ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Daske, Thomas

Working Paper

Efficient Incentives in Social Networks: "Gamification" and the Coase Theorem

Suggested Citation: Daske, Thomas (2019) : Efficient Incentives in Social Networks: "Gamification" and the Coase Theorem, ZBW – Leibniz Information Centre for Economics, Kiel, Hamburg

This Version is available at:
https://hdl.handle.net/10419/193148

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.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

Efficient Incentives in Social Networks: "Gamification" and the Coase Theorem

Thomas Daske*

February 23, 2019

This study explores mechanism design for networks of interpersonal relationships. Agents' social (i.e., altruistic or spiteful) preferences and private payoffs are all subject to asymmetric information; utility is (quasi-)linear, types are independent.

I show that any network of at least three agents can resolve any allocation problem with a mechanism that is Bayesian incentive-compatible, ex-interim individually rational, and ex-post Pareto-efficient (also ex-post budget-balanced). By contrast, a generalized Myerson-Satterthwaite theorem is established for two agents.

The central tool to exploit the *asymmetry of information* about agents' social preferences is "gamification": Resolve the agents' allocation problem with an efficient *social-preference robust* mechanism; ensure agents' participation with the help of a mediator, some network member, who complements that mechanism with an *unrelated* hawk-dove like game between the others, a game that effectively rewards (sanctions) strong (poor) cooperation at the expense (to the benefit) of the mediator. Ex interim, agents (and the mediator) desire this game to be played, for it provides them with a platform to live out their propensities to cooperate or compete.—A figurative example is a fund-raiser, hosted by the "mediator", complemented with awarding the best-dressed guest.

JEL Classification: C70, D62, D64, D82, D85

Keywords: Networks, social preferences, mechanisms, gamification, Coase theorem.

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

For their helpful comments and critical remarks, I want to thank Christian Feilcke, Aart Gerritsen, Julian Hackinger, Kai Konrad, Christoph March, Matthias Regier, Salmai Qari, Marco Sahm, Klaus Schmidt, Johannes Schneider, and Roland Strausz. I am also grateful for helpful comments by participants of the European Winter Meeting of the Econometric Society in Milan, the 5th World Congress of the Game Theory Society in Maastricht, and workshops at the Universities of Augsburg, Berlin, and Leipzig. Financial support of the Max Planck Society is gratefully acknowledged.

1 Introduction

How can people resolve their conflicts of interest in an efficient and mutually acceptable way? The Coase theorem (1960) asserts: As long as property rights are well-defined, contracts are enforceable, and transaction costs are sufficiently low, they can resolve any such conflict through 'negotiations', thereby making outside interventions superfluous.

This claim has been attacked on the grounds of asymmetric information. In its typical mechanism-design interpretation, the Coase theorem reads as follows: *Any group of agents can resolve any allocation problem with a mechanism that is incentive-compatible, ex-interim individually rational, and ex-post Pareto-efficient.* This very statement has been falsified for many allocation problems of practical importance.[1]

In this paper, I shed light on a strategic dimension that is part of the picture in many real-world allocation problems but has just recently gained the attention of economists, namely, *asymmetric information about agents' social preferences.* I show that this sort of information asymmetry does not necessarily cause adverse effects but that it can rather be exploited to satisfy agents' participation constraints and thereby serve the common good. Specifically, I explore mechanism design for groups of agents whose social (i.e., altruistic or spiteful) preferences and private payoffs are *all* subject to asymmetric information, with utility being (quasi-)linear and types being independent. Within this framework, I provide a constructive proof of the Coase theorem for groups of at least three agents.

