from Lemma 1, Eqs. (1) and (4): these imply that when $\beta_i \leq a^*$, $\beta_i \leq a_i^I \leq a^*$ and when $\beta_i > a^*$, $\beta_i > a_i^I > a^*$. □

Proof (**Remark** 2). Part (i): Using Equations (1) and (3), we have the following. If $a^* \geq \beta_j > \beta_i$, meaning that both agents with β_i and β_j will be under-contributing insiders by the proof of Remark 1, then

$$a^B\left(\beta_j\right)-a^B\left(\beta_i\right)=\beta_j-\beta_i>\frac{\beta_j-\beta_i}{1+\theta}=a^I\left(\beta_j\right)-a^I\left(\beta_i\right).$$

If $a^* < \beta_i < \beta_j$, meaning that both agents with β_i and β_j will be over-contributing insiders by the proof of Remark 1, then

$$a^B\left(\beta_j\right)-a^B\left(\beta_i\right)=\beta_j-\beta_i>\frac{\beta_j-\beta_i}{1+\gamma}=a^I\left(\beta_j\right)-a^I\left(\beta_i\right).$$

If $\beta_j > a^* \geq \beta_i$, meaning that the agent with β_i will be an under-contributor while the other an over-contributor as insiders, then

$$\begin{aligned} & a^B(\beta_j) - a^B(\beta_i) \\ & = \beta_j - \beta_i \\ & > \left[\beta_j + \frac{\gamma(a^* - \beta_j)}{1+\gamma}\right] - \left[\beta_i + \frac{\theta(a^* - \beta_i)}{1+\theta}\right] \\ & = a^I(\beta_j) - a^I(\beta_i), \end{aligned}$$

because $\frac{\gamma(a^*-\beta_j)}{1+\gamma} - \frac{\theta(a^*-\beta_i)}{1+\theta} < 0$.

Part (ii): If $\beta_i < a^* < \beta_j$, then $\left|a^B(\beta_j) - a^I(\beta_i)\right| = a^B(\beta_j) - a^I(\beta_i) < a^B(\beta_j) - a^B(\beta_i) = \left|a^B(\beta_j) - a^B(\beta_i)\right|$ because $a^I(\beta_i) > a^B(\beta_i)$. If $\beta_j < a^* < \beta_i$, then $\left|a^B(\beta_j) - a^I(\beta_i)\right| = -a^B(\beta_j) + a^I(\beta_i) < -a^B(\beta_j) + a^B(\beta_i) = \left|a^B(\beta_j) - a^B(\beta_i)\right|$ because $a^I(\beta_i) < a^B(\beta_i)$. If $\beta_j < \beta_i < a^*$, $\left|a^B(\beta_j) - a^I(\beta_i)\right| = -a^B(\beta_j) + a^I(\beta_i) > -a^B(\beta_j) + a^B(\beta_i) = \left|a^B(\beta_j) - a^B(\beta_i)\right|$ because $a^I(\beta_i) > a^B(\beta_i)$. If $\beta_j > \beta_i > a^*$, $\left|a^B(\beta_j) - a^I(\beta_i)\right| = a^B(\beta_j) - a^I(\beta_i) > a^B(\beta_j) - a^B(\beta_i) = \left|a^B(\beta_j) - a^B(\beta_i)\right|$ because $a^I(\beta_i) < a^B(\beta_i)$. □

Proof (**Remark** 3). All results follow from direct inspections. The only exception is the result on $g(\bar{a})$ being (weakly) increasing in θ. Clearly, if $\overline{\beta} - \underline{\beta} > B$ then $g(\bar{a}) = \frac{V}{\overline{\beta}-\underline{\beta}}$ is independent of θ. If $\overline{\beta} - \underline{\beta} \leq B$, then we can use $B = \left(\frac{2V(1+\theta)}{\theta}\right)^{1/2}$ to calculate the derivative of $g(\bar{a}) = \frac{\theta}{1+\theta}\left(B - \frac{\overline{\beta}-\underline{\beta}}{2}\right)$ with respect to θ as

$$\frac{dg(\bar{a})}{d\theta} = \left(\frac{1}{2}(2V)^{1/2}\left(\frac{\theta}{1+\theta}\right)^{-1/2} - \frac{\overline{\beta} - \underline{\beta}}{2}\right)\frac{1}{(1+\theta)^2} = \frac{B - (\overline{\beta} - \underline{\beta})}{2(1+\theta)^2} \geq 0.$$

□

A.3: Contribution maximization (Propositions 1, 3, 4)

This subsection proves Propositions 1, 3, and 4.

Before proving the propositions, we first establish some claims. We rewrite $g(a^*)$ in (7) as

$$\begin{aligned} g(a^*) = & \frac{\theta}{1+\theta}\int_{\max\{a^*-B,\underline{\beta}\}}^{\max\left\{\underline{\beta},\min\{a^*,\overline{\beta}\}\right\}} (a^* - \beta_i)f(\beta_i)d\beta_i \\ & + \frac{\gamma}{1+\gamma}\int_{\max\left\{\underline{\beta},\min\{a^*,\overline{\beta}\}\right\}}^{\min\{a^*+D,\overline{\beta}\}} (a^* - \beta_i)f(\beta_i)d\beta_i. \end{aligned}$$

Note that $g(a^*)$, being composed of continuous functions (integrals and min/max functions), is continuous. We first narrow down the possible range for $\bar{a}$ in the following series of claims.

Claim 1 $\bar{a} \in \left(\underline{\beta}, \overline{\beta} + B\right)$.

Proof For all $a^* \le \underline{\beta}$, $g(a^*) \le 0$ because all insiders (if any) are over-contributors. For all $a^* \ge \overline{\beta} + B$, $g(a^*) = 0$ because no agents join as insiders by Lemma 2. All these ideals are dominated by $a^* = \overline{\beta}$, where

$$g\left(\overline{\beta}\right) = \frac{\theta}{1+\theta} \int_{\max\{\overline{\beta}-B,\underline{\beta}\}}^{\overline{\beta}} \left(\overline{\beta} - \beta_i\right) f\left(\beta_i\right) d\beta_i > 0.$$

Hence, any $a^* \notin \left(\underline{\beta}, \overline{\beta} + B\right)$ cannot be contribution-maximizing. □

Claim 2 *Suppose* $\overline{\beta} - \underline{\beta} \le B$, *then* $\bar{a} \ge \underline{\beta} + B$.

Proof By Claim 1, any $a^* < \underline{\beta}$ can never be optimal. For all $\underline{\beta} \le a^* < \underline{\beta} + B$,

$$\begin{aligned} g(a^*)|_{\underline{\beta} \le a^* < \underline{\beta}+B} & \\ &= \frac{\theta}{1+\theta} \int_{\underline{\beta}}^{\min\{a^*,\overline{\beta}\}} \left(a^* - \beta_i\right) f\left(\beta_i\right) d\beta_i \\ &\quad + \frac{\gamma}{1+\gamma} \int_{\min\{a^*,\overline{\beta}\}}^{\min\{a^*+D,\overline{\beta}\}} \left(a^* - \beta_i\right) f\left(\beta_i\right) d\beta_i \end{aligned} \tag{19}$$

$$\le \frac{\theta}{1+\theta} \int_{\underline{\beta}}^{\min\{a^*,\overline{\beta}\}} \left(a^* - \beta_i\right) f\left(\beta_i\right) d\beta_i. \tag{20}$$

$$\begin{aligned} &< \frac{\theta}{1+\theta} \int_{\underline{\beta}}^{\overline{\beta}} \left(\underline{\beta} + B - \beta_i\right) f\left(\beta_i\right) d\beta_i \\ &= g\left(\underline{\beta} + B\right). \end{aligned}$$

The first inequality follows from the fact that the second integral in (19) is nonpositive. The second inequality uses the fact that (20) is increasing in a^*. The last equality follows from the definition of g, using $\underline{\beta} + B \ge \overline{\beta}$. Since $g(a^*)|_{\underline{\beta} \le a^* < \underline{\beta}+B} < g\left(\underline{\beta} + B\right)$ and by Claim 1, any $a^* < \underline{\beta} + B$ cannot be optimal, and so $\bar{a} \ge \underline{\beta} + B$. □

Claim 3 *Suppose* $\overline{\beta} - \underline{\beta} > B$. *If* $f' \ge 0$, *then* $\bar{a} \ge \overline{\beta}$. *If* $f' \le 0$ *or* γ *converges to* 0 *then* $\bar{a} \ge \underline{\beta} + B$.

Proof By Claim 1, any $a^* < \underline{\beta}$ can never be optimal. Suppose $f' \geq 0$ and consider $a^* \in \left[\underline{\beta}, \overline{\beta}\right)$,

$$\begin{aligned} g(a^*)|_{a^* \in \left[\underline{\beta}, \overline{\beta}\right)} = & \frac{\theta}{1+\theta} \int_{\max\{a^*-B, \underline{\beta}\}}^{a^*} \left(a^* - \beta_i\right) f\left(\beta_i\right) d\beta_i \\ & + \frac{\gamma}{1+\gamma} \int_{a^*}^{\min\{a^*+D, \overline{\beta}\}} \left(a^* - \beta_i\right) f\left(\beta_i\right) d\beta_i. \end{aligned} \tag{21}$$

$$\begin{aligned} & < \frac{\theta}{1+\theta} \int_{\max\{a^*-B, \underline{\beta}\}}^{a^*} \left(a^* - \beta_i\right) f\left(\beta_i\right) d\beta_i \\ & \leq \frac{\theta}{1+\theta} \int_{\overline{\beta}-B}^{\overline{\beta}} \left(\overline{\beta} - \beta_i\right) f\left(\beta_i\right) d\beta_i \\ & = g(\overline{\beta}). \end{aligned} \tag{22}$$

The first inequality follows from the fact that the second integral in (21) is negative (since $\beta_i \geq a^*$ in the term and $\min\{a^* + D, \overline{\beta}\} > a^*$). For the second inequality, we note that if $a^* \leq \underline{\beta} + B$, then (22) has derivative $\frac{\theta}{1+\theta} F(a^*) \geq 0$; while if $a^* \in \left[\underline{\beta} + B, \overline{\beta}\right)$ then (22) has derivative

$$\frac{\theta}{1+\theta}(F(a^*) - F(a^* - B) - Bf(a^* - B)) \geq 0$$

given that $f' \geq 0$ (hence F is convex). The final equality follows from the definition of g, using $\overline{\beta} > \underline{\beta} + B$. Since $g(a^*)|_{a^* \in \left[\underline{\beta}, \overline{\beta}\right)} < g(\overline{\beta})$ and by Claim 1, any $a^* < \overline{\beta}$ cannot be optimal, and so $\bar{a} \geq \overline{\beta}$.

Now consider $a^* \in \left[\underline{\beta}, \underline{\beta} + B\right)$,

$$\begin{aligned} g(a^*)|_{a^* < \underline{\beta}+B} = & \frac{\theta}{1+\theta} \int_{\underline{\beta}}^{a^*} \left(a^* - \beta_i\right) f\left(\beta_i\right) d\beta_i \\ & + \frac{\gamma}{1+\gamma} \int_{a^*}^{\min\{a^*+D, \overline{\beta}\}} \left(a^* - \beta_i\right) f\left(\beta_i\right) d\beta_i. \end{aligned} \tag{23}$$

The first term in the right side of (23) is increasing in a^*. As for the second term, if $a^* + D > \overline{\beta}$ the derivative is $\frac{\gamma}{1+\gamma}(1 - F(a^*)) \geq 0$; if $a^* + D \leq \beta$, the derivative is

$$\frac{\gamma}{1+\gamma}(F(a^* + D) - F(a^*) - Df(a^* + D)) \geq 0,$$

provided that $f' \leq 0$ (i.e. F is concave). Therefore, (23) is increasing in $a^* \in \left[\underline{\beta}, \underline{\beta} + B\right)$ as long as either the second term in the right side of (23) is increasing (a sufficient condition of which is $f' \leq 0$) or converges to 0 (a sufficient condition of which is γ converging 0). Thus, if $f' \leq 0$ or γ converges 0, then

$g(a^*)|_{a^*\in[\underline{\beta},\underline{\beta}+B)} < g\left(\underline{\beta}+B\right)$, so any $a^* \in \left[\underline{\beta},\underline{\beta}+B\right)$ cannot be optimal, which, together with Claim 1, implies $\bar{a} \geq \underline{\beta}+B$. □

A.3.1: Completing the proofs

Proposition 1 is a special case of Propositions 3 and 4. Therefore, we first prove Propositions 3 and 4 and then get back to Proposition 1.

Proof (**Proposition** 3) By Claims 1 and 2, when $\overline{\beta}-\underline{\beta}\leq B$, we can focus on $a^* \in \left[\underline{\beta}+B,\overline{\beta}+B\right]$, in which

$$g(a^*) = \frac{\theta}{1+\theta}\int_{a^*-B}^{\overline{\beta}}\left(a^*-\beta_i\right)f\left(\beta_i\right)d\beta_i, \tag{24}$$

the derivative of which is

$$\frac{dg(a^*)}{da^*} = \frac{\theta}{1+\theta}\left(\frac{1-F(a^*-B)}{f(a^*-B)}-B\right)f(a^*-B).$$

The increasing hazard rate assumption (Assumption 1) implies $\frac{dg(a^*)}{da^*}$ is single-crossing from above, i.e. if $\frac{dg(a^*)}{da^*}|_{a^*=a_1}\leq 0$ then $\frac{dg(a^*)}{da^*}|_{a^*=a_2}<0$ for all $a_2>a_1$. Therefore, $g(a^*)$ is quasiconcave, meaning that the solution to the first-order condition indeed maximizes g whenever it exists. If $\frac{1-F(\underline{\beta})}{f(\underline{\beta})}=\frac{1}{f(\underline{\beta})}\leq B$, then the single-crossing property implies $\frac{dg(a^*)}{da^*}<0$ for all $a^*>\underline{\beta}+B$, so $\bar{a}=\underline{\beta}+B$. If $\frac{1}{f(\underline{\beta})}>B$, by the intermediate value theorem there exists a solution $\bar{a}\in(\underline{\beta}+B,\overline{\beta}+B]$ that solves the first-order condition $\frac{1-F(\bar{a}-B)}{f(\bar{a}-B)}-B=0$, or equivalently, condition (12). The uniqueness of the solution to the first-order condition follows from the single-crossing property. □

Proof (**Proposition** 4) (i) Suppose $f'\geq 0$. By Claims 1 and 3, when $\overline{\beta}-\underline{\beta}>B$ and $f'\geq 0$, we can focus on $a^*\in\left[\overline{\beta},\overline{\beta}+B\right]$, in which the expression of $g(a^*)$takes the same form as (24). Using the same argument in the proof of Proposition 3, we have the following: If $\frac{1-F(\overline{\beta}-B)}{f(\overline{\beta}-B)}\leq B$, then $\frac{dg(a^*)}{da^*}<0$ for all $a^*>\overline{\beta}$, so $\bar{a}=\overline{\beta}$. If $\frac{1-F(\overline{\beta}-B)}{f(\overline{\beta}-B)}>B$, we have $\bar{a}\in(\overline{\beta},\overline{\beta}+B]$, which is the solution of the first-order condition $\frac{1-F(\bar{a}-B)}{f(\bar{a}-B)}-B=0$.

(ii) Suppose $f'<0$. By Claims 1 and 3, when $\overline{\beta}-\underline{\beta}>B$ and $f'<0$, we can focus on $a^*\in\left[\underline{\beta}+B,\overline{\beta}+B\right]$, in which

$$g(a^*) = \frac{\theta}{1+\theta}\int_{a^*-B}^{\min\{a^*,\overline{\beta}\}} \left(a^* - \beta_i\right) f\left(\beta_i\right) d\beta_i + \frac{\gamma}{1+\gamma}\int_{\min\{a^*,\overline{\beta}\}}^{\min\{a^*+D,\overline{\beta}\}} \left(a^* - \beta_i\right) f\left(\beta_i\right) d\beta_i. \tag{25}$$

Let $\chi_1(a^*)$ denote the derivative of the first term in the right side of (25). We have

$$\chi_1(a^*) = \begin{cases} \frac{\theta}{1+\theta}(F(a^*) - F(a^*-B) - Bf(a^*-B)) \text{ if } \underline{\beta} + B \le a^* \le \overline{\beta} \\ \frac{\theta}{1+\theta}(1 - F(a^*-B) - Bf(a^*-B)) \text{ if } \overline{\beta} < a^* \le \overline{\beta} + B \end{cases}, \tag{26}$$
$$< 0,$$

where the inequality is due to $f' < 0$ (i.e. F is strictly concave). Let $\frac{1+\gamma}{\gamma}\chi_2(a^*)$ denote the derivative of the second term in the right side of (25). We have

$$\chi_2(a^*) = \begin{cases} 0 \text{ if } a^* + D > \overline{\beta} \text{ and } a^* > \bar{\beta} \\ 1 - F(a^*) \text{ if } a^* + D > \overline{\beta} \text{ and } a^* \le \bar{\beta} \\ F(a^*+D) - F(a^*) - Df(a^*+D) \text{ if } a^* + D \le \overline{\beta} \end{cases}$$
$$\in [0, 1],$$

where $F(a^*+D) - F(a^*) - Df(a^*+D) \ge 0$ due to $f' < 0$ (i.e. F is strictly concave). Further define threshold $\bar{\gamma}$ such that

$$\frac{\bar{\gamma}}{1+\bar{\gamma}} \equiv - \max_{a^* \in \left[\underline{\beta}+B, \overline{\beta}+B\right]} \{\chi_1(a^*)\}, \tag{27}$$

in which the maximizer exists because χ_1 is continuous in the compact interval. Note that $\bar{\gamma} > 0$ since $\chi_1(a^*) < 0$ for all $a^* \in \left[\underline{\beta} + B, \overline{\beta} + B\right]$.

To prove $\bar{a} = \underline{\beta} + B$, by Claim 1, it is sufficient to show that $\frac{dg(a^*)}{da^*} < 0$ for all $\bar{a} \in \left[\underline{\beta} + B, \overline{\beta} + B\right]$ whenever $\gamma < \bar{\gamma}$:

$$\begin{aligned} \frac{dg(a^*)}{da^*} &= \chi_1(a^*) + \frac{1+\gamma}{\gamma}\chi_2(a^*) \\ &\le \max_{a^* \in \left[\underline{\beta}+B, \overline{\beta}+B\right]} \left\{\chi_1(a^*)\right\} + \frac{\gamma}{1+\gamma} \\ &< \max_{a^* \in \left[\underline{\beta}+B, \overline{\beta}+B\right]} \left\{\chi_1(a^*)\right\} + \frac{\bar{\gamma}}{1+\bar{\gamma}} \\ &= 0. \end{aligned}$$

The first inequality uses $\chi_2(a^*) \le 1$, while the second inequality uses $0 \le \gamma < \bar{\gamma}$ and the fact that $\frac{\gamma}{1+\gamma}$ is increasing in γ for any nonnegative γ; the final equality uses the definition of $\bar{\gamma}$. □

Proof (**Proposition** 1) When F is uniform, for any $\beta \in \left[\underline{\beta}, \overline{\beta}\right]$, $\frac{1-F(\beta)}{f(\beta)} = \overline{\beta} - \beta$. Therefore, if $\overline{\beta} - \underline{\beta} \leq B$, then $\frac{1}{f(\underline{\beta})} = \overline{\beta} - \underline{\beta} \leq B$; hence Proposition 3 implies $\bar{a} = \underline{\beta} + B$. If $\overline{\beta} - \underline{\beta} > B$, then $\frac{1-F(\overline{\beta}-B)}{f(\overline{\beta}-B)} = B$; hence Proposition 4(i) implies $\bar{a} = \overline{\beta}$. □

In what follows, we explain the intuition of the proof of Proposition 1. Claim 1 implies that we should focus on $a^* \in \left(\underline{\beta}, \overline{\beta} + B\right)$, where there is a positive mass of under-contributing insiders.

Suppose $\overline{\beta} - \underline{\beta} \leq B$. Claim 2 further restricts the possible range of the contribution-maximizing ideal to $a^* \in [\underline{\beta} + B, \overline{\beta} + B)$. This is because there are no overcontributing insiders for all $a^* \leq \overline{\beta} + B$ (given $\overline{\beta} - \underline{\beta} \leq B$), and $a^* = \underline{\beta} + B$ leads to higher contribution of undercontributing insiders than any $a^* < \underline{\beta} + B$. Next, for all $a^* \in [\underline{\beta} + B, \overline{\beta} + B)$, we have

$$g(a^*) = \frac{\theta}{1+\theta} \int_{a^*-B}^{\overline{\beta}} \left(\frac{a^* - \beta_i}{\overline{\beta} - \underline{\beta}} \right) d\beta_i.$$

The derivative is $\frac{dg(a^*)}{da^*} = \frac{\theta}{1+\theta} \left(\frac{\overline{\beta} - a^*}{\overline{\beta} - \underline{\beta}} \right) \leq 0$, where the inequality follows from $a^* \geq \underline{\beta} + B \geq \overline{\beta}$ (given $\overline{\beta} - \underline{\beta} \leq B$). The strict inequality holds as long as $a^* > \underline{\beta} + B$. Therefore, starting from $a^* = \underline{\beta} + B$, a marginal increase in the ideal decreases contribution.

Suppose $\overline{\beta} - \underline{\beta} > B$. Claim 3 restricts the possible range of the contribution-maximizing ideal to $a^* \in [\overline{\beta}, \overline{\beta} + B)$. This is because there are no overcontributing insiders for all $a^* < \overline{\beta}$, and $a^* = \overline{\beta}$ leads to higher contribution of undercontributing insiders than any $a^* < \overline{\beta}$. Next, for all $a^* \in [\overline{\beta}, \overline{\beta} + B)$, the expressions of $g(a^*)$ and $\frac{dg(a^*)}{da^*}$ are exactly the same as the previous case. This implies that, again, starting from $a^* = \overline{\beta}$, a marginal increase in the ideal decreases contribution.

A.3.2: Large preference heterogeneity with non-monotonic *f*

As an extension of Proposition 4, the following proposition characterizes the contribution-maximizing ideal under large preference heterogeneity for a class of non-monotonic density functions in a special case of γ converging to 0. We assume that there exists a unique $\hat{\beta} \in \left(\underline{\beta}, \overline{\beta}\right)$ such that $f'\left(\beta_i\right) > 0$ for $\beta_i \in \left(\underline{\beta}, \hat{\beta}\right)$, while $f'\left(\beta_i\right) < 0$ for $\beta_i \in \left(\hat{\beta}, \overline{\beta}\right)$. That is, the distribution has a single peak at $\hat{\beta}$. Some commonly used distribution functions, such as normal distributions, extreme value distributions are special cases of these single-peaked distributions.

Given γ converges to 0, agents with $\beta_i \geq a^*$ identify with the ideal but contribute their baseline, and incur no disutility. Thus, the ideal does not discourage high-demand agents from contributing, and we can focus on the trade-off between attracting more under-contributing members and increasing their contributions.

Proposition 6 *Suppose $\overline{\beta} - \underline{\beta} > B$, the density function f has a unique single peak $\hat{\beta} \in \left(\underline{\beta}, \overline{\beta}\right)$, and γ converges to 0. If $F(\underline{\beta} + B) \leq Bf(\underline{\beta})$, then $\bar{a} = \underline{\beta} + B$; otherwise, $\bar{a} > \underline{\beta} + B$ and it is the unique solution to $F(\bar{a}) - F(\bar{a} - B) = Bf(\bar{a} - B)$.*

Proof Given γ converges to 0, by Claims 1 and 3, we can focus on $a^* \in \left[\underline{\beta} + B, \overline{\beta} + B\right]$, in which

$$g(a^*) = \frac{\theta}{1+\theta} \int_{a^*-B}^{\min\{a^*, \overline{\beta}\}} \left(a^* - \beta_i\right) f\left(\beta_i\right) d\beta_i.$$

Therefore, $\frac{dg(a^*)}{da^*}$ is the same as $\chi_1(a^*)$ defined in (26). Since $F(a^*) = 1$ for all $a^* \geq \overline{\beta}$, $\frac{dg(a^*)}{da^*}$ can be simplified as

$$\frac{dg(a^*)}{da^*} = \frac{\theta}{1+\theta}(F(a^*) - F(a^* - B) - Bf(a^* - B)), \tag{28}$$

for all $a^* \in \left[\underline{\beta} + B, \overline{\beta} + B\right]$.

We first narrow down the possible range for $\bar{a}$. For all $a^* < \hat{\beta}$, we have $f'\left(\beta_i\right) > 0$ for all $\beta_i \in [a^* - B, a^*]$, which implies $\frac{dg(a^*)}{da^*} > 0$ by (28). For all $a^* > \hat{\beta} + B$, we have $f'\left(\beta_i\right) < 0$ for all $\beta_i \in [a^* - B, a^*]$, which implies $\frac{dg(a^*)}{da^*} < 0$ by (28). These two observations imply

$$\bar{a} \in \left[\max\left\{\underline{\beta} + B, \hat{\beta}\right\}, \hat{\beta} + B\right] \equiv H.$$

Next, we want to show (28) is single-crossing for all $a^* \in H$. Formally, we want to prove that, for all $a_L \in H$ such that $\frac{dg(a_L)}{da^*} \leq 0$, we have $\frac{dg(a^*)}{da^*} < 0$ for all $a^* \in H$ that satisfy $a^* > a_L$. To prove this, we first note $\frac{dg(a_L)}{da^*} \leq 0$ implies $f\left(a_L\right) \leq f(a_L - B)$. To see this, suppose by contradiction $f\left(a_L\right) > f(a_L - B)$; this implies $F\left(a_L\right) > F(a_L - B) + Bf\left(a_L - B\right)$ given single-peakedness, which further implies $\frac{dg(a_L)}{da^*} > 0$, a contradiction. Next, we note that $f\left(a_L\right) \leq f(a_L - B)$ implies $f(a^*) < f(a^* - B)$ given a_L, $a^* \in H$, and $a^* > a_L$. Then, for all $a^* \in H$ such that $a^* > a_L$, we have $\frac{dg(a^*)}{da^*} < 0$ because the derivative of (28) satisfies

$$\begin{aligned}
\frac{d^2 g(a^*)}{da^{*2}}|_{a^* \in H \text{ and } a^* > a_L} &= \frac{\theta}{1+\theta}\left(f(a^*) - f(a^* - B) - Bf'(a^* - B)\right) \\
&\leq \frac{\theta}{1+\theta}(f(a^*) - f(a^* - B)) \\
&< 0,
\end{aligned}$$

where the first inequality is due to $f'(a^* - B) \geq 0$ (since $a^* \in H$ implies $a^* - B \leq \hat{\beta}$), while the second inequality uses the fact that $f(a^*) < f(a^* - B)$.

We are now ready to prove the proposition. Suppose $F(\underline{\beta} + B) \leq Bf(\underline{\beta})$ so that $\frac{dg\left(\underline{\beta}+B\right)}{da^*} \leq 0$ by (28). The single-crossing property of (28) then implies $\frac{dg(a^*)}{da^*} < 0$ for all $a^* > \underline{\beta} + B$, so the contribution maximizing ideal $\bar{a} = \underline{\beta} + B$. Suppose instead $F(\underline{\beta} + B) > Bf(\underline{\beta})$, then by (28) the intermediate value theorem guarantees the existence of a solution to the first-order condition $F(\bar{a}) - F(\bar{a} - B) = Bf(\bar{a} - B)$ in compact interval H. Moreover, the single-crossing property of (28) guarantees that the first-order condition is sufficient for maximizing $g(a^*)$ and has a unique solution.

□

A.4: Welfare maximization (Propositions 2 and 5)

This subsection proves Propositions 2 and 5.

Denote $n = (N-1)E(\beta_i)$, so condition (11) implies

$$n > \left(\frac{2V\theta}{1+\theta}\right)^{1/2} = \frac{\theta}{1+\theta}B$$

by the definition of B. Since β_i's are i.i.d, (9) can be written as $M(a^*) = E\left[\beta_i a_i - \frac{1}{2}a_i^2\right] + nE\left(a_i\right)$. Given an insider's contribution in Equation (3) and an outsider's contribution β_i, after some simple algebra we obtain

$$\begin{aligned} E\left(a_i\right) = &\int_{\underline{\beta}}^{\bar{\beta}} \beta_i dF\left(\beta_i\right) \\ &+ \frac{\theta}{1+\theta}\int_{\max\{a^*-B,\underline{\beta}\}}^{\max\left\{\underline{\beta},\min\{a^*,\bar{\beta}\}\right\}} \left(a^* - \beta_i\right)dF\left(\beta_i\right) + \frac{\gamma}{1+\gamma}\int_{\max\left\{\underline{\beta},\min\{a^*,\bar{\beta}\}\right\}}^{\min\{a^*+D,\bar{\beta}\}} \left(a^* - \beta_i\right)dF\left(\beta_i\right), \end{aligned}$$

and

$$\begin{aligned} E\left[\beta_i a_i - \frac{1}{2}a_i^2\right] = &\frac{1}{2}\int_{\underline{\beta}}^{\bar{\beta}} \beta_i^2 dF\left(\beta_i\right) \\ &- \frac{1}{2}\int_{\max\{a^*-B,\underline{\beta}\}}^{\max\left\{\underline{\beta},\min\{a^*,\bar{\beta}\}\right\}} \left(\frac{\theta\left(a^* - \beta_i\right)}{1+\theta}\right)^2 dF\left(\beta_i\right) \\ &- \frac{1}{2}\int_{\max\left\{\underline{\beta},\min\{a^*,\bar{\beta}\}\right\}}^{\min\{a^*+D,\bar{\beta}\}} \left(\frac{\gamma\left(a^* - \beta_i\right)}{1+\gamma}\right)^2 dF\left(\beta_i\right). \end{aligned}$$

Substitute both expressions into $M(a^*)$ to obtain:

$$M(a^*) = M_0 + M_1(a^*) + M_2(a^*), \tag{29}$$

where the first term

$$M_0 \equiv \int_{\underline{\beta}}^{\bar{\beta}} \left(\frac{1}{2}\beta_i^2 + n\beta_i\right)dF\left(\beta_i\right)$$

is independent of a^*, while

$$M_1(a^*) \equiv \frac{\theta}{1+\theta} \int_{\max\{a^*-B,\underline{\beta}\}}^{\max\left\{\underline{\beta},\min\{a^*,\overline{\beta}\}\right\}} \left[n\left(a^* - \beta_i\right) - \frac{1}{2}\frac{\theta\left(a^* - \beta_i\right)^2}{1+\theta} \right] dF\left(\beta_i\right). \quad (30)$$

$$M_2(a^*) \equiv \frac{\gamma}{1+\gamma} \int_{\max\left\{\underline{\beta},\min\{a^*,\overline{\beta}\}\right\}}^{\min\{a^*+D,\overline{\beta}\}} \left[n\left(a^* - \beta_i\right) - \frac{1}{2}\frac{\gamma\left(a^* - \beta_i\right)^2}{1+\gamma} \right] dF\left(\beta_i\right) \quad (31)$$

respectively represent the welfare change of under- and over-contributing insiders.

A.4.1: Preliminaries

We first prove a series of preliminary claims concerning M_1 and M_2.

Claim 4 *For all a^*, denote the integrands of M_1 and M_2 as $m_1\left(\beta_i;a^*\right)$ and$m_2\left(\beta_i;a^*\right)$ respectively. Then,*

- *for all β_i in the integration interval of M_1, $m_1\left(\beta_i;a^*\right) > 0$ if $\beta_i < a^*$; $m_1(a^*;a^*) = 0$; moreover, m_1 is increasing in a^*;*
- *for all β_i in the integration interval of M_2, $m_2\left(\beta_i;a^*\right) < 0$ if $\beta_i > a^*$; $m_2(a^*;a^*) = 0$; moreover, m_2 is increasing in a^*.*

Proof Consider m_1. If $a^* \leq \underline{\beta}$, then the integration interval of M_1 is empty. Suppose instead $a^* > \underline{\beta}$, then the integration interval is $[\max\{a^* - B, \underline{\beta}\}, \min\left\{a^*, \overline{\beta}\right\}]$, and

$$\begin{aligned} \frac{dm_1\left(\beta_i;a^*\right)}{d\beta_i}\Big|_{\beta_i\in[\max\{a^*-B,\underline{\beta}\},\min\left\{a^*,\overline{\beta}\right\}]} &= -n + \frac{\theta}{1+\theta}\left(a^* - \beta_i\right) \\ &\leq -n + \frac{\theta}{1+\theta}B \\ &< 0 \text{(by condition (11))}, \end{aligned}$$

where the first inequality comes from $\beta_i \geq a^* - B$. For all $\beta_i < a^*$ in the integration interval, $\frac{dm_1}{d\beta_i} < 0$ implies $m_1\left(\beta_i;a^*\right) > m_1(a^*;a^*) = 0$, where the equality follows from the definition of m_1. Finally,

$$\frac{dm_1}{da^*} = -\frac{dm_1}{d\beta_i} = n - \frac{\theta}{1+\theta}\left(a^* - \beta_i\right) > 0.$$

Consider m_2. If $a^* \geq \overline{\beta}$, then the integration interval of M_2 is empty. Suppose instead $a^* < \overline{\beta}$, then the integration interval is $[\max\left\{\underline{\beta}, a^*\right\}, \min\{a^* + D, \overline{\beta}\}]$, and

$$\frac{dm_2\left(\beta_i;a^*\right)}{d\beta_i}\Big|_{\beta_i\in[\max\left\{\underline{\beta},a^*\right\},\min\{a^*+D,\overline{\beta}\}]} = -n + \frac{\gamma}{1+\gamma}\left(a^* - \beta_i\right) < 0,$$

where the inequality follows from $\beta_i \geq a^*$.

For all $\beta_i > a^*$ in the integration interval, $\frac{dm_2}{d\beta_i} < 0$ implies $m_2\left(\beta_i;a^*\right) < m_2(a^*;a^*) = 0$, where the equality follows from the definition of m_2. Finally, we have $\frac{dm_2}{da^*} = -\frac{dm_2}{d\beta_i} > 0$. □

Claim 5 *For all a^*, $M_1(a^*) \geq 0$ and $M_2(a^*) \leq 0$, and each of these inequalities is strict when the corresponding integration interval is non-empty.*

Proof The claim follows immediately from $m_1\left(\beta_i;a^*\right) \geq 0$ and $m_2\left(\beta_i;a^*\right) \leq 0$ by Claim 4. □

Claim 6 *For all $a^* \in (\underline{\beta}, \underline{\beta} + B]$, $\frac{dM_1(a^*)}{da^*} > 0$.*

Proof For the specified range of a^*,

$$M_1(a^*) = \frac{\theta}{1+\theta}\int_{\underline{\beta}}^{\min\{a^*,\overline{\beta}\}} m_1\left(\beta_i;a^*\right)dF\left(\beta_i\right).$$

$$\frac{dM_1(a^*)}{da^*} = \left\{\begin{array}{ll} \frac{\theta}{1+\theta}\int_{\underline{\beta}}^{a^*}\frac{dm_1\left(\beta_i;a^*\right)}{da^*}dF\left(\beta_i\right) + m_1(a^*;a^*)f(a^*) & \text{if } a^* \leq \overline{\beta} \\ \frac{\theta}{1+\theta}\int_{\underline{\beta}}^{\overline{\beta}}\frac{dm_1\left(\beta_i;a^*\right)}{da^*}dF\left(\beta_i\right) & \text{if } a^* > \overline{\beta} \end{array}\right\} > 0,$$

where we have utilized $m_1(a^*;a^*) = 0$ and $\frac{dm_1\left(\beta_i;a^*\right)}{da^*} > 0$ from Claim 4. □

Claim 7 *Suppose f is monotonic, i.e. $f' \leq 0$ or $f' \geq 0$ for all β_i. For all $a^* \in (\underline{\beta} + B, \overline{\beta} + B)$:*

- (i) *if $a^* \leq \overline{\beta}$, then $\frac{dM_1(a^*)}{da^*}$ has the same sign as f';*
- (ii) *if $a^* \geq \overline{\beta}$, then*

$$\frac{dM_1(a^*)}{da^*} = \frac{1}{2}\Phi_1 + \frac{\theta}{1+\theta}\left(n - \frac{1}{2}\frac{\theta B}{1+\theta}\right)(1 - F(a^* - B) - Bf(a^* - B)),$$

where Φ_1 is defined in (33) and has exactly the same sign as f'.

Proof For the specified range of a^*,

$$M_1(a^*) = \frac{\theta}{1+\theta}\int_{a^*-B}^{\min\{a^*,\overline{\beta}\}} m_1\left(\beta_i;a^*\right)dF\left(\beta_i\right).$$

$$\frac{dM_1(a^*)}{da^*} = \left\{\begin{array}{ll} \frac{\theta}{1+\theta}\left[\int_{a^*-B}^{a^*}\frac{dm_1\left(\beta_i;a^*\right)}{da^*}dF\left(\beta_i\right) - m_1(a^*-B;a^*)f(a^*-B) + m_1(a^*;a^*)f(a^*)\right] & \text{if } a^* \leq \overline{\beta} \\ \frac{\theta}{1+\theta}\left[\int_{a^*-B}^{\overline{\beta}}\frac{dm_1\left(\beta_i;a^*\right)}{da^*}dF\left(\beta_i\right) - m_1(a^*-B;a^*)f(a^*-B)\right] & \text{if } a^* > \overline{\beta} \end{array}\right\}.$$

Applying $m_1(a^*;a^*) = 0$ and substituting the definition of m_1 and $\frac{dm_1}{da^*}$ in Claim 4 and its proof to the above, we get

$$\begin{aligned}\frac{dM_1(a^*)}{da^*} =& \frac{\theta}{1+\theta}\left(\int_{a^*-B}^{\min\{a^*,\bar{\beta}\}}\left(n - \frac{\theta\left(a^* - \beta_i\right)}{1+\theta}\right)dF\left(\beta_i\right)\right. \\ &\left. -\left(n - \frac{1}{2}\frac{\theta B}{1+\theta}\right)Bf(a^* - B)\right).\end{aligned} \tag{32}$$

Define

$$\begin{aligned}\Phi_1 &= \left(\frac{\theta}{1+\theta}\right)^2 \int_{a^*-B}^{\min\{a^*,\bar{\beta}\}} \left[B - 2\left(a^* - \beta_i\right)\right]dF\left(\beta_i\right), \\ \Phi_2 &= F\left(\min\{a^*, \bar{\beta}\}\right) - F(a^* - B) - Bf(a^* - B).\end{aligned} \tag{33}$$

By adding and subtracting $\frac{1}{2}\frac{\theta B}{1+\theta}$ to the integrand in (32) and some rearrangements, we get:

$$\begin{aligned}\frac{dM_1(a^*)}{da^*} &= \frac{\theta}{1+\theta}\int_{a^*-B}^{\min\{a^*,\bar{\beta}\}}\left[n - \frac{\theta\left(a^* - \beta_i\right)}{1+\theta} + \frac{1}{2}\frac{\theta B}{1+\theta} - \frac{1}{2}\frac{\theta B}{1+\theta}\right] \\ &\quad - \frac{\theta}{1+\theta}\left(n - \frac{1}{2}\frac{\theta B}{1+\theta}\right)Bf(a^* - B) \\ &= \frac{\theta}{1+\theta}\int_{a^*-B}^{\min\{a^*,\bar{\beta}\}}\left[-\frac{\theta\left(a^* - \beta_i\right)}{1+\theta} + \frac{1}{2}\frac{\theta B}{1+\theta}\right]dF\left(\beta_i\right) \\ &\quad + \frac{\theta}{1+\theta}\left(n - \frac{1}{2}\frac{\theta B}{1+\theta}\right)\int_{a^*-B}^{\min\{a^*,\bar{\beta}\}} dF\left(\beta_i\right) \\ &\quad - \frac{\theta}{1+\theta}\left(n - \frac{1}{2}\frac{\theta B}{1+\theta}\right)Bf(a^* - B) \\ &= \frac{\theta}{1+\theta}\int_{a^*-B}^{\min\{a^*,\bar{\beta}\}}\frac{\theta}{1+\theta}\left[\frac{B - 2\left(a^* - \beta_i\right)}{2}\right]dF\left(\beta_i\right) \\ &\quad + \frac{\theta}{1+\theta}\left(n - \frac{1}{2}\frac{\theta B}{1+\theta}\right)\left(F\left(\min\{a^*,\bar{\beta}\}\right) - F(a^* - B) - Bf(a^* - B)\right) \\ &= \frac{1}{2}\Phi_1 + \frac{\theta}{1+\theta}\left(n - \frac{1}{2}\frac{\theta B}{1+\theta}\right)\Phi_2.\end{aligned}$$

Case (i): $a^* \le \bar{\beta}$. By condition (11), $n > \frac{1}{2}\frac{\theta B}{1+\theta}$. Therefore, the sign of $\frac{dM_1(a^*)}{da^*}$ is pinned down whenever Φ_1 and Φ_2 have the same sign. Using integration by parts,

$$\Phi_1 = \left(\frac{\theta}{1+\theta}\right)^2\left(BF(a^*) + BF(a^* - B) - 2\int_{a^*-B}^{a^*} F\left(\beta_i\right)d\beta_i\right). \tag{34}$$

By inspection, we find that $\Phi_1 = 0$ if F is linear (i.e. $f' = 0$ for all β_i); $\Phi_1 \le (<)0$ if F is (strictly) concave (i.e. $f' \le (<)0$ for all β_i); and $\Phi_1 \ge (>)0$ if F is (strictly) convex (i.e. $f' \geq (>)0$ for all β_i). Therefore, Φ_1 has the same sign as f' when f is monotonic. Moreover,

$$\Phi_2 = F(a^*) - F(a^* - B) - Bf(a^* - B),$$

which also has the same sign as f' based on direct inspection. Putting the results together, we know $\frac{dM_1(a^*)}{da^*}$ has the same sign as f'.

Case (ii): $a^* \geq \overline{\beta}$. We note

$$\begin{aligned}\Phi_1 &= \left(\frac{\theta}{1+\theta}\right)^2 \int_{a^*-B}^{\overline{\beta}} \left[B - 2\left(a^* - \beta_i\right)\right] dF\left(\beta_i\right) \\ &= \left(\frac{\theta}{1+\theta}\right)^2 \int_{a^*-B}^{a^*} \left[B - 2\left(a^* - \beta_i\right)\right] dF\left(\beta_i\right)\end{aligned}$$

because distribution F has zero density for $\beta_i > \overline{\beta}$. Using integration by parts, Φ_1 becomes (34) exactly, and so Φ_1 has the same sign as f', as shown in Case (i) above. Meanwhile, $\Phi_2 = 1 - F(a^* - B) - Bf(a^* - B)$. Part (ii) of the claim is thus proved. □

A.4.2: Narrowing down the range of $\hat{a}$

We now prove a series of claims to narrow down the possible range of welfare-maximizing ideal, $\hat{a}$.

Claim 8 $\hat{a} \in \left(\underline{\beta}, \overline{\beta} + B\right)$.

Proof For all $a^* \leq \underline{\beta}$, all insiders (if any) are over-contributors so $M(a^*) = M_0 + M_2(a^*) \leq M_0$, where the inequality is due to Claim 5. For all $a^* \geq \overline{\beta} + B$, $M(a^*) = M_0$ because no agents join as insiders. All these ideals are dominated by $a^* = \overline{\beta}$, where $M(a^*) = M_0 + M_1(a^*) > M_0$, where the inequality is due to Claim 5. □

Claim 9 *Suppose* $\overline{\beta} - \underline{\beta} \leq B$, *then* $\hat{a} \geq \underline{\beta} + B$.

Proof By Claim 8, any $a^* \leq \underline{\beta}$ can never be optimal. For all $a^* \in \left(\underline{\beta}, \underline{\beta} + B\right)$

$$\begin{aligned}M(a^*)|_{a^* < \underline{\beta}+B} &= M_0 + M_1(a^*) + M_2(a^*) \\ &\leq M_0 + M_1(a^*) \\ &< M_0 + M_1\left(\underline{\beta} + B\right) \\ &= M\left(\underline{\beta} + B\right)\end{aligned}$$

where the first inequality uses $M_2(a^*) \leq 0$ by Claim 5, the second inequality uses $M_1(a^*)$ being increasing by Claim 6, the final equality is due to $M_2\left(\underline{\beta} + B\right) = 0$ utilizing the supposition that $\overline{\beta} \leq \underline{\beta} + B$. □

Claim 10 *Suppose $\overline{\beta} - \underline{\beta} > B$. If $f' \geq 0$ for all β_i, then $\hat{a} \geq \overline{\beta}$.*

Proof By Claim 8, any $a^* \leq \underline{\beta}$ can never be optimal. Suppose $f' \geq 0$ for all β_i and consider $a^* \in \left(\underline{\beta}, \overline{\beta}\right)$,

$$\begin{aligned} M(a^*)|_{a^*<\overline{\beta}} &= M_0 + M_1(a^*) + M_2(a^*) \\ &< M_0 + M_1(a^*) \\ &\leq M_0 + M_1\left(\overline{\beta}\right) \\ &= M\left(\overline{\beta}\right), \end{aligned}$$

where the first inequality uses $M_2(a^*) < 0$ by Claim 5 (note $a^* < \overline{\beta}$ means M_2 has a non-empty integration interval). The second inequality is due to $M_1(a^*)$ being (weakly) increasing for all $a^* \in (\underline{\beta}, \overline{\beta}]$, which is because of Claim 6, 7(i) and the assumption that $f' \geq 0$ for all β_i. The final equality follows from $M_2\left(\overline{\beta}\right) = 0$ by the definition of M_2. □

A.4.3: Completing the proofs

We assume that condition (11) is satisfied throughout this subsection, that is, $n > \frac{1}{2}\frac{\theta B}{1+\theta}$.

Claim 11 *Suppose $\overline{\beta} - \underline{\beta} \leq B$. If $f' \leq 0$ for all β_i, then $\hat{a} = \underline{\beta} + B = \bar{a}$. If $f' \geq 0$ for all β_i, then $\hat{a} \geq \bar{a}$.*

Proof By Claims 8 and 9, we can focus on $a^* \in \left[\underline{\beta} + B, \overline{\beta} + B\right]$, for which $M(a^*) = M_0 + M_1(a^*)$ and so $\frac{dM(a^*)}{da^*} = \frac{dM_1(a^*)}{da^*}$. For all $a^* \geq \underline{\beta} + B \geq \overline{\beta}$, Claim 7(ii) implies

$$\begin{aligned} \frac{dM(a^*)}{da^*} &= \frac{dM_1(a^*)}{da^*} \\ &= \frac{1}{2}\Phi_1 + \frac{\theta}{1+\theta}\left(n - \frac{1}{2}\frac{\theta B}{1+\theta}\right)(1 - F(a^* - B) - Bf(a^* - B)), \end{aligned}$$

where Φ_1 has the same sign as f'. Therefore, if $f' \leq 0$ for all β_i, then $\Phi_1 \leq 0$. Meanwhile $f' \leq 0$ implies F is concave, so for all $a^* \geq \underline{\beta} + B$,

$$\begin{aligned}
&1 - F(a^* - B) - Bf(a^* - B) \\
&\leq 1 - F(a^* - B) - \left(\overline{\beta} - \underline{\beta}\right) f(a^* - B) \\
&\leq 1 - F(a^* - B) - \left[\overline{\beta} - (a^* - B)\right] f(a^* - B) \\
&\leq 0,
\end{aligned}$$

where the first inequality follows from $\overline{\beta} - \underline{\beta} \leq B$, the second follows from $a^* - B \geq \underline{\beta}$, and the third from $f' \leq 0$ for all β_i. For all $a^* > \underline{\beta} + B$, the second inequality is strict and so $1 - F(a^* - B) - Bf(a^* - B) < 0$. These results imply that $\frac{dM(a^*)}{da^*}(\leq) < 0$ for all $a^*(\geq) > \underline{\beta} + B$. Therefore, $\hat{a} = \underline{\beta} + B$, which coincides with $\bar{a} = \underline{\beta} + B$ in Proposition 3 (by Footnote 14).

If $f' \geq 0$ for all β_i, then $\Phi_1 \geq 0$ by Claim 7(ii). Since $f' \geq 0$ for all β_i, $1 \geq \left(\overline{\beta} - \underline{\beta}\right) f(\underline{\beta}) \geq Bf(\underline{\beta})$, implying $\frac{1}{f(\underline{\beta})} \geq B$.

If $\frac{1}{f(\underline{\beta})} > B$, Proposition 3 and its proof show that the contribution maximizing $\bar{a}$ is determined by

$$1 - F(\bar{a} - B) - Bf\left(\bar{a} - B\right) = 0, \tag{35}$$

while for all $a^* \in \left[\underline{\beta} + B, \bar{a}\right)$,

$$1 - F(a^* - B) - Bf(a^* - B) > 0.$$

These imply that $\frac{dM(a^*)}{da^*} > 0$ for all $a^* \in \left[\underline{\beta} + B, \bar{a}\right)$ and $\frac{dM(a^*)}{da^*} \geq 0$ for $a^* = \bar{a}$, and so $\hat{a} \geq \bar{a}$.

If $\frac{1}{f(\underline{\beta})} = B$, Proposition 3 shows that $\bar{a} = \underline{\beta} + B$, and so $\hat{a} \geq \bar{a}$ by Claim 9. □

Claim 12 *Suppose $\overline{\beta} - \underline{\beta} > B$. If $f' = 0$ for all β_i then $\hat{a} = \overline{\beta} = \bar{a}$. If $f' \geq 0$ for all β_i then $\hat{a} \geq \bar{a}$.*

Proof By Claims 8 and 10, we can focus on $a^* \in \left[\overline{\beta}, \overline{\beta} + B\right]$, for which $M(a^*) = M_0 + M_1(a^*)$. For all $a^* \geq \overline{\beta}$, Claim 7(ii) implies

$$\begin{aligned}
\frac{dM(a^*)}{da^*} = \frac{dM_1(a^*)}{da^*} = \frac{1}{2}\Phi_1 \\
+ \frac{\theta}{1+\theta}\left(n - \frac{1}{2}\frac{\theta B}{1+\theta}\right)(1 - F(a^* - B) - Bf(a^* - B)),
\end{aligned}$$

where Φ_1 has the same sign as f'. If $f' = 0$ for all β_i, then $\Phi_1 = 0$. Meanwhile $f' = 0$ implies $1 - F(a^* - B) - Bf(a^* - B) = \frac{\overline{\beta} - (a^* - B)}{\overline{\beta} - \underline{\beta}} - \frac{B}{\overline{\beta} - \underline{\beta}} = \frac{\overline{\beta} - a^*}{\overline{\beta} - \underline{\beta}} \leq 0$ for all $a^* \geq \overline{\beta}$, where the inequality is strict for all $a^* > \overline{\beta}$. Therefore, $\frac{dM(a^*)}{da^*} < 0$ for all $a^* > \overline{\beta}$ and so $\hat{a} = \overline{\beta}$, which coincides with $\bar{a} = \overline{\beta}$ as stated in Proposition 1 when $\overline{\beta} - \underline{\beta} > B$.

Suppose $f' \geq 0$ for all β_i. If $\frac{1 - F(\overline{\beta} - B)}{f(\overline{\beta} - B)} \leq B$, Proposition 4(i) shows that $\bar{a} = \overline{\beta}$; by Claim 10, $\hat{a} \geq \overline{\beta} = \bar{a}$. Otherwise, Proposition 4(i) and its proof show that $\bar{a} > \overline{\beta}$, which is determined by $1-F(\bar{a}-B)=Bf(\bar{a}-B)$, and that for all $a^*\in\left[\overline{\beta},\bar{a}\right)$, $1-F(a^*-B)-Bf(a^*-B)>0$. Moreover, $\Phi_1\geq 0$ since $f'\geq 0$ for all β_i by Claim 7(ii). These results show that $\frac{dM(a^*)}{da^*}>0$ for all $a^*\in\left[\overline{\beta},\bar{a}\right)$ and so $\hat{a}\geq\bar{a}$. □

Claim 13 *Suppose $\overline{\beta}-\underline{\beta}>B$. If $f'<0$ for all β_i and γ coverges to 0 then $\hat{a}=\underline{\beta}+B=\bar{a}$.*

Proof By Claim 8, we can focus on $a^*\in\left[\underline{\beta},\overline{\beta}+B\right]$. Given $\lim_{\gamma\to 0}M_2(a^*)=0$, we have $M(a^*)=M_0+M_1(a^*)$. For $a^*\leq\underline{\beta}+B$, $M_1(a^*)$ is increasing in a^* by Claim 6. For $a^*\in(\underline{\beta}+B,\overline{\beta}]$, $\frac{dM_1(a^*)}{da^*}$ has the same sign as $f'<0$ for all β_i by Claim 7(i). For $a^*\geq\overline{\beta}$,

$$
\begin{aligned}
&1-F(a^*-B)-Bf(a^*-B)\\
&\quad\leq 1-F(a^*-B)-\left[\overline{\beta}-(a^*-B)\right]f(a^*-B)<0,
\end{aligned}
$$

where the first inequality follows from $a^*\geq\overline{\beta}$, and the second from $f'<0$ for all β_i. Therefore,

$$
\frac{dM_1(a^*)}{da^*}=\frac{1}{2}\Phi_1+\frac{\theta}{1+\theta}\left(n-\frac{1}{2}\frac{\theta B}{1+\theta}\right)(1-F(a^*-B)-Bf(a^*-B))\ <0,
$$

where $\Phi_1<0$ since $f'<0$ for all β_i by Claim 7(ii). Therefore, $M_1(a^*)$ is maximized at $\hat{a}=\underline{\beta}+B$, which coincides with $\bar{a}=\overline{\beta}$ in the second part of Proposition 4. □

We are now ready to prove Propositions 2 and 5.

Proof (**Proposition** 2) Given $f'=0$ for all β_i, combining Claims 11 and 12 immediately yields the proposition. □

Proof (**Proposition** 5) For the first part, we divide (30) and (31) by n so that M_1+M_2 is proportional to

$$
\begin{aligned}
&\frac{\theta}{1+\theta}\int_{\max\{a^*-B,\underline{\beta}\}}^{\max\left\{\underline{\beta},\min\{a^*,\overline{\beta}\}\right\}}\left[\left(a^*-\beta_i\right)-\frac{1}{2n}\frac{\theta\left(a^*-\beta_i\right)^2}{1+\theta}\right]dF\left(\beta_i\right)\\
&+\frac{\gamma}{1+\gamma}\int_{\max\left\{\underline{\beta},\min\{a^*,\overline{\beta}\}\right\}}^{\min\{a^*+D,\overline{\beta}\}}\left[\left(a^*-\beta_i\right)-\frac{1}{2n}\frac{\gamma\left(a^*-\beta_i\right)^2}{1+\gamma}\right]dF\left(\beta_i\right).
\end{aligned}
$$

When n coverges to ∞, the expression above converges to

$$
\begin{aligned}
&\frac{\theta}{1+\theta}\int_{\max\{a^*-B,\underline{\beta}\}}^{\max\left\{\underline{\beta},\min\{a^*,\overline{\beta}\}\right\}}\left(a^*-\beta_i\right)dF\left(\beta_i\right)\\
&+\frac{\gamma}{1+\gamma}\int_{\max\left\{\underline{\beta},\min\{a^*,\overline{\beta}\}\right\}}^{\min\{a^*+D,\overline{\beta}\}}\left(a^*-\beta_i\right)dF\left(\beta_i\right)\equiv g(a^*).
\end{aligned}
$$

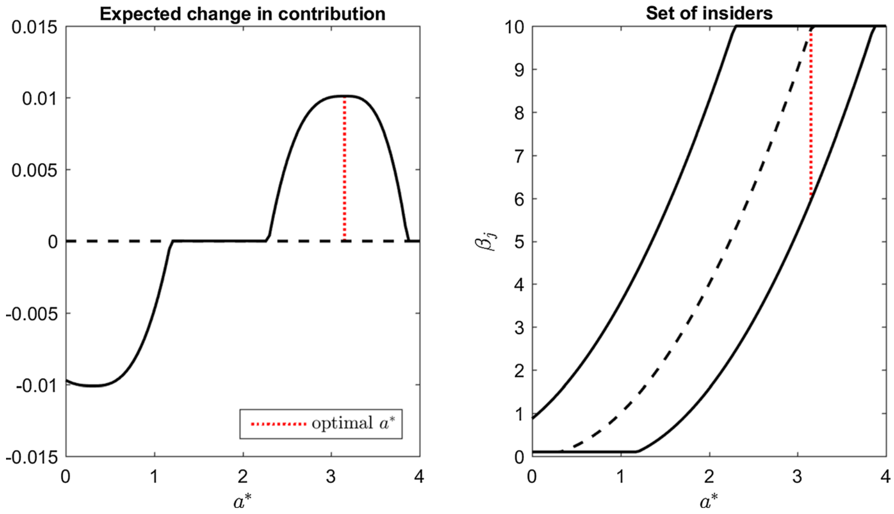

Fig. 4 Cubic cost functions: large preference heterogeneity ($\bar{\beta} = 10$). Left: expected change in contribution. Right: set of insiders

It follows that the maximizer of $M(a^*)$ is the same as the maximizer of $g(a^*)$ when n coverges to ∞, which is implied by N converging to ∞. The first part of the proposition is thus proved. Combining Claims 11 - 13 yields the last two parts of the proposition. □

Appendix B: Extension: alternative functional forms

In this appendix, we discuss two extensions of Sect. 5.1 using alternative functional forms. Section B.1 replaces quadratic cost functions by alternative convex functions and show that the results in Sect. 5.1 still hold. Section B.2 replaces the linear benefit function by a concave function, and uses an example to show that the main insights in Sect. 5.1 remain unchanged. We maintain the assumption of uniform distribution of types.

B.1: Alternative cost functions

Let the quadratic cost function of contributing to the public good $\frac{1}{2}a^2$ be replaced by $\frac{1}{p}a^p$, and let the disutility of an undercontributing insider i be $\frac{\theta}{p}\left(a^* - a_i\right)^p$ and that of an overcontributing insider be $\frac{\gamma}{p}\left(a_i - a^*\right)^p$, where $p > 1$.

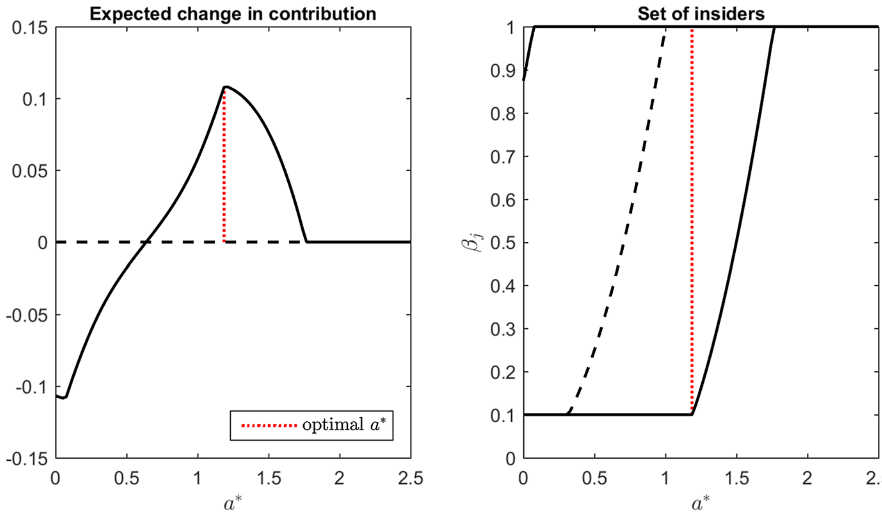

Fig. 5 Cubic cost functions: small preference heterogeneity ($\bar{\beta} = 1$). Left: expected change in contribution. Right: set of insiders

The contribution of an outsider with type β_i, denoted by $a_i^B \equiv a^B(\beta_i)$, is determined by the following first order condition:

$$\left(a_i^B\right)^{p-1} = \beta_i.$$

Calculation similar to that in the proof of Lemma 1 shows that the contribution of an insider with type β_i, denoted by $a_i^I \equiv a^I(\beta_i)$, is determined by the following first order conditions.

$$\begin{aligned} \left(a_i^I\right)^{p-1} &= \beta_i + \theta\left(a^* - a_i^I\right)^{p-1} \text{ if } \beta_i < (a^*)^{p-1} \\ \left(a_i^I\right)^{p-1} &= \beta_i - \gamma\left(a_i^I - a^*\right)^{p-1} \text{ if } \beta_i \geq (a^*)^{p-1} \end{aligned} \tag{36}$$

It is straightforward to verify from (36) that

$$\begin{aligned} a_i^I > a_i^B \text{ if } \beta_i < (a^*)^{p-1} \\ a_i^I = a_i^B \text{ if } \beta_i = (a^*)^{p-1} \\ a_i^I < a_i^B \text{ if } \beta_i > (a^*)^{p-1} \end{aligned}.$$

When $p = 2$, we get Eq. (3) from Equation (36).

For any agent i, substituting the optimal contributions as an insider and as an outsider back to the utility functions and using the envelope theorem, we can show that the difference in the utility of being an insider and that of being an outsider, denoted by $U_i^I - U_i^O$, satisfies

$$\frac{d\left[U_i^I - U_i^O\right]}{d\beta_i} = \beta_i\left[a_i^I - a_i^B\right],$$

which is positive if and only if $\beta_i < (a^*)^{p-1}$, implying that $U_i^I - U_i^O$ is increasing in β_i for $\beta_i < (a^*)^{p-1}$ and decreasing in β_i for $\beta_i > (a^*)^{p-1}$. Therefore,

$$\max_{\beta_i} \left(U_i^I - U_i^O\right) = \left(U_i^I - U_i^O\right)|_{\beta_i=(a^*)^{p-1}} = V > 0.$$

Suppose $(a^*)^{p-1} \in \left[\underline{\beta}, \overline{\beta}\right]$. Then, we can define I^- as the root of $U_i^I - U_i^O = 0$ over interval $\left[\underline{\beta}, (a^*)^{p-1}\right]$, which is unique (whenever it exists) because $U_i^I - U_i^O$ is monotonic over this interval. If the root does not exist then we define $I^- = \underline{\beta}$. Likewise, we can define I^+ as the root of $U_i^I - U_i^O = 0$ over interval $\left[(a^*)^{p-1}, \bar{\beta}\right]$, and if the root does not exist then we define $I^+ = \bar{\beta}$. We thus have

$$U_i^I \geq U_i^O \iff \beta_i \in \left[I^-, I^+\right].$$

This implies that any agent with type β_i will choose to be an insider if and only if $\beta_i \in \left[I^-, I^+\right]$. This result corresponds to Lemma 2.

We now turn to stage 0, when the social leader chooses a^*. We rely on numerical simulations to look for a^* that maximizes the expected contribution. Let $p = 3$, meaning that the cost functions are cubic. Let $\theta = 1$, $\gamma = 0.5$, $V = 0.1$, and $\underline{\beta} = 0.1$. We consider two cases: (i) $\bar{\beta} = 10$ (large preference heterogeneity) and (ii) $\bar{\beta} = 1$ (small preference heterogeneity).

Figure 4 shows the case with $\bar{\beta} = 10$ (large preference heterogeneity). In both graphs of Fig. 4, the horizontal axis represents a^*. The left graph of Fig. 4 plots expected change in contribution as a^* varies. In the right graph, the vertical axis indicates player type β_i; the vertical interval between the two solid black curves indicates the set of insiders. Among these insiders, those with β_i above the black dashed curve are over-contributors, while those with β_i below the black dashed curve are under-contributors. In both graphs, the red dotted line corresponds to the contribution maximizing ideal. Figure 4 shows that, with large preference heterogeneity, the contribution maximizing ideal attracts only high-type agents to be insiders. Moreover, under this ideal, there are no over-contributing insiders.

Figure 5 shows the expected change in contribution in the left graph and the set of insiders in the right graph, with $\bar{\beta} = 1$ (small preference heterogeneity). Under the contribution-maximizing ideal (denoted by the red dotted line), all agents choose to be insiders. Moreover, under this ideal, an agent with type $\bar{\beta}$ is indifferent between becoming an insider or remaining an outsider. When the ideal increases slightly, the agent will drop out and the set of insiders shrinks.

All these results are consistent with our analysis with quadratic cost functions in Sect. 5.1.

B.2: Concave benefits

This subsection discusses an alternative direction of extending our model in Sect. 5.1: using alternative benefit functions. The assumptions of linear benefit and convex cost of the public good adopted in our model drastically simplify our analysis, as under these assumptions, each individual has a dominant strategy in

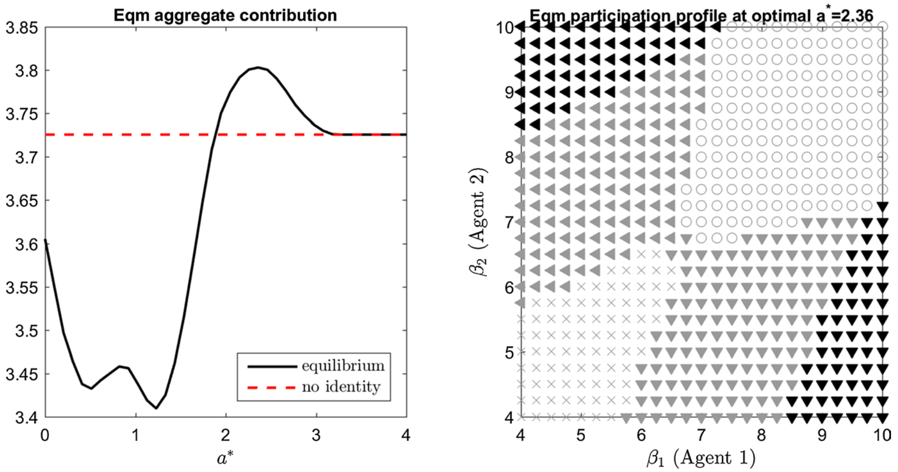

Fig. 6 Logarithm benefits: large preference heterogeneity ($\overline{\beta} = 10$). Left: expected equilibrium aggregate contribution; right: participation profile under the contribution-maximizing ideal

contribution. In this section, we will use an alternative benefit function for the public good, thereby introducing strategic interdependence in contribution between players. Since introducing strategic interdependence largely complicates the analysis, to keep tractability, we will consider the simplest case of only two agents and assume $\theta = \gamma$. We will use a numerical example to show that our main results in Sect. 5.1 largely remain unchanged.

Let the benefit of the public good to individual i be a concave function $\beta_i \Gamma(a_i + \sum_{j \neq i} a_j)$ instead of the linear function $\beta_i(a_i + \sum_{j \neq i} a_j)$. Let Γ' be the derivative of Γ. Consider a simplest case of only two agents ($N = 2$), agent 1 and agent 2. Further assume symmetric disutility of deviating from the deal for insiders: $\theta = \gamma$, as in Akerlof and Kranton (2002). The types of the two agents, β_1 and β_2, are independent and follow the uniform distribution in $\left[\underline{\beta}, \overline{\beta}\right]$ from the stage 0's point of view, as in Sect. 5.1. The agents observe each other's type before their respective decisions. In what follows, we denote $A \equiv a_1 + a_2$ as the aggregate contribution.

Consider a given realized pair of $\left(\beta_1, \beta_2\right)$. Simple calculation shows that, in the contribution stage, the best response function of any outsider j is determined by $a^O\left(\beta_j\right) = \beta_j \Gamma'(A)$, where superscript O denotes outsiders, while the best response function of any insider i is determined by

$$a_i^I = \frac{\theta}{1+\theta}\left(a^* - \beta_i \Gamma'(A)\right) + \beta_i \Gamma'(A).$$

Both outsiders' and insiders' individual contributions thus depend on the contribution of the other player. Contributions are strategic substitutes. Using the two best response functions, we can pin down the aggregate contribution in the equilibrium, which is determined by

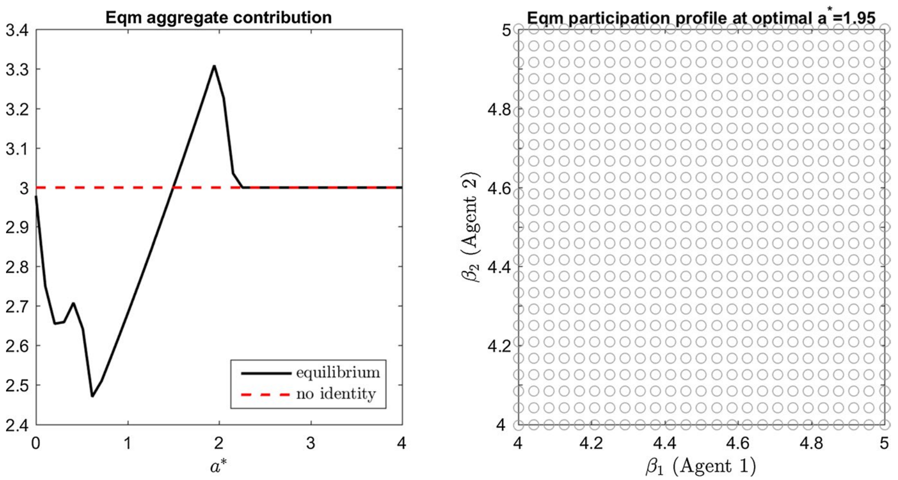

Fig. 7 Logarithm benefits: small preference heterogeneity ($\overline{\beta} = 5$). Left: expected equilibrium aggregate contribution; right: participation profile under the contribution-maximizing ideal

$$A = \left(\beta_1 + \beta_2\right)\Gamma'(A) + \frac{\theta}{1+\theta}\sum_{i\in I}\left(a^* - \beta_i\Gamma'(A)\right), \tag{37}$$

where I is the set of insiders. In other words, the equilibrium aggregate contribution is characterized by the fixed point A that solves (37). An inspection on (37) suggests that aggregate contribution is increasing in a^* as long as I is not affected.

In the stage of self-selection (stage 1), a player will compare the utility of becoming an insider and that of remaining an outsider, conditional on her belief about what the other player will do. This causes tricky strategic interdependence as a player may prefer to become an insider when she expects the other player to stay outside, and may prefer to stay outside when she expects the other player will become an insider. In general, for each realized pair of $\left(\beta_1, \beta_2\right)$, we have the following 2×2 payoff matrix:

Agent 1	Agent 2	
	In	Out
In	$U_1(In, In), U_2(In, In)$	$U_1(In, Out), U_2(In, Out)$
Out	$U_1(Out, In), U_2(Out, In)$	$U_1(Out, Out), U_2(Out, Out)$

where, for example, $U_1(In, Out)$ and $U_2(In, Out)$ denote the utility for agent 1 and agent 2 respectively, if agent 1 chooses to become an insider and agent 2 chooses to remain an outsider.

Given a specific functional form of $\Gamma(.)$ and parameters θ, V, β_1 and β_2, we will be able to solve the equilibrium numerically for any given a^*.

We now consider the social leader's choice at Stage 0, when she cannot observe β_1 and β_2 but knows that β_1 and β_2 are independent and each of them follows the uniform distribution in $\left[\underline{\beta}, \overline{\beta}\right]$. Let $\Gamma(A) = \ln(A)$, and let $\theta = 1$, $V = 0.2$, and $\underline{\beta} = 4$. We will consider two cases: $\bar{\beta} = 10$ (large preference heterogeneity) and $\bar{\beta} = 5$ (small preference heterogeneity) and look for the ideal that maximizes the expected equilibrium aggregate contribution to the public good:

$$\max_{a^*} \int_{[\underline{\beta},\bar{\beta}]} \int_{[\underline{\beta},\bar{\beta}]} \frac{A}{\left(\overline{\beta} - \underline{\beta}\right)^2} d\beta_1 d\beta_2.$$

In the following analysis, we will focus on the equilibrium with the largest insider set, whenever there are multiple equilibria. In the case that there are equilibria where the high-type agent is the only insider and where the low-type agent is the only insider, we will focus on the former equilibrium.

B.2.1: Large preference heterogeneity ($\overline{\beta} = 10$)

In the left graph of Fig. 6, the black solid curve plots the expected equilibrium aggregate contribution for each ideal level, while the (horizontal) red dashed curve plots the contribution under the baseline case without a^*. We see that the expected contribution is maximized at $a^* = 2.36$. Under this ideal, the right graph of Fig. 6 shows the participation profile of agents, in which the horizontal and vertical axes represent the types of agent 1 and agent 2 respectively. In this graph (and in the right graph of Fig. 7),

- symbol "×" denotes realization of types where both agents are outsiders in equilibrium;
- symbol "○" denotes realization of types where both agents are insiders in equilibrium;
- triangle "◁" denotes realization of types where agent 2 is the only insider in equilibrium;
- triangle "▽" denotes realization of types where agent 1 is the only insider in equilibrium; and
- the gray (black) color on the shapes (○, ◁ or ▽) denotes under-contributing (over-contributing) insider(s).

This graph shows that, under the contribution maximizing ideal, the agents with high enough types will choose to be insiders, while agents with low types will remain outsiders. This observation is consistent with the finding of our model in the main text that the optimal ideal attracts only the high-type agents when preference heterogeneity is large. On the other hand, however, the contribution-maximizing ideal here allows the possibility of overcontributors (those in black color). In our model with linear benefits (and thus no strategic interdependence), there are no overcontributing insiders under the contribution maximizing ideal.

B.2.2: Small preference heterogeneity ($\overline{\beta} = 5$)

The left graph of Fig. 7 plots the expected equilibrium aggregate contribution as the ideal level varies. We see that the contribution is maximized at $a^* = 1.95$. In the right graph which shows the participation profile of the agents under the contribution-maximizing ideal, all realizations are covered by the "○" symbol with gray color, meaning that all the agents are under-contributing insiders. Moreover, the left graph shows that the expected contribution drops sharply when a^* is slightly above the contribution-maximizing ideal, meaning that insiders with the lowest type drop out. Altogether these findings show that with small preference heterogenity, the contribution-maximizing ideal will make the agent with the lowest type indifferent between becoming an insider and not, and thus attract all agents as under-contributing insiders and encourage them to contribute more. These results are consistent with our model with linear benefits in the main text.

References

Aidt T (1998) Political internalization of economic externalities and environmental policy. J Public Econ 69:1–16

Aimone J, Iannaccone L, Makowsky M (2013) Endogenous group formation via unproductive costs. Rev Econ Stud 80(4):1215–1236

Akerlof R (2017) Value formation: the role of esteem. Games Econ Behav 102:1–19

Akerlof G, Kranton R (2000) Economics and identity. Q J Econ CXV(3):715–751

Akerlof G, Kranton R (2002) Identity and schooling: some lessons from the economics of education. J Econ Lite 40(4):1167–1201

Akerlof G, Kranton R (2005) Identity and the economics of organizations. J Econ Perspect 19(1):9–32

Akerlof G, Kranton R (2008) Identity, supervision, and work groups. Am Econ Rev Pap Proc 98(2):212–217

Akerlof G, Kranton R (2010) Identity economics: how our identities shape our work, wages and well-being. Princeton University Press, Princeton

Ali SN, Bénabou R (2016) Image versus information: changing societal norms and optimal privacy. NBER working paper 22203

Almudi I, Chóliz JS (2011) Sustainable use of renewable resources: an identity approach. J Bioecon 13:97–123

Andreoni J (1990) Impure altruism and donations to public goods: a theory of warm-glow giving. Econ J 100:464–477

Bagnoli M, Bergstrom T (2005) Log-concave probability and its applications. Econ Theor 26(2):445–469

Barbieri S, Malueg D (2008) Private provision of a discrete public good: continuous-strategy equilibria in the private-information subscription game. J Public Econ Theory 10(4):529–545

Barrett S (1994) Self-enforcing international environmental agreements. Oxf Econ Pap 46:878–894

Bénabou R, Tirole J (2011) Identity, dignity and taboos: beliefs as assets. Q J Econ 126:805–855

Benjamin D, Choi J, Fisher G (2016) Religious identity and economic behavior. Rev Econ Stat 98(4):617–637

Benjamin D, Choi J, Strickland GJ (2010) Social identity and preferences. Am Econ Rev 100(4):1913–28

Bernard M, Hett F, Mechtel M (2016) Social identity and social free-riding. Eur Econ Rev 90:4–17

Bernheim BD, Rangel A (2007) Behavioral public economics: welfare and policy analysis with non-standard decision-makers. In: Peter Diamond, Hannu Vartiainen (eds) Behavioral economics and its applications. Princeton University Press, Princeton, pp 7–77

Bisin A, Patacchini E, Verdier T, Zenou Y (2011) Formation and persistence of oppositional identities. Eur Econ Rev 55:1046–1071
Brekke K, Kverndokk S, Nyborg K (2003) An economic model of moral motivation. J Public Econ 87:1967–1983
Burlando R, Hey J (1997) Do Anglo-Saxons free-ride more? J Public Econ 64(1):41–60
Carvalho J-P (2016) Identity-based organizations. Am Econ Rev 106(5):410–414
Chen Y, Li SX (2009) Group identity and social preferences. Am Econ Rev 99(1):431–457
Conconi P (2003) Green lobbies and transboundary pollution in large open economies. J Int Econ 59:399–422
Costa D, Kahn M (2013) Energy conservation 'nudges' and environmentalist ideology: evidence from a randomized residential electricity field experiment. J Eur Eco Assoc 11(3):680–702
Costa-i-Font J, Cowell F (2014) Social identity and redistributive preferences: a survey. J Econ Surv 29(2):357–374
Croson R, Marks M, Snyder J (2003) Groups work for women: gender and group identity in the provision of public goods. Negot J 24(4):411–427
Diederich J, Goeschl T (2014) Willingness to pay for voluntary climate action and its determinants: field experimental evidence. Environ Resour Econ 57(3):405–429
Fehr E, Schmidt K (1999) A theory of fairness, competition and cooperation. Q J Econ 114(3):817–868
Fehr E, Gächter S (2002) Altruistic punishment in humans. Science 415:137–140
Goeschl T, Perino G (2017) The climate policy hold-up: green technologies, intellectual property rights, and the abatement incentives of international agreements. Scand J Econ 119(3):709–732
Goldstein N, Cialdini R, Griskevicius V (2008) A room with a viewpoint: using social norms to motivate environmental conservation in hotels. J Consum Res 18(5):429–434
Hardin R (2005) From communal norms to networks. Chin Public Aff Q 1(2):93–102
Hiller V, Verdier T (2014) Corporate culture and identity investment in an industry equilibrium. J Econ Behav Organ 103:93–102
Holländer H (1990) A social exchange approach to voluntary cooperation. Am Econ Rev 80(5):1157–1167
Hsiaw A (2013) Goal-setting and self-control. J Econ Theory 148:601–626
Iannaccone LR (1992) Sacrifice and stigma: reducing free-riding in cults, communes, and other collectives. J Polit Econ 100(2):271–291
Ito K, Ida T, Tanaka M (2018) Moral suasion and economic incentives: field experimental evidence from energy demand. Am Econ J Econ Policy 10(1):240–267
Kahn M (2007) Do greens drive hummers or hybrids? Environmental ideology as a determinant of consumer choice. J Environ Econ Manag 54(2):129–145
Kosfeld M, von Siemens FA (2011) Competition, cooperation, and corporate culture. Rand J Econ 42(1):23–43
Kotchen M, Moore M (2008) Conservation: from voluntary restraint to a voluntary pice premium. Environ Resour Econ 40(2):195–215
Kotchen M (2009) Voluntary provision of public goods for bads: a theory of environmental offsets. Econ J 119:883–899
Knack S, Keefer P (1997) Does social capital have an economic payoff? A cross-country investigation. Q J Econ 112(4):1251–1288
Laffont J-J, Martimort D (2001) The theory of incentives: the principal-agent model. Princeton University Press, Princeton
List J, Sturm D (2006) How elections matter: theory and evidence from environmental policy. Q J Econ 121(4):1249–1281
Monin B, Sawyer P, Marquez M (2008) The rejection of moral rebels: resenting those who do the right thing. J Pers Soc Psychol 95(1):76–93
Pareto V (1980) Compendium of general sociology, abridged in Italian by Giulio Farina, 1920, from Trattato di Sociologia Generale, Elisabeth Abbott, trans. University of Minnesota Press, Minneapolis
Rege M (2004) Social norms and private provision of public goods. J Public Econ Theory 6(1):65–77
Schultz P, Nolan J, Cialdini R, Goldstein N, Griskevicius V (2007) The constructive, destructive, and reconstructive power of social norms. Psychol Sci 18:429–34

Shayo M (2009) A model of social identity with an application to political economy: nation, class, and redistribution. Am Polit Sci Rev 103:147–174

Solow J, Kirkwood N (2002) Group identity and gender in public goods experiments. J Econ Behav Organ 48:403–412

Tabarrok A (1998) The private provision of public goods via dominant assurance contracts. Public Choice 96:345–362

Thaler R, Sunstein C (2008) Nudge: improving decisions about health, wealth and happiness. Yale University Press, London