

Daske, Thomas; March, Christoph

Working Paper

Efficient Incentives with Social Preferences

Suggested Citation: Daske, Thomas; March, Christoph (2022) : Efficient Incentives with Social Preferences, ZBW - Leibniz Information Centre for Economics, Kiel, Hamburg

This Version is available at:

<https://hdl.handle.net/10419/254263>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Efficient Incentives with Social Preferences*

Thomas Daske[‡]

Christoph March[§]

May 13, 2022

This study explores mechanism design with allocation-based social preferences. Agents' social preferences and private payoffs are all subject to asymmetric information. We assume quasi-linear utility and independent types. We show that the asymmetry of information about agents' social preferences can be operationalized to satisfy agents' participation constraints. Our main result is a possibility result for groups of *at least three* agents: If endowments are sufficiently large, any such group can resolve any given allocation problem with an ex-post budget-balanced mechanism that is Bayesian incentive-compatible, interim individually rational, and ex-post Pareto-efficient.

JEL Classification: C72; C78; D62; D82

Keywords: Mechanism design; social preferences; Bayesian implementation; participation constraints

*An earlier version circulated under the title “Efficient incentives in social networks: Gamification and the Coase theorem” and is available under <http://hdl.handle.net/10419/222527>.

For their helpful comments and critical remarks, we thank Benny Moldovanu, Marco Sahn, Klaus Schmidt, Johannes Schneider, Roland Strausz, and Robert von Weizsäcker as well as participants of the European Winter Meeting of the Econometric Society in Milan, the World Congress of the Game Theory Society in Maastricht, the Annual Meeting of the Association for Public Economic Theory in Strasbourg, the Annual Congress of the International Institute of Public Finance in Glasgow, the European Meeting of the Econometric Society in Manchester, the Annual Congress of the German Economic Association in Leipzig, and the virtual Econometric Society World Congress.

[‡]Technical University of Munich, TUM School of Management, Arcisstr. 21, 80333 Munich, Germany. Email: thomas.daske@tum.de

[§]University of Bamberg, Department of Economics, Feldkirchenstr. 21, 96047 Bamberg, Germany; and CESifo, Poschingerstr. 5, 81679 Munich, Germany. Email: christoph.march@uni-bamberg.de

1 Introduction

An incentive mechanism should have four properties: Incentive compatibility, ex-post Pareto efficiency, ex-post budget balance, and interim individual rationality.

Bayesian implementation has become an accepted way to achieve the first three properties.¹ Often, however, Bayesian mechanisms violate agents' participation constraints.²

Bayesian mechanisms that reconcile all four properties exist if agents' private signals are sufficiently *correlated*: [Crémer and McLean \(1985, 1988\)](#) show that the designer may exploit this correlation to validate the agents' reports, extract all information rents, and ensure participation en passant.³ [Mezzetti \(2004\)](#) shows that the logic of [Crémer and McLean \(1985, 1988\)](#) can be extended to the case of independent private signals if the designer is permitted to implement a two-stage mechanism: The designer can resolve the allocation problem and ensure participation by *sequentially* administering a social alternative and transfers.

The present paper extends the class of possibility results. Contrary to [Crémer and McLean \(1985, 1988\)](#) and [Mezzetti \(2004\)](#), we neither assume that private signals are correlated, nor that reporting is sequential. Instead, we explore mechanism design for agents with outcome-based social preferences that are privately known (next to privately known payoff types). That is, agents care about the *distributive effects* of a mechanism, and their distributive preferences are private information. We show how this kind of information asymmetry can be operationalized to satisfy agents' participation constraints.

Our main result, Theorem 1, states that any group of at least three agents can resolve any given allocation problem with an ex-post budget-balanced mechanism that is Bayesian incentive-compatible, interim individually rational, and ex-post Pareto-efficient, provided endowments are sufficiently large.

Until recently, the literature on efficient design has either neglected social preferences or assumed them to be common knowledge.⁴ An exception is [Bierbrauer and Netzer \(2016\)](#), who study mechanism design when agents have intention-based social preferences that are private information. They show that social preferences enhance the opportunities for efficient, individually rational design if and only if none but conditionally pro-social

¹E.g., [Arrow \(1979\)](#), [d'Aspremont and Gérard-Varet \(1979\)](#).

²For settings with independent private signals see, e.g., [Myerson and Satterthwaite \(1983\)](#), [Mailath and Postlewaite \(1990\)](#), [Williams \(1999\)](#), and [Segal and Whinston \(2016\)](#).

³Likewise, [McAfee and Reny \(1992\)](#), [McLean and Postlewaite \(2004\)](#), [Kosenok and Severinov \(2008\)](#).

⁴E.g., [Desiraju and Sappington \(2007\)](#), [Kucuksenel \(2012\)](#), [Tang and Sandholm \(2012\)](#).

types are present. Our result differs in that we consider different kinds of social preferences, allowing for anti-social preferences such as spite, and identify a more subtle rationale for participation, building on information asymmetry rather than convenient social-type combinations.

Our possibility result builds on the following insights: In quasi-linear environments, a mechanism can be designed such that the incentives to reveal payoff types and social types are separated. While the allocation problem can be resolved through payoff-type conditional transfers, agents' participation can be attracted through budget-balanced transfers that condition on social types; the latter is possible whenever there are at least three agents.

The paper proceeds as follows. Section 2 outlines the model framework. Section 3 states and interprets our main result. Section 4 details the proof. Section 5 illustrates the economic intuition behind our participation-stimulating transfers. Section 6 concludes.

2 The Model

2.1 The Allocation Problem

There is a group $\mathcal{I} = \{1, \dots, n\}$ of $n \geq 2$ agents and there is a finite set K of social alternatives. From alternative $k \in K$ and a transfer $t_i \in \mathbb{R}$, agent i gains a *private payoff* $\Pi_i(k, t_i | \theta_i) = \pi_i(k | \theta_i) + t_i$, with $\pi_i : K \times \Theta_i \rightarrow \mathbb{R}$. Agent i 's *payoff type* θ_i belongs to a finite set Θ_i , with $|\Theta_i| \geq 2$. The collection of agents' payoff types is denoted by $\theta = (\theta_i, \theta_{-i})$, where $\theta_{-i} = (\theta_j)_{j \neq i}$. Agents exhibit social preferences in the form of altruism or spite:⁵ From the allocation of private payoffs, agent i derives ex-post *utility*

$$u_i(k, (t_j)_{j \in \mathcal{I}}, \theta_{-i} | \theta_i, \delta_i) = \sum_{j \in \mathcal{I}} \delta_{ij} \Pi_j(k, t_j | \theta_j),$$

where the value δ_{ij} that i assigns to j 's payoff, $j \neq i$, belongs to a closed (proper) interval $\Delta_{ij} = [\delta_{ij}^{\min}, \delta_{ij}^{\max}] \subset (\frac{-1}{n-1}, 1)$. For convenience, $\delta_{ii} = 1$ for all i . Notice that $(\frac{-1}{n-1}, 1)$ is the maximum range of interpersonal altruism, or spite, for which agents care about overall

⁵For evidence on altruism, see Andreoni and Miller (2002), Charness and Rabin (2002), and Bruhin, Fehr, and Schunk (2019). For evidence on spite, see Saijo and Nakamura (1995), Fehr, Hoff, and Kshetramade (2008), and Prediger, Vollan, and Herrmann (2014).

material efficiency while still being selfish to the extent that every one of them prefers a dollar to be her own rather than having that same dollar distributed among the others.

We refer to δ_{ij} as *i*'s degree of altruism towards *j*, to the collection $\delta_i = (\delta_{ij})_{j \neq i}$ as *i*'s social type, and to the pair (θ_i, δ_i) as *i*'s type. The collection of social types is denoted by $\delta = (\delta_i, \delta_{-i})$, with $\delta_{-i} = (\delta_j)_{j \neq i}$, and Cartesian products of type sets by $\Theta = \prod_i \Theta_i$, $\Delta_i = \prod_{j \neq i} \Delta_{ij}$, and $\Delta = \prod_i \Delta_i$.

Our key assumption is that each agent is privately informed about her payoff type and social type. Hence, in any bilateral relationship there remains, to some extent, uncertainty about who (dis-)likes whom how much.⁶ While each agent *i*'s payoff type and social type realize independently according to strictly positive densities, the various degrees of altruism determining *i*'s social type may correlate. At the interpersonal level, *agents' types are independent*.

We further assume that agents do not have access to an outside source of money, such that transfers must be *weakly budget-balanced*: $\sum_{i \in \mathcal{I}} t_i \leq 0$.

The agents' problem is to choose a social alternative *k* and transfers $(t_i)_{i \in \mathcal{I}}$ such that the resulting allocation is ex-post Pareto-efficient.⁷

2.2 Revelation Mechanisms

A *direct* revelation mechanism involves the agents in a strategic game of incomplete information in which they are asked to report their types truthfully. Types are reported *simultaneously*. Based on their reports, a social alternative is chosen and transfers are made. As the *revelation principle* applies to the present setup (Myerson, 1979), there is no loss of generality in considering only direct mechanisms.

Formally, a mechanism is given by a pair $\langle k, T \rangle$ with allocation function $k : \Theta \times \Delta \rightarrow K$ and transfer scheme $T = (t_i)_{i \in \mathcal{I}} : \Theta \times \Delta \rightarrow \mathbb{R}^n$. Denote by $U_i(\hat{\theta}_i, \hat{\delta}_i | \theta_i, \delta_i)$ agent *i*'s interim-expected utility from reporting $(\hat{\theta}_i, \hat{\delta}_i)$ if her true type is (θ_i, δ_i) while all the other agents report their types truthfully: $U_i(\hat{\theta}_i, \hat{\delta}_i | \theta_i, \delta_i) = \sum_{j \in \mathcal{I}} \delta_{ij} [\bar{\pi}_{ij}(\hat{\theta}_i, \hat{\delta}_i) + \bar{t}_{ij}(\hat{\theta}_i, \hat{\delta}_i)]$, where $\bar{\pi}_{ij}(\theta_i, \delta_i) = \mathbb{E}_{\theta_{-i}, \delta_{-i}} [\pi_j(k(\theta, \delta) | \theta_j)]$ and $\bar{t}_{ij}(\theta_i, \delta_i) = \mathbb{E}_{\theta_{-i}, \delta_{-i}} [t_j(\theta, \delta)]$. For convenience,

⁶Despite the asymmetry of information, it can still be common knowledge who is 'friends' and who is 'foes.' For instance, if $\delta_{k\ell}^{\max}, \delta_{\ell k}^{\max} < 0 < \delta_{ij}^{\min}, \delta_{ji}^{\min}$, then, in comparison, *i* and *j* are friends, whereas *k* and *l* are foes. Similarly, it can be common knowledge that *i* likes *j* more than *k*, which is the case if $\delta_{ik}^{\max} < \delta_{ij}^{\min}$. While we assume that the variance of every δ_{ij} is strictly positive, it is also allowed to be arbitrarily small. Reciprocal social preferences can be captured by letting $\Delta_{ij} = \Delta_{ji}$ and $\delta_{ij}^{\min} \approx \delta_{ij}^{\max}$.

⁷Formally, our model is one of *one-dimensional* allocative and informational externalities. Jehiel and Moldovanu (2001) prove the generic impossibility of efficient design if externalities are *multi-dimensional*.

$U_i(\theta_i, \delta_i) = U_i(\theta_i, \delta_i | \theta_i, \delta_i)$. The mechanism $\langle k, T \rangle$ is *Bayesian incentive-compatible* if, for all $i \in \mathcal{I}$ and all $(\theta_i, \delta_i) \in \Theta_i \times \Delta_i$, we have $U_i(\theta_i, \delta_i) = \max_{(\hat{\theta}_i, \hat{\delta}_i) \in \Theta_i \times \Delta_i} U_i(\hat{\theta}_i, \hat{\delta}_i | \theta_i, \delta_i)$.⁸

2.3 Efficiency and Participation

The following Lemma links material efficiency (the maximum surplus of private payoffs) to Pareto efficiency. It allows us to focus on allocations that are *ex-post materially efficient*: $k^*(\theta) \in \arg \max_{k \in K} \sum_{i \in \mathcal{I}} \pi_i(k | \theta_i)$ while transfers $(t_i)_{i \in \mathcal{I}}$ are (strictly, or *ex-post*) budget-balanced, $\sum_{i \in \mathcal{I}} t_i = 0$.

Lemma 1 *An allocation is ex-post Pareto-efficient only if transfers are ex-post budget-balanced. If $|\delta_{ij}| < \frac{1}{2n-3}$ for all i and all $j \neq i$, then an ex-post materially efficient allocation is also ex-post Pareto-efficient; moreover, no ex-post budget-balanced transfer scheme Pareto-dominates another.*

Proof. See Appendix A. ■

The intuition behind Lemma 1 is the following: If agents switch from a social alternative that is materially efficient to one that is not, or from one budget-balanced transfer scheme to another, then at least one agent must incur a material loss. Consider the agent whose material loss is largest. If this agent i is sufficiently selfish, $|\delta_{ij}| < \frac{1}{2n-3}$ for all $j \neq i$, then she would also incur a loss utility-wise.

Notice that the Pareto frontier can be indefinite for combinations of social types satisfying $|\delta_{ij}| \geq \frac{1}{2n-3}$, in which case a subgroup of agents might be willing to transfer arbitrary amounts of money to their joint favorite agent.⁹

Finally, the mechanism $\langle k, T \rangle$ is *interim individually rational* if it gains all agents' approval at the interim stage. Following Segal and Whinston (2016), we represent agents' reservation utilities by the interim-expected utilities that agents' derive from a Bayesian mechanism $\langle k^\circ, T^\circ \rangle$, with $k^\circ : \Theta \times \Delta \rightarrow K$ representing "property rights" and $T^\circ = (t_i^\circ)_{i \in \mathcal{I}} : \Theta \times \Delta \rightarrow \mathbb{R}^n$ representing "liability rules."

⁸Bayesian implementation has been criticized for assuming that the distribution of agents' types is common knowledge. Bergemann and Morris (2005) have proposed *ex-post implementation* for environments with interdependent utilities, requiring that truthful revelation of types constitutes a Nash equilibrium. Jehiel et al. (2006) show that ex-post implementation is 'generically' not feasible in the presence of informational externalities; a finding extended by Zik (2021) to our present context.

⁹An example is the group of three agents with $\delta_{13} = \delta_{23} > 1/3$, $\delta_{12} = \delta_{21} = -1/3$, and $\delta_{31} = \delta_{32} = 0$, in which agents 1 and 2 are willing to *jointly* transfer arbitrary individual amounts of $t > 0$ to agent 3.

3 A Possibility Result

We establish our main result with the help of two concepts, *preference-separating mechanisms* and *participation-stimulating transfers*:

Definition 1 (Preference Separation and Participation Stimulation)

A preference-separating mechanism $\langle k^*, T^* \rangle$ consists of the ex-post materially efficient allocation function $k^* : \Theta \rightarrow K$, with $k^*(\theta) \in \arg \max_{k \in K} \sum_{i \in \mathcal{I}} \pi_i(k | \theta_i)$, and an ex-post budget-balanced transfer scheme $T^* = (t_i^*)_{i \in \mathcal{I}} : \Theta \times \Delta \rightarrow \mathbb{R}^n$ defined by

$$t_i^*(\hat{\theta}, \hat{\delta}) = \underbrace{\sum_{j \neq i} \left[\mathbb{E}_{\theta_{-i}} [\pi_j(k^*(\hat{\theta}_i, \theta_{-i}) | \theta_j)] - \mathbb{E}_{\theta_{-j}} [\pi_i(k^*(\hat{\theta}_j, \theta_{-j}) | \theta_i)] \right]}_{\text{the terms of trade}} + \underbrace{s_i^*(\hat{\delta})}_{\text{participation-stimulating transfers}},$$

where participation-stimulating (PS) transfers $s^* = (s_i^*)_{i \in \mathcal{I}} : \Delta \rightarrow \mathbb{R}^n$ are defined by jointly satisfying the following conditions:

(i) s^* is strategy-proof: For all $i \in \mathcal{I}$, all $\delta \in \Delta$, and all $\hat{\delta}_i \in \Delta_i$,

$$\sum_{j \in \mathcal{I}} \delta_{ij} s_j^*(\delta) \geq \sum_{j \in \mathcal{I}} \delta_{ij} s_j^*(\hat{\delta}_i, \delta_{-i}).$$

(ii) s^* is ex-post budget-balanced: For all $\delta \in \Delta$,

$$\sum_{j \in \mathcal{I}} s_j^*(\delta) = 0.$$

(iii) From unanimous participation in s^* , each agent derives strictly positive interim-expected utility: For all $i \in \mathcal{I}$ and all $\delta_i \in \Delta_i$,

$$\sum_{j \in \mathcal{I}} \delta_{ij} \mathbb{E}_{\delta_{-i}} [s_j^*(\delta)] > 0.$$

Theorem 1 (Efficient Implementation With At Least Three Agents)

If $n \geq 3$ and endowments are sufficiently large, then there exists a preference-separating mechanism $\langle k^*, T^* \rangle$ that is Bayesian incentive-compatible, interim individually rational, ex-post budget-balanced, and ex-post materially efficient. If $|\delta_{ij}| < \frac{1}{2n-3}$ for all i and all $j \neq i$, then $\langle k^*, T^* \rangle$ is necessarily ex-post Pareto-efficient.

Before we prove Theorem 1, we shall discuss the inner logic of our mechanism.

Consider first the *terms of trade*. Those operate on agents' payoff types and, as we will see, are *social-preference robust* in that they leave agents' social preferences strategically irrelevant. This is achieved by applying the mutual-concessions principle of the dyadical AGV-mechanism (Arrow, 1979, d'Aspremont and Gérard-Varet, 1979) to each and every single dyad: For materially efficient social alternatives $k^*(\theta)$, every agent i pays to every other j the monetary equivalent of what j believes to contribute to i 's material well-being when reporting her payoff type $\hat{\theta}_j$; that is, i transfers $\mathbb{E}_{\theta_{-j}}[\pi_i(k^*(\hat{\theta}_j, \theta_{-j}) | \theta_i)]$ to j and receives $\mathbb{E}_{\theta_{-i}}[\pi_j(k^*(\hat{\theta}_i, \theta_{-i}) | \theta_j)]$ from j .

For *two* other-regarding agents, Bierbrauer and Netzer (2016) show that the AGV-mechanism is *social-preference robust*. Agents are incentivized to behave *as if* they are selfish: If $-i$ reports her payoff type truthfully, then $\mathbb{E}_{\theta_{-i}}[\pi_{-i}(k^*(\hat{\theta}_i, \theta_{-i}) | \theta_{-i}) + t_{-i}^*(\hat{\theta}_i, \theta_{-i})] = \mathbb{E}_{\theta}[\pi_i(k^*(\theta) | \theta_i)]$; thereby, i 's degree of altruism is rendered strategically irrelevant.¹⁰ We establish social-preference robustness for groups of arbitrary size. Consequently, the *terms of trade* preserve agents' privately known social preferences as a strategic degree of freedom.

This strategic degree of freedom is utilized by *participation-stimulating transfers*. Those are independent of the actual allocation problem and serve the purpose of attracting (or, stimulating) agents' participation in the terms of trade. While being *ex-post budget balanced*, PS transfers yield agents an *interim-expected Pareto improvement upon the terms of trade*. If this interim-expected Pareto improvement is amplified sufficiently, by uniformly scaling up the PS transfers, then agents' interim-expected utilities from unanimous participation will outweigh their interim reservation utilities.

Finally, we note that our preference-separation approach cannot succeed in dyads:

Proposition 1 *Participation-stimulating transfer schemes do not exist if $n = 2$.*

Proof. Suppose the opposite is true. Then Definition 1(iii) requires that $0 < \mathbb{E}_{\delta_{-i}}[s_i^*(\delta)] + \delta_i \mathbb{E}_{\delta_{-i}}[s_{-i}^*(\delta)]$ for both $i \in \{1, 2\}$ and all $\delta_i \in (-1, 1)$, while $s_{-i}^*(\delta) = -s_i^*(\delta)$ due to ex-post budget balance. Hence, $0 < (1 - \delta_i) \mathbb{E}_{\delta_{-i}}[s_i^*(\delta)]$, implying that $0 < \mathbb{E}_{\delta_{-i}}[s_i^*(\delta)]$ for all i, δ_i . But then, $0 < \mathbb{E}_{\delta}[s_i^*(\delta)]$ for both i , contradicting ex-post budget balance. ■

¹⁰Bierbrauer and Netzer (2016) coin this property the 'insurance property,' as it insures agents against the other-regarding concerns of one another.

4 Proof of Theorem 1

The proof of Theorem 1 proceeds in a series of Lemmas. Throughout, $n \geq 3$.

Lemma 2 Preference-separating mechanisms are *Bayesian incentive-compatible and ex-post materially efficient*. If $|\delta_{ij}| < \frac{1}{2n-3}$ for all i and all $j \neq i$, they are also *ex-post Pareto-efficient*.

Proof. *Efficiency:* Preference-separating mechanisms are ex-post materially efficient by construction; hence, by Lemma 1, they are also ex-post Pareto-efficient if $|\delta_{ij}| < \frac{1}{2n-3}$ for all i and all $j \neq i$.

Incentive compatibility: Suppose the agents other than i reveal their types truthfully. Then the transfers that i interim-expects for herself and every other j are given by:

$$\begin{aligned} \bar{t}_{ii}(\hat{\theta}_i, \hat{\delta}_i) &= \sum_{\ell \neq i} \mathbb{E}_{\theta_{-i}} [\pi_\ell(k^*(\hat{\theta}_i, \theta_{-i}) | \theta_\ell)] - (n-1) \mathbb{E}_\theta [\pi_i(k^*(\theta) | \theta_i)] + \mathbb{E}_{\delta_{-i}} [s_i^*(\hat{\delta}_i, \delta_{-i})], \\ \bar{t}_{ij}(\hat{\theta}_i, \hat{\delta}_i) &\stackrel{j \neq i}{=} \sum_{\ell \neq j} \mathbb{E}_{\theta_{-i}, \theta_{-j}} [\pi_\ell(k^*(\theta) | \theta_\ell)] - \sum_{\ell \neq i, j} \mathbb{E}_{\theta_{-i}, \theta_{-\ell}} [\pi_j(k^*(\theta) | \theta_j)] \\ &\quad - \mathbb{E}_{\theta_{-i}} [\pi_j(k^*(\hat{\theta}_i, \theta_{-i}) | \theta_j)] + \mathbb{E}_{\delta_{-i}} [s_j^*(\hat{\delta}_i, \delta_{-i})] \\ &= \sum_{\ell \in \mathcal{I}} \mathbb{E}_\theta [\pi_\ell(k^*(\theta) | \theta_\ell)] - (n-1) \mathbb{E}_\theta [\pi_j(k^*(\theta) | \theta_j)] \\ &\quad - \mathbb{E}_{\theta_{-i}} [\pi_j(k^*(\hat{\theta}_i, \theta_{-i}) | \theta_j)] + \mathbb{E}_{\delta_{-i}} [s_j^*(\hat{\delta}_i, \delta_{-i})]. \end{aligned}$$

Agent i 's interim-expected utility from reporting $(\hat{\theta}_i, \hat{\delta}_i)$ thus satisfies

$$\begin{aligned} (1) \quad U_i(\hat{\theta}_i, \hat{\delta}_i | \theta_i, \delta_i) &= \sum_{j \in \mathcal{I}} \delta_{ij} \left[\mathbb{E}_{\theta_{-i}} [\pi_j(k^*(\hat{\theta}_i, \theta_{-i}) | \theta_j)] + \bar{t}_{ij}(\hat{\theta}_i, \hat{\delta}_i) \right] \\ &= \mathbb{E}_{\theta_{-i}} \left[\sum_{\ell \in \mathcal{I}} \pi_\ell(k^*(\hat{\theta}_i, \theta_{-i}) | \theta_\ell) \right] + \left(\sum_{j \neq i} \delta_{ij} \right) \mathbb{E}_\theta \left[\sum_{\ell \in \mathcal{I}} \pi_\ell(k^*(\theta) | \theta_\ell) \right] \\ &\quad - (n-1) \mathbb{E}_\theta \left[\sum_{j \in \mathcal{I}} \delta_{ij} \pi_j(k^*(\theta) | \theta_j) \right] + \sum_{j \in \mathcal{I}} \delta_{ij} \mathbb{E}_{\delta_{-i}} [s_j^*(\hat{\delta}_i, \delta_{-i})]. \end{aligned}$$

By equation (1), the incentives to reveal payoff types and social types are additively separated. As participation-stimulating transfers s^* are strategy-proof by Definition 1(i), preference-separating mechanisms are (dominant-strategy) incentive-compatible with respect to social types. On the other hand, if truthful revelation of her payoff type θ_i was inferior for some agent i , then there would exist $\hat{\theta}_i$ and θ_{-i} such that $\sum_{\ell \in \mathcal{I}} \pi_\ell(k^*(\hat{\theta}_i, \theta_{-i}) | \theta_\ell) >$

$\sum_{\ell \in \mathcal{I}} \pi_\ell(k^*(\theta_i, \theta_{-i}) | \theta_\ell)$, implying that $\sum_{\ell \in \mathcal{I}} \pi_\ell(k | \theta_\ell) > \sum_{\ell \in \mathcal{I}} \pi_\ell(k^*(\theta) | \theta_\ell)$ for some social alternative k , in contradiction to the definition of k^* . ■

By equation (1), the terms of trade are social-preference robust: Agents' social preferences are rendered strategically irrelevant when it comes to implementing the materially efficient allocation function k^* . This opens up the possibility to operationalize the asymmetry of information about agents' social preferences to satisfy their interim participation constraints.

We construct participation-stimulating transfer schemes as follows. Let $M \in \mathcal{I}$ denote one (arbitrarily chosen) agent and define transfers $s^* = (s_i^*)_{i \in \mathcal{I}} : \Delta \rightarrow \mathbb{R}^n$ by

$$(2) \quad s_M^*(\delta) = - \sum_{j \neq M} s_j^*(\delta),$$

$$(3) \quad s_i^*(\delta) = -C + g_i(\delta_i^*) - \delta_i^* g_i'(\delta_i^*) + \sum_{\ell \neq i, M} g_\ell'(\delta_\ell^*), \quad \text{for } i \neq M,$$

$$(4) \quad g_i(\delta_i^*) = \text{Var}_{\delta_i}[\delta_i^*] + (\delta_i^* - \mathbb{E}_{\delta_i}[\delta_i^*])^2,$$

$$(5) \quad \delta_i^* = \frac{\sum_{\ell \neq i, M} (\delta_{i\ell} - \delta_{iM})}{\delta_{ii} - \delta_{iM}},$$

for some constant $C > 0$.

In order to establish that this transfer scheme is participation-stimulating, we first consider the functions $(g_i)_i$:

Lemma 3 *Be $X_i : \Delta_i \rightarrow \mathbb{R}$ a continuous non-constant random variable. Then $\mathbb{E}_{\delta_i} [X_i]$ and $\text{Var}_{\delta_i} [X_i]$ exist, and $g_i : \mathbb{R} \rightarrow \mathbb{R}$ defined by $g_i(X_i) = \text{Var}_{\delta_i} [X_i] + (X_i - \mathbb{E}_{\delta_i} [X_i])^2$ satisfies $g_i(X_i) > 0$, $g_i''(X_i) > 0$, and $\mathbb{E}_{\delta_i} [g_i'(X_i)] = 0 = \mathbb{E}_{\delta_i} [g_i(X_i) - X_i g_i'(X_i)]$.*

Proof. $\mathbb{E}_{\delta_i} [X_i]$ and $\text{Var}_{\delta_i} [X_i]$ exist, since Δ_i is compact and convex while X_i and the density of δ_i are continuous. Obviously, $g_i(X_i) > 0$, $g_i''(X_i) > 0$, and $\mathbb{E}_{\delta_i} [g_i'(X_i)] = 2 \mathbb{E}_{\delta_i} [X_i - \mathbb{E}_{\delta_i} [X_i]] = 0$. On the other hand, as $\text{Var}_{\delta_i} [X_i] = \mathbb{E}_{\delta_i} [X_i^2] - \mathbb{E}_{\delta_i} [X_i]^2$, one has $g_i(X_i) - X_i g_i'(X_i) = \mathbb{E}_{\delta_i} [X_i^2] - \mathbb{E}_{\delta_i} [X_i]^2 + X_i^2 - 2X_i \mathbb{E}_{\delta_i} [X_i] + \mathbb{E}_{\delta_i} [X_i]^2 - 2X_i (X_i - \mathbb{E}_{\delta_i} [X_i]) = \mathbb{E}_{\delta_i} [X_i^2] - X_i^2$; hence, $\mathbb{E}_{\delta_i} [g_i(X_i) - X_i g_i'(X_i)] = 0$. ■

We obtain that participation-stimulating transfer schemes exist if $n \geq 3$:

Lemma 4 *The transfer scheme s^* defined by (2)–(5) is participation-stimulating in the manner of Definition 1 if $C > 0$ is chosen sufficiently small.*

Proof. *Strategy proofness:* Under s^* , each agent $j \neq M$ reports a social type $\hat{\delta}_j$, which is strategically equivalent to reporting some signal $\hat{\delta}_j^* \in \mathbb{R}$. Her ex-post utility is given by

$$\begin{aligned} \sum_{\ell \neq M} (\delta_{j\ell} - \delta_{jM}) s_\ell^*(\hat{\delta}) &= (\delta_{jj} - \delta_{jM}) \left[g_j(\hat{\delta}_j^*) - \hat{\delta}_j^* g'_j(\hat{\delta}_j^*) + \sum_{\ell \neq j, M} g'_\ell(\hat{\delta}_\ell^*) \right] \\ &+ \sum_{\ell \neq j, M} (\delta_{j\ell} - \delta_{jM}) \left[g_\ell(\hat{\delta}_\ell^*) - \hat{\delta}_\ell^* g'_\ell(\hat{\delta}_\ell^*) + \sum_{\ell' \neq \ell, j, M} g'_{\ell'}(\hat{\delta}_{\ell'}^*) \right] \\ &+ \left[\sum_{\ell \neq j, M} (\delta_{j\ell} - \delta_{jM}) \right] g'_j(\hat{\delta}_j^*) - C \sum_{\ell \neq M} (\delta_{j\ell} - \delta_{jM}). \end{aligned}$$

Hence, when substituting for $\delta_j^* = \sum_{\ell \neq j, M} (\delta_{j\ell} - \delta_{jM}) / (\delta_{jj} - \delta_{jM})$, agent j maximizes $g_j(\hat{\delta}_j^*) + (\delta_j^* - \hat{\delta}_j^*) g'_j(\hat{\delta}_j^*)$ over the choice of $\hat{\delta}_j^*$. As $g''_j > 0$, each $j \neq M$ has the strictly dominant strategy to report $\hat{\delta}_j^* = \delta_j^*$. As agent M is not involved strategically, she has the weakly dominant strategy to report her true social type δ_M .

Ex-post budget balance: Immediate from equation (2).

Interim-expected Pareto improvement: When substituting for δ_j^* and $\mathbb{E}_{\delta_\ell} [g'_\ell(\delta_\ell^*)] = 0 = \mathbb{E}_{\delta_\ell} [g_\ell(\delta_\ell^*) - \delta_\ell^* g'_\ell(\delta_\ell^*)]$, due to Lemma 3, then j 's interim-expected utility from s^* is

$$\begin{aligned} \sum_{\ell \neq M} (\delta_{j\ell} - \delta_{jM}) \mathbb{E}_{\delta_{-j}} [s_\ell^*(\delta)] &= (\delta_{jj} - \delta_{jM}) g_j(\delta_j^*) - C \sum_{\ell \neq M} (\delta_{j\ell} - \delta_{jM}) \\ &= (\delta_{jj} - \delta_{jM}) g_j(\delta_j^*) - C(\delta_{jj} - \delta_{jM}) - C \sum_{\ell \neq j, M} (\delta_{j\ell} - \delta_{jM}) \\ &= (\delta_{jj} - \delta_{jM}) [g_j(\delta_j^*) - C(1 + \delta_j^*)]. \end{aligned}$$

Recall that $\delta_{jj} = 1 > \delta_{jM}$ and $g_j(\delta_j^*) \geq \text{Var}_{\delta_j}[\delta_j^*] > 0$. Notice that $\delta_j^* < n - 2$, since $\delta_{jj} - \delta_{jM} > \delta_{j\ell} - \delta_{jM}$ for all $\ell \neq j, M$. Hence, each agent $j \neq M$ derives positive interim-expected utility from unanimous participation if $C \leq \min_{j \neq M} \text{Var}_{\delta_j}[\delta_j^*] / (n - 1)$. Due to Lemma 3 again, also M 's interim-expected utility is positive if all agents participate: $\sum_{i \in \mathcal{I}} \delta_{Mi} \mathbb{E}_{\delta_{-M}} [s_i^*(\delta)] = \sum_{j \neq M} (\delta_{Mj} - 1) \mathbb{E}_\delta [s_j^*(\delta)] = C \sum_{j \neq M} (1 - \delta_{Mj}) > 0$. ■

Several remarks on the PS scheme (2)–(5) are in order. First, s^* is independent of agent M 's social type, $(\delta_{Mj})_{j \neq M}$, such that M has no strategic role to play under s^* . This feature is not a prerequisite for preference-separating implementation. Furthermore, each agent $i \neq M$ has the strictly dominant strategy to report $\delta_i^* = \sum_{\ell \neq i, M} (\delta_{i\ell} - \delta_{iM}) / (\delta_{ii} - \delta_{iM})$ which is thus a one-dimensional sufficient statistic for i 's social type. This fact allows for

implementing the PS scheme by having players reveal the necessary information about their social types via the choice of one-dimensional strategic variables, such as efforts.¹¹

The PS scheme implicitly assumes that the mean and variance of every δ_j^* are commonly known. This assumption is sufficient but not necessary. As s^* is strategy-proof while the resulting interim-expected Pareto improvement is strict, it suffices that agents (and the designer) have sufficiently good estimates of those means and variances.

With Lemmas 1 to 4 at hand, we can establish Theorem 1:

Proof of Theorem 1. Consider the preference-separating mechanism $\langle k^*, T^* \rangle$ with

$$t_i^*(\hat{\theta}, \hat{\delta}) = \underbrace{\sum_{j \neq i} \left[\mathbb{E}_{\theta_{-i}} [\pi_j(k^*(\hat{\theta}_i, \theta_{-i}) | \theta_j)] - \mathbb{E}_{\theta_{-j}} [\pi_i(k^*(\hat{\theta}_j, \theta_{-j}) | \theta_i)] \right]}_{\text{the terms of trade}} + \underbrace{\alpha^* \cdot s_i^*(\hat{\delta})}_{\text{PS transfers}},$$

where $(s_i^*)_{i \in \mathcal{I}}$ is defined by equations (2) to (5) while $\alpha^* > 0$. Notice that the conditions of Definition 1 are invariant under scaling all the components s_i^* with the same factor.

By Lemmas 2 and 4, this mechanism is Bayesian incentive-compatible. The mechanism is ex-post budget-balanced and ex-post materially efficient by construction. By Lemma 1, it is ex-post Pareto-efficient if $|\delta_{ij}| < \frac{1}{2n-3}$ for all i and all $j \neq i$. By equation (1), and since $\langle k^*, T^* \rangle$ is Bayesian incentive-compatible, agent i 's interim-expected utility from unanimous participation in $\langle k^*, T^* \rangle$ is given by

$$\begin{aligned} U_i(\theta_i, \delta_i) &= \mathbb{E}_{\theta_{-i}} \left[\sum_{\ell \in \mathcal{I}} \pi_\ell(k^*(\theta) | \theta_\ell) \right] + \left(\sum_{j \neq i} \delta_{ij} \right) \mathbb{E}_\theta \left[\sum_{\ell \in \mathcal{I}} \pi_\ell(k^*(\theta) | \theta_\ell) \right] \\ &\quad - (n-1) \mathbb{E}_\theta \left[\sum_{j \in \mathcal{I}} \delta_{ij} \pi_j(k^*(\theta) | \theta_j) \right] + \alpha^* \cdot \sum_{j \in \mathcal{I}} \delta_{ij} \mathbb{E}_{\delta_{-i}} [s_j^*(\delta)], \end{aligned}$$

where $\sum_{j \in \mathcal{I}} \delta_{ij} \mathbb{E}_{\delta_{-i}} [s_j^*(\delta)] > 0$ due to Lemma 4. Hence, if α^* is chosen sufficiently large, agents' interim participation constraints are satisfied for any given collection of interim reservation utilities $(U_i^\circ(\theta_i, \delta_i))_{i \in \mathcal{I}}$, as specified in Section 2.3. The assumption of sufficiently large endowments guarantees that agents can afford the respective transfers $(t_i^*(\theta, \delta))_{i \in \mathcal{I}}$ whenever those are negative. ■

¹¹We will illustrate these aspects in a separate paper. For details see the earlier draft of the this paper, available under <http://hdl.handle.net/10419/222527>.

5 The Intuition Behind Participation Stimulation

We wish to outline the construction of and intuition behind the participation-stimulating transfer scheme defined by equations (2)–(5).

We start out by looking at only three agents and dedicate to one of those a strategically inoperative (or, mediating) role. We refer to this agent as M and to the others as 1 and 2. With all else equal, we assume for the moment that it is common knowledge that $\delta_{1M} = 0 = \delta_{2M}$, so we can write $\delta_1 = \delta_{12}$ and $\delta_2 = \delta_{21}$.

The idea is to begin with a strategy-proof social-type dependent transfer scheme s^* that yields 1 and 2 an *ex-ante* transfer of zero while ex-post transfers are paid (received) by M . Then transfers are *ex-post budget-balanced* among $\{1, 2, M\}$, and *interim*-expected utility to M , who has no strategic role to play, is zero. If s^* yields 1 and 2 positive *interim*-expected utility, then M can be given a (sufficiently small) monetary rent by demanding a uniform participation fee from 1 and 2 while preserving positive *interim*-expected utility for 1 and 2. Thereby, also M obtains positive *interim*-expected utility.

We thus look for a (smooth) transfer scheme $s^* = (s_1^*, s_2^*)$ that is *strategy-proof*,

$$(6) \quad \frac{\partial s_i^*(\delta)}{\partial \delta_i} + \delta_i \frac{\partial s_{-i}^*(\delta)}{\partial \delta_i} = 0,$$

yields agents 1 and 2 *ex-ante transfers of zero*,

$$(7) \quad \mathbb{E}_\delta [s_i^*(\delta)] = 0,$$

and yields agents 1 and 2 *positive interim-expected utility*,

$$(8) \quad \mathbb{E}_{\delta_{-i}} [s_i^*(\delta)] + \delta_i \mathbb{E}_{\delta_{-i}} [s_{-i}^*(\delta)] = g_i(\delta_i)$$

for some function $g_i : \Delta_i \rightarrow (0, \infty)$ determining i 's interim-expected *utility gain*.

We can derive s^* from $(g_i)_i$ for appropriate functions $(g_i)_i$.¹²

¹²The sufficient conditions of Proposition 2 can be obtained as follows: By differentiating (6) with respect to δ_{-i} one obtains that $\partial^2 s_i^* / \partial \delta_1 \partial \delta_2 = 0$, implying that s_i^* is additively separable: $s_i^*(\delta) = a_i(\delta_i) + b_i(\delta_{-i})$ for appropriate functions $a_i : \Delta_i \rightarrow \mathbb{R}$ and $b_i : \Delta_{-i} \rightarrow \mathbb{R}$. Hence, by condition (6) again, $a_i'(\delta_i) + \delta_i b_{-i}'(\delta_i) = 0$, such that partial integration yields $a_i(\delta_i) = -\delta_i b_{-i}(\delta_i) + \int_{\delta_{\min}^i}^{\delta_i} b_{-i}(x) dx + C$, for a constant C . Write $g_i(\delta_i) = \int_{\delta_{\min}^i}^{\delta_i} b_{-i}(x) dx + C$. Then, $a_i(\delta_i) = g_i(\delta_i) - \delta_i g_i'(\delta_i)$ and $b_i(\delta_{-i}) = g_{-i}'(\delta_{-i})$, yielding s^* of Proposition 2. Condition (9) can now be imposed to satisfy (7) and (8).

Proposition 2 For smooth functions $g_i : \Delta_i \rightarrow (0, \infty)$ satisfying $g_i'' > 0$ and

$$(9) \quad \mathbb{E}_{\delta_i} [g_i'(\delta_i)] = 0 = \mathbb{E}_{\delta_i} [g_i(\delta_i) - \delta_i g_i'(\delta_i)]$$

define the transfer scheme $s^* = (s_1^*, s_2^*)$ by $s_i^*(\delta) = g_i(\delta_i) - \delta_i g_i'(\delta_i) + g_{-i}'(\delta_{-i})$. Then s^* satisfies conditions (6)–(8). From unanimous participation in s^* , agent i derives an interim-expected utility gain of $g_i(\delta_i) > 0$ while interim-expecting a transfer of $\mathbb{E}_{\delta_{-i}} [s_i^*(\delta)] = g_i(\delta_i) - \delta_i g_i'(\delta_i)$ to herself and a transfer of $\mathbb{E}_{\delta_{-i}} [s_{-i}^*(\delta)] = g_{-i}'(\delta_{-i})$ to agent $-i$.

Proof. We have $d[s_i^*(\hat{\delta}_i, \delta_{-i}) + \delta_i s_{-i}^*(\hat{\delta}_i, \delta_{-i})]/d\hat{\delta}_i = (\delta_i - \hat{\delta}_i)g_i''(\hat{\delta}_i)$; hence, $\hat{\delta}_i = \delta_i$. By (9), $\mathbb{E}_{\delta} [s_i^*(\delta)] = 0$. By (9) again, $\mathbb{E}_{\delta_{-i}} [s_i^*(\delta)] + \delta_i \mathbb{E}_{\delta_{-i}} [s_{-i}^*(\delta)] = [g_i(\delta_i) - \delta_i g_i'(\delta_i)] + \delta_i [g_{-i}'(\delta_{-i})] = g_i(\delta_i) > 0$. Hence, s^* satisfies (6)–(8). All else is obvious. ■

Under s^* of Proposition 2, the transfer that an agent interim-expects for herself is maximal, and positive, if that agent is a pure-payoff maximizer ($\delta_i = 0$), as $d\mathbb{E}_{\delta_{-i}} [s_i^*(\delta)]/d\delta_i = -\delta_i g_i''(\delta_i)$ while $g_i'' > 0$. Money is thus *ex-interim* redistributed to those agents who ‘care least’ about others. On the other hand, the transfer that an agent interim-expects for her opponent increases in her own social type, since $d\mathbb{E}_{\delta_{-i}} [s_{-i}^*(\delta)]/d\delta_i = g_{-i}''(\delta_i) > 0$, and is zero ex ante, since $\mathbb{E}_{\delta_i} [g_{-i}'(\delta_i)] = 0$. Hence, least (most) altruistic types interim-expect to impose a negative (positive) externality on their opponent. This interim-expected externality, weighted with an agent’s social type, overcompensates for interim-expected monetary losses: $\mathbb{E}_{\delta_{-i}} [s_i^*(\delta)] + \delta_i \mathbb{E}_{\delta_{-i}} [s_{-i}^*(\delta)] = g_i(\delta_i) > 0$.

The functions $(g_i)_{i=1,2}$ should be chosen such that common-knowledge assumptions about social-type distributions are as weak as possible. In fact, it suffices to assume common knowledge about mean and variance. It is easy to see that the functions

$$(10) \quad g_i(\delta_i) = \text{Var}_{\delta_i}[\delta_i] + (\delta_i - \mathbb{E}_{\delta_i}[\delta_i])^2$$

satisfy the conditions of Proposition 2. The corresponding transfer scheme becomes

$$(11) \quad s_i^*(\delta) = \mathbb{E}_{\delta_i}[\delta_i^2] - \delta_i^2 + 2(\delta_{-i} - \mathbb{E}_{\delta_{-i}}[\delta_{-i}]).$$

Figure 1 depicts the interim-expected distributive effects of (11): Social types satisfying $|\delta_i| > \sqrt{\mathbb{E}_{\delta_i}[\delta_i^2]}$ incur interim-expected monetary losses (blue), $\mathbb{E}_{\delta_{-i}} [s_i^*(\delta)] < 0$, for which they are overcompensated through sufficiently strong interim-expected exter-

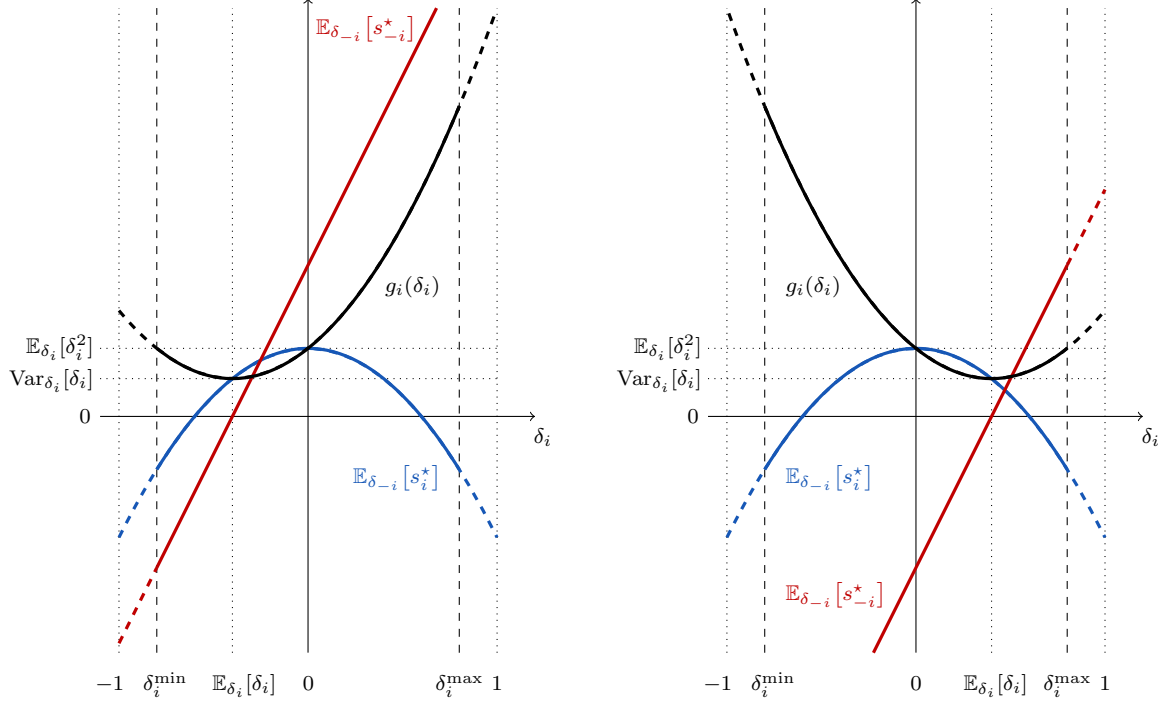


Figure 1: The utility gain $g_i(\delta_i) = \mathbb{E}_{\delta_{-i}}[s_i^*(\delta)] + \delta_i \mathbb{E}_{\delta_{-i}}[s_{-i}^*(\delta)] > 0$ that a social type δ_i interim-expects under the transfer scheme s^* of equation (11), for two different type distributions: $\delta_i \in [\delta_i^{\min}, \delta_i^{\max}] = [-4/5, 4/5]$, $\mathbb{E}_{\delta_i}[\delta_i] = \mp 2/5$, and $\text{Var}_{\delta_i}[\delta_i] = 1/5$, such that $\mathbb{E}_{\delta_{-i}}[s_i^*(\delta)] = 9/25 - \delta_i^2$, $\mathbb{E}_{\delta_{-i}}[s_{-i}^*(\delta)] = 2\delta_i \pm 4/5$, and $g_i(\delta_i) = (\delta_i \pm 2/5)^2 + 1/5$.

nalities (red), $\mathbb{E}_{\delta_{-i}}[s_{-i}^*(\delta)] = 2\delta_i - 2\mathbb{E}_{\delta_i}[\delta_i]$. These interim-expected monetary losses of relatively strong social types are the source for attracting relatively selfish agents with interim-expected monetary gains (blue): $\mathbb{E}_{\delta_{-i}}[s_i^*(\delta)] > 0$ for social types $|\delta_i| < \sqrt{\mathbb{E}_{\delta_i}[\delta_i^2]}$.

From here, we obtain our participation-stimulating transfers (2)–(5) as follows: The interim-expected distributive effects of $s_i^*(\delta) = g_i(\delta_i) - \delta_i g'_i(\delta_i) + g'_{-i}(\delta_{-i})$, discussed above, suggest that participation stimulation is driven by the externality that i imposes on $-i$ through the term $g'_{-i}(\delta_{-i})$. Hence, for the n -agents case, we let $s_i^*(\hat{\delta}) = -C + g_i(\hat{\delta}_i^*) - \hat{\delta}_i^* g'_i(\hat{\delta}_i^*) + \sum_{\ell \neq i, M} g'_\ell(\hat{\delta}_\ell^*)$ for each $i \neq M$, with C the uniform fee given to M .

Under this scheme, now re-accounting for the privately known social preferences toward M , agent i has the dominant strategy to report $\hat{\delta}_i^* = \delta_i^*$ of equation (5).¹³

Finally, the functions g_i of (10) must now be chosen with respect to the random variables δ_i^* . We thus obtain equation (4).

¹³The term $\delta_i^* = \sum_{\ell \neq i, M} (\delta_{i\ell} - \delta_{iM}) / (\delta_{ii} - \delta_{iM})$ gives i 's relative marginal utility from a redistribution of M 's money either to the others, who obtain equal shares, or to i herself. It can be referred to as i 's relative spite towards M , since δ_i^* decreases in δ_{iM} and increases in i 's pro-sociality toward the others, given by $\sum_{\ell \neq i, M} \delta_{i\ell}$.

6 Concluding Remarks

An important question regarding possibility results concerns their practical relevance; whether they serve to show how efficient design can be achieved in practice or to point out practical difficulties in the manner of a “reductio ad absurdum critique.”¹⁴

Our preferred interpretation is the former, as we believe that Theorem 1 captures, in an abstract way, solutions to allocation problems that can be seen in practice. Observe that our participation-stimulating transfer scheme only requires agents to report a one-dimensional sufficient statistic for their social type. The reporting of social types thus translates into players selecting one-dimensional strategies in a strategic game. It is this strategic game that renders participation attractive. A case in point is fundraisers, which are often complemented with unrelated auctions, raffles, or contests (such as awarding the best-dressed guest).

We show in a separate paper how participation can be attracted via various game forms, involving relative- or team-performance incentives and even team contests.

A Appendix

Proof of Lemma 1

Having required weak budget balance, Pareto efficiency implies strict budget balance: Suppose $\sum_{i \in \mathcal{I}} t_i = -\epsilon$ for some $\epsilon > 0$. Then a Pareto improvement can be achieved through transfers $(t_i + \epsilon/n)_{i \in \mathcal{I}}$, since $\sum_{j \in \mathcal{I}} \delta_{ij} > 0$ by assumption.

In the following, let $|\delta_{ij}| < \frac{1}{2n-3}$ for all i and all $j \neq i$. Suppose that, for any fixed transfers $(t_i)_{i \in \mathcal{I}}$, there exists a social alternative $k^\circ(\theta)$ that Pareto-dominates $k^*(\theta)$ while $\sum_{i \in \mathcal{I}} \pi_i(k^\circ | \theta_i) < \sum_{i \in \mathcal{I}} \pi_i(k^* | \theta_i)$. Then there must exist agents i who make strict material losses when switching from k^* to k° : $\pi_i(k^\circ | \theta_i) - \pi_i(k^* | \theta_i) = -\epsilon_i < 0$. Be i^* one of the agents for whom this material loss is largest. Then i^* is *not worse off* under k° than under k^* if and only if she is ‘emotionally’ compensated through the distributive effects on all the others: $\sum_{j \neq i^*} \delta_{i^*j} [\pi_j(k^\circ | \theta_j) - \pi_j(k^* | \theta_j)] \geq \epsilon_{i^*}$. We show that this is impossible.

First suppose $\delta_{i^*j} \leq 0$ for all $j \neq i^*$. Then i^* obtains the maximum ‘emotional’ compensation feasible if also each $j \neq i^*$ realizes the maximum material loss of $-\epsilon_{i^*}$ when

¹⁴We are grateful to an anonymous referee for pointing this out.

switching from k^* to k° ; that is, if $\pi_j(k^\circ | \theta_j) - \pi_j(k^* | \theta_j) = -\epsilon_{i^*} < 0$. But even then, $\sum_{j \neq i^*} \delta_{i^*j} [\pi_j(k^\circ | \theta_j) - \pi_j(k^* | \theta_j)] = \sum_{j \neq i^*} \delta_{i^*j} (-\epsilon_{i^*}) < \epsilon_{i^*}$, since $0 \geq \delta_{i^*j} > \frac{-1}{2n-3} \geq \frac{-1}{n-1}$.

Now suppose $\max_{j \neq i^*} \delta_{i^*j} > 0$, and let $j^* \in \arg \max_{j \neq i^*} \delta_{i^*j}$ be the favorite agent of i^* . Then i^* obtains the maximum ‘emotional’ compensation feasible if j^* realizes a maximum material gain when switching from k^* to k° , under the constraint that $\sum_{j \in \mathcal{I}} \pi_j(k^\circ | \theta_j) < \sum_{j \in \mathcal{I}} \pi_j(k^* | \theta_j)$. This is the case if each $j \neq i^*, j^*$ also realizes the maximum material loss of $-\epsilon_{i^*}$ while aggregate losses, amounting to $(n-1)\epsilon_{i^*}$, serve as a subsidy to agent j^* ; that is, if $\pi_j(k^\circ | \theta_j) - \pi_j(k^* | \theta_j) = -\epsilon_{i^*} < 0$ for all $j \neq i^*, j^*$ while $\pi_{j^*}(k^\circ | \theta_{j^*}) - \pi_{j^*}(k^* | \theta_{j^*}) = (n-1)\epsilon_{i^*}$. But even then, $\sum_{j \neq i^*} \delta_{i^*j} [\pi_j(k^\circ | \theta_j) - \pi_j(k^* | \theta_j)] = \sum_{j \neq i^*, j^*} \delta_{i^*j} (-\epsilon_{i^*}) + \delta_{i^*j^*} (n-1)\epsilon_{i^*} < \frac{n-2}{2n-3} \epsilon_{i^*} + \frac{n-1}{2n-3} \epsilon_{i^*} = \epsilon_{i^*}$, since $|\delta_{i^*j}| < \frac{1}{2n-3}$ for all $j \neq i^*$.

Hence, agent i^* is worse off under k° than under k^* , implying k^* is Pareto-efficient.

The reasoning is exactly the same when showing that, for any fixed social alternative k , no ex-post budget-balanced transfer scheme Pareto-dominates another. ■

References

- Andreoni, James and John Miller. 2002. “Giving according to GARP: An experimental test of the consistency of preferences for altruism.” *Econometrica* 70 (2):737–753.
- Arrow, Kenneth. 1979. “The property rights doctrine and demand revelation under incomplete information.” In *Economics and Human Welfare*, edited by M. J. Boskin. New York, NY: Academic Press.
- Bergemann, Dirk and Stephen Morris. 2005. “Robust mechanism design.” *Econometrica* 73 (6):1771–1813.
- Bierbrauer, Felix and Nick Netzer. 2016. “Mechanism design and intentions.” *Journal of Economic Theory* 163:557–603.
- Bruhin, Adrian, Ernst Fehr, and Daniel Schunk. 2019. “The many faces of human sociality: Uncovering the distribution and stability of social preferences.” *Journal of the European Economic Association* 17 (4):1025–1069.
- Charness, Gary and Matthew Rabin. 2002. “Understanding social preferences with simple tests.” *Quarterly Journal of Economics* 117 (3):817–869.
- Crémer, Jacques and Richard McLean. 1985. “Optimal selling strategies under uncertainty for a discriminating monopolist when demands are interdependent.” *Econometrica* 53 (2):345–361.
- . 1988. “Full extraction of surplus in Bayesian and dominant strategy auctions.” *Econometrica* 56 (6):1247–1257.

- d'Aspremont, Claude and Louis-André Gérard-Varet. 1979. "Incentives and incomplete information." *Journal of Public Economics* 11 (1):25–45.
- Desiraju, Ramarao and David Sappington. 2007. "Equity and adverse selection." *Journal of Economics and Management Strategy* 16 (2):285–318.
- Fehr, Ernst, Karla Hoff, and Mayuresh Kshetramade. 2008. "Spite and development." *American Economic Review: Papers & Proceedings* 98 (2):494–99.
- Jehiel, Philippe, Moritz Meyer-ter Vehn, Benny Moldovanu, and William Zame. 2006. "The limits of ex post implementation." *Econometrica* 74 (3):585–610.
- Jehiel, Philippe and Benny Moldovanu. 2001. "Efficient design with interdependent valuations." *Econometrica* 69 (5):1237–1259.
- Kosenok, Grigory and Sergei Severinov. 2008. "Individually rational, budget-balanced mechanisms and allocation of surplus." *Journal of Economic Theory* 140 (1):126–161.
- Kucuksenel, Serkan. 2012. "Behavioral mechanism design." *Journal of Public Economic Theory* 14 (5):767–789.
- Mailath, George and Andrew Postlewaite. 1990. "Asymmetric information bargaining problems with many agents." *Review of Economic Studies* 57 (3):351–367.
- McAfee, Preston and Philip Reny. 1992. "Correlated information and mechanism design." *Econometrica* 60 (2):395–421.
- McLean, Richard and Andrew Postlewaite. 2004. "Informational size and efficient auctions." *Review of Economic Studies* 71 (3):809–827.
- Mezzetti, Claudio. 2004. "Mechanism design with interdependent valuations: Efficiency." *Econometrica* 72 (5):1617–1626.
- Myerson, Roger. 1979. "Incentive compatibility and the bargaining problem." *Econometrica* 47 (1):61–73.
- Myerson, Roger and Mark Satterthwaite. 1983. "Efficient mechanisms for bilateral trading." *Journal of Economic Theory* 29 (2):265–281.
- Prediger, Sebastian, Björn Vollan, and Benedikt Herrmann. 2014. "Resource scarcity and antisocial behavior." *Journal of Public Economics* 119:1–9.
- Saijo, Tatsuyoshi and Hideki Nakamura. 1995. "The spite dilemma in voluntary contribution mechanism experiments." *Journal of Conflict Resolution* 39 (3):535–560.
- Segal, Ilya and Michael Whinston. 2016. "Property rights and the efficiency of bargaining." *Journal of the European Economic Association* 14 (6):1287–1328.
- Tang, Pingzhong and Tuomas Sandholm. 2012. "Optimal auctions for spiteful bidders." In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*.
- Williams, Steven. 1999. "A characterization of efficient, Bayesian incentive compatible mechanisms." *Economic Theory* 14 (1):155–180.
- Zik, Boaz. 2021. "Ex-post implementation with social preferences." *Social Choice and Welfare* 56 (3):467–485.