A figurative example helps clarify the basic idea: Think of a small-town community organizing a fund-raiser in support of their elementary school. The hard-core allocation problem underlying this event is obviously one of public-good provision, and the mechanism to resolve it is actually quite simple, realistically speaking: "Once you're in, you have to give", as a matter of social norm. Events like these are often complemented with some soft-core incentive device, like awarding the best-dressed guest (with a paper crown). The major purpose of such add-on mechanism is not to make guests dress well, but rather to suppress free-riding-at-the-doorstep by compensating participants for their material "losses" (the lost returns from free-riding) with the "fun" they derive from playing games. Awarding the best-dressed guest provides participants with a platform to

[1] Counterexamples have been provided by, e.g., Myerson and Satterthwaite (1983) in the range of bilateral trade, Williams (1999) for multilateral trade, Mailath and Postlewaite (1990) for the provision of public goods, and Segal and Whinston (2016) in a general bargaining framework. For a broader discussion of the Coase theorem in the light of incomplete information, see Farrell (1987).

live out their propensities to compete, and it is this attraction that pulls them over the doorstep.[2]

Based on asymmetric information about agents' social preferences, and in a unified incentive-theoretical framework, I provide a micro-economic foundation for how agents derive "fun" from participating in cooperative or competitive games. Specifically, I show how mutual participation in a game that merely redistributes money between agents can yield them an interim-expected Pareto improvement upon a status quo in which such game is not played.[3] Based on this result, I develop a "gamification" approach to the design of mechanisms:[4] Resolve the agents' allocation problem with an efficient *social-preference robust* mechanism, thereby leaving their social preferences strategically inoperative. Ensure agents' participation with the help of a mediator, some group member, who complements that mechanism with an *unrelated* hawk-dove like game between the others, a game that effectively rewards (sanctions) strong (poor) cooperation at the expense (to the benefit) of the mediator. Ex interim, agents (and the mediator) desire this game to be played, for it provides them with a platform to live out their propensities to cooperate or compete, which are determined by their social preferences.[5]

The paper proceeds as follows. Section 2 outlines the model framework. I introduce *social networks* of interpersonal altruistic (or spiteful) preferences in the manner of Bourlès, Bramoullé, and Perez-Richet (2017); here, however, these preferences are private information.[6] Agents' private payoffs are linear in transfers and their valuations of social alternatives, their *payoff types*. The mechanism-design problem is one of *informational externalities*.[7] As a first step, I characterize the ex-post Pareto frontier of the agents' al-

[2]This point, among others, is frequently made in conceptual research on how to organize fund-raising events; e.g., Webber (2004) and Peloza and Hassay (2007).

[3]The skeptical reader might now want to jump to Section 6.1 where a simple example illustrates how this is possible.

[4]The abundance of characterizations of "gamification" all surround the idea of applying "game design principles in non-gaming contexts" (Robson et al., 2015). Economically speaking, the idea is simply to provide people with additional incentives, often the 'behavioral' type of. This is essentially what I am proposing in a mechanism-design context.

[5]It is important that this add-on game exploits a *different* strategic dimension (namely, agents' social preferences) than the 'basic' mechanism that is to resolve the allocation problem (and works on preferences for 'consumption'). In this respect, I deviate from the literature on how to finance public goods with the revenues from auctions or raffles; e.g., Franke and Leininger (2018).

[6]Recent empirical evidence for people exhibiting altruistic preferences is provided by Bruhin, Fehr, and Schunk (In press). The linear framework considered here is a technically-motivated simplification of their more nuanced findings.

[7]The model is close to the one by Jehiel and Moldovanu (2001), the two differences being that (1) the extent to which an agent perceives those externalities (here called: his social preferences) is common

location problem. Assuming agents have no access to an outside source of money, ex-post efficient mechanisms are ex-post budget-balanced.

Section 3 identifies necessary and sufficient conditions for a mechanism to be Bayesian incentive-compatible.[8] Based on these results, Section 4 establishes the incentive-theoretical bedrock of this paper: Theorem 1 characterizes fully the set of ex-post efficient Bayesian incentive-compatible mechanisms. These mechanisms take a specific form among the renowned expected-externality mechanisms (due to Arrow, 1979, and d'Aspremont and Gérard-Varet, 1979), and from an interim perspective, they are unique with respect to payoff types. This result is interesting because it constitutes an example of a global-games approach (as introduced by Carlsson and Van Damme, 1993) to the design of mechanisms: The introduction of incomplete information about agents' social preferences diminishes tremendously the set of solutions to the mechanism designer's problem; intuitively, an 'appropriate' mechanism must take into account that agents internalize to some privately known extent the distributive effects that it unfolds. Most importantly, the mechanisms of Theorem 1 can (in dyads: must) be social-preference robust (Corollary 1).[9] I also show that all sufficiency results extend to allocation problems with *quasi*-linear private payoffs.

Theorem 1 is particularly useful when it comes to agents' participation constraints, which I take into account in Section 5. For dyads, I derive a clear-cut criterion for when ex-post efficiency is *incentive-feasible* (i.e., for when there exists an ex-post efficient incentive-compatible mechanism that is ex-interim individually rational; Theorem 2). I interpret this result as a generalized Myerson-Satterthwaite theorem.

I then consider networks of arbitrary size. Under the assumption that mechanisms must be social-preference robust, I derive a positive-monotonic relationship between the lower bounds of agents' social preferences and the 'number' of allocation problems in which ex-post efficiency is incentive-feasible (Theorem 3). Loosely speaking, the incentive

knowledge in their model, and (2) that this perception now extends to other agents' transfers, as agents, through their social preferences, internalize the distributive effects that a mechanism unfolds.

[8]Bayesian implementation has been criticized for assuming that agents' type distributions are common knowledge. In order to avoid such strong assumption, Bergemann and Morris (2005) have proposed ex-post implementation for model economies with interdependent utilities (in which dominant-strategy implementation is not feasible), requiring truthful revelation of types to constitute a Nash equilibrium under the respective mechanism. However, as shown by Jehiel et al. (2006), if agents impose informational externalities on each other, then ex-post implementation is no longer feasible, thus leaving Bayesian implementation as the only one among the well-established equilibrium concepts.

[9]This result generalizes the observation of Bartling and Netzer (2016), Bierbrauer and Netzer (2016), and Bierbrauer et al. (2017) that mechanisms for dyads can be social-preference robust: I show they can be so in larger networks, too, and even must be so in dyads.

feasibility of *efficiency gains* ranges from 'almost never' to 'almost always', depending on the *least* altruistic types potentially present in the network.[10]

Theorem 3 suggests that there might be something to gain from considering mechanisms that do not leave social preferences strategically inoperative. At this point, I break with the conventional wisdom in the mechanism-design literature that social-preference robustness would be desirable. The conviction behind this desire is that Wilson's (1987) call for avoiding common-knowledge assumptions (here: about social-type distributions) would be best replied to by rendering social preferences entirely irrelevant.[11] What I propose instead is to look at the trade-off between making common-knowledge assumptions about *some* higher moments of social-type distributions and the incentive feasibility of ex-post efficiency. In this spirit, Sections 6 and 7 are dedicated to gamification.

In Section 6, I define gamification formally. For expositional reasons, I first replace ex-post budget balance by the weaker notion of ex-ante budget balance, which I will interpret as the introduction of a mediator who is an *outsider* to the agents' network. This simplification allows me to construct a *gamification mechanism* successively and thereby make the economic intuition behind it transparent. Theorem 4 shows that gamification is feasible with an outside mediator. This finding has some interesting implications. One is that ex-ante and ex-post budget balance are not "equivalent" in dyads (Corollary 2), contrary to the result of Börgers and Norman (2009) which I will discuss in detail then. Another one is that the mediator can even use gamification to reap profits (Corollary 3), an insight on which I build a simple rationale for casino gambling (Corollary 4).

In Section 7, I adapt the gamification mechanism derived in Section 6 in such a way that the role of the mediator can be taken by some network member, implying that ex-post efficiency is incentive-feasible in *autonomous* social networks (Theorem 5). This is the major result of this paper, as it gives proof to the Coase theorem for networks of at least three agents.

Finally, in Section 8, I discuss some limitations to the practical scope of gamification. One of those is risk aversion, for which I provide a possibility result.

[10]This result generalizes the observation of Kucuksenel (2012) that ex-post efficiency is incentive-feasible for bilateral trade and public-goods provision if agents are sufficiently altruistic, an observation obtained under the assumption that agents' social preferences are common knowledge.

[11]See, e.g., Bartling and Netzer (2016), Bierbrauer and Netzer (2016), and Bierbrauer et al. (2017).

2 The Model

2.1 The Allocation Problem

There is a group $I = \{1, ..., n\}$ of $n \geq 2$ agents and there is a compact set $K \subset \mathbb{R}^m$ of social alternatives, with $m \geq 1$. From alternative $k \in K$ and a transfer $t_i \in \mathbb{R}$, agent i gains a *private payoff*

$$\pi_i(k, t_i \,|\, \theta_i) \quad = \quad \theta_i v_i(k) + w_i(k) + t_i,$$

where the functions $v_i, w_i : \text{conv}(K) \to \mathbb{R}$ are continuous on the convex hull, $\text{conv}(K)$, of K. Agent i's *payoff type* θ_i belongs to a closed interval $\Theta_i = [\theta_i^{\min}, \theta_i^{\max}]$. The collection of agents' payoff types is denoted by θ; by convention, $\theta = (\theta_i, \theta_{-i})$, with $\theta_{-i} = (\theta_j)_{j \neq i}$.

Agents exhibit interpersonal preferences in the manner of altruism or spite: From the allocation of private payoffs (and the observation of others' payoff types), agent i derives *utility*

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

where the value δ_{ij} that i assigns to j's payoff belongs to a closed interval $\Delta_{ij} = [\delta_{ij}^{\min}, \delta_{ij}^{\max}] \subset \left(\frac{-1}{n-1}, 1\right)$, while $\delta_{ii} = 1$. I 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 agents' types specify a structure that can be represented by a complete double-directed graph in which nodes (one for each agent) are weighted by payoff types and edges are weighted by the interpersonal preferences between agents. I refer to this structure as the agents' *social network*. Figure 1 gives an illustration.

Each agent i is privately informed about his type. However, it is common knowledge that i's payoff type and social type realize independently according to continuous densities $f_i : \Theta_i \to (0, \infty)$ and $g_i : \Delta_i \to (0, \infty)$, with Cartesian product $\Delta_i = \Pi_{j \in I \setminus \{i\}} \Delta_{ij}$. Agents' types are independent. On the intra-personal level, the various degrees of altruism determining an agent's social type may correlate.[12]

[12] Despite the asymmetry of information, it can still be common knowledge who is 'friends' and who is 'foes'. For instance, if $\delta_{ij}^{\min}, \delta_{ji}^{\min} > 1/2$ and $\delta_{kl}^{\max}, \delta_{lk}^{\max} < 0$, then, in comparison, i and j are friends, whereas k and l are foes. Similarly, it can be common knowledge that i dislikes j more than ℓ, which is

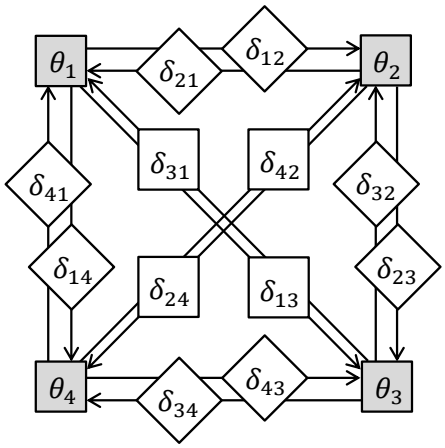

Figure 1: *A social network of $n = 4$ agents with payoff types $(\theta_i)_i$ and interpersonal altruistic or spiteful preferences $(\delta_{ij})_{i,j\neq i}$, all of which are private information.*

I make use the following notation: I denote the collection of social types by $\delta = (\delta_i, \delta_{-i})$ and write the collection of agents' types as $(\theta, \delta) = (\theta_i, \theta_{-i}, \delta_i, \delta_{-i})$. Cartesian products of type sets are denoted by $\Theta = \Pi_{i\in I}\Theta_i$ and $\Delta = \Pi_{i\in I}\Delta_i$. Let also $\pi_i(k \,|\, \theta_i) = \theta_i v_i(k) + w_i(k)$ and $u_i(k, \theta_{-i} \,|\, \theta_i, \delta_i) = \sum_{j\in I} \delta_{ij}\pi_j(k \,|\, \theta_j)$.

I assume (and relax this assumption only in Section 6) that agents do not have access to an outside source of money, such that transfers must be weakly budget-balanced: $\sum_{i\in I} t_i \leq 0$. The agents' problem is to choose a social alternative k and transfers $(t_i)_i$ such that the resulting allocation is ex-post Pareto-efficient:

Lemma 1 *The ex-post Pareto frontier of the agents' allocation problem is characterized by a social alternative $k^*(\theta) \in \arg\max_{k\in K} \sum_{i\in I} \pi_i(k \,|\, \theta_i)$ and budget-balanced transfers $(t_i)_{i\in I}$, such that $\sum_{i\in I} t_i = 0$. If $\Delta_{ij} \subset \left(\frac{-1}{2n-3}, \frac{1}{2n-3}\right)$ for all $i \in I$ and all $j \in I \setminus \{i\}$, then no budget-balanced transfer scheme Pareto-dominates another.*

Proof. See Appendix A.1. ■

Notice that the Pareto frontier can be indefinite for some combinations of social types satisfying $|\delta_{ij}| > 1/(2n-3)$ if $n \geq 3$, in which case a subgroup of agents might be willing to transfer arbitrary amounts of money to their joint favorite agent.[13] This possibility stems from the assumption that utility is linear in transfers and that there are no endowment

the case if $\delta_{ij}^{\max} < \delta_{i\ell}^{\min}$. The key assumption is rather that in any bilateral relationship, there remains some uncertainty about who (dis)likes whom how much.

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

constraints. In what follows, I focus on social networks that are *generic* in the sense that the Pareto frontier of the agents' allocation problem is determinate, characterized by Lemma 1. Notice, however, that the rare case of unlimited ex-post Pareto improvements through interpersonal transfers does not resolve the agents' information problem.

2.2 Efficient Incentives

A *direct revelation mechanism* involves the agents in a strategic game of incomplete information in which they are asked to report their types truthfully; based on their reports, a social alternative will be implemented and transfers will be made.[14] An *efficient mechanism* is determined by a social-choice function (k^*, T), consisting of an ex-post efficient allocation function $k^* : \Theta \to K$, with $k^*(\theta) \in \arg\max_{k \in K} \sum_{i \in I} \pi_i(k \,|\, \theta_i)$, and a budget-balanced transfer scheme $T = (t_i)_{i \in I} : \Theta \times \Delta \to \mathbb{R}^n$, just in line with Lemma 1.

For a random variable $X : \Theta \times \Delta \to \mathbb{R}$, denote by $\mathbb{E}_{\theta_{-i}, \delta_{-i}}\big[X(\theta, \delta)\big]$ the expected value of X for a given type (θ_i, δ_i). For some allocation function $k : \Theta \to K$, some transfer scheme $T : \Theta \times \Delta \to \mathbb{R}^n$, and $x_i \in \{v_i, w_i\}$, define $\bar{x}_i(\theta_i) = \mathbb{E}_{\theta_{-i}}\big[x_i(k(\theta))\big]$, $\bar{\pi}_{ij}(\theta_i) = \mathbb{E}_{\theta_{-i}}\big[\pi_j(k(\theta) \,|\, \theta_j)\big]$, and $\bar{t}_{ij}(\theta_i, \delta_i) = \mathbb{E}_{\theta_{-i}, \delta_{-i}}\big[t_j(\theta, \delta)\big]$. (Payoffs and transfers are assumed to be Lebesgue-integrable with respect to type distributions.) 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 his true type is (θ_i, δ_i) and if all the other agents report their types truthfully:

$$U_i(\hat{\theta}_i, \hat{\delta}_i \,|\, \theta_i, \delta_i) \;=\; \theta_i \bar{v}_i(\hat{\theta}_i) + \bar{w}_i(\hat{\theta}_i) + \bar{t}_{ii}(\hat{\theta}_i, \hat{\delta}_i) + \sum_{j \in I \setminus \{i\}} \delta_{ij} \left[\bar{\pi}_{ij}(\hat{\theta}_i) + \bar{t}_{ij}(\hat{\theta}_i, \hat{\delta}_i)\right].$$

For convenience, write $U_i(\theta_i, \delta_i) = U_i(\theta_i, \delta_i \,|\, \theta_i, \delta_i)$. An allocation function is *Bayesian implementable* if there exists a transfer scheme such that, for all i and all (θ_i, δ_i),

$$U_i(\theta_i, \delta_i) \;=\; \max_{\hat{\theta}_i, \hat{\delta}_i} U_i(\hat{\theta}_i, \hat{\delta}_i \,|\, \theta_i, \delta_i).$$

In this case, the mechanism defined by (k, T) is (*Bayesian*) *incentive-compatible*.

[14] By the *revelation principle*, which applies to the present setup (Myerson, 1979), there is no loss of generality in considering only direct mechanisms.

3 Incentive-Compatible Mechanisms

In order to characterize the entire set of efficient incentive-compatible mechanisms, I first ask for properties that incentive-compatible mechanisms have in common. Due to Lemma 1, I restrict attention to allocation functions that are independent from social types.

Lemma 2 *A social-choice function (k, T), with $k : \Theta \to K$, is incentive-compatible only if it satisfies the following conditions for all $i \in I$ and all $j \in I \setminus \{i\}$:*

(i) $\bar{v}_i(\theta_i)$ *is non-decreasing in* θ_i.

(ii) $\bar{t}_{ij}(\theta_i, \delta_i)$ *is non-decreasing in* δ_{ij}.

(iii) There exist a partially differentiable function $p_i : \Delta_i \to \mathbb{R}$ *and functions* $q_{ij} : \Theta_i \times \Pi_{\ell \in I \setminus \{i,j\}} \Delta_{i\ell} \to \mathbb{R}$, *partially differentiable in* θ_i, *such that, for* $\delta_i^{-j} = (\delta_{il})_{\ell \in I \setminus \{i,j\}}$,

$$U_i(\theta_i, \delta_i) = p_i(\delta_i) + \int_{\theta_i^{\min}}^{\theta_i} \bar{v}_i(r)\, dr \tag{1}$$

$$= q_{ij}(\theta_i, \delta_i^{-j}) + \int_{\delta_{ij}^{\min}}^{\delta_{ij}} \left[\bar{\pi}_{ij}(\theta_i) + \bar{t}_{ij}(\theta_i, r, \delta_i^{-j})\right] dr. \tag{2}$$

Proof. See Appendix A.2. ■

Conditions (i) and (1) are the well-known results for Bayesian implementation in linear settings with independent valuations (Myerson, 1981), and conditions (ii) and (2) are their social-preference equivalents. Condition (2) implies in particular that an incentive-compatible mechanism must take into account that agents internalize to some extent its distributive effects; consequently, expected individual transfers are now linked.[15]

Contrary to linear settings with independent valuations, where conditions (i) and (1) are even sufficient (Myerson, 1981), the conditions of Lemma 2 do not yet ensure incentive compatibility:

Proposition 1 *A social-choice function (k, T), with $k : \Theta \to K$, is incentive-compatible if and only if it satisfies the following conditions for all $i \in I$ and all $j \in I \setminus \{i\}$:*

[15] Nevertheless, asymmetric information about agents' social preferences does not much restrict the set of Bayesian implementable allocation functions if no further constraints are imposed on transfers: Zik (2019), in a slightly more general framework, shows that any allocation function $k : \Theta \to K$ that can be Bayesian implemented in the commonly known absence of social preferences is also Bayesian implementable if social preferences are private information.

(i) $\bar{v}_i(\theta_i)$ *is non-decreasing in* θ_i.

(ii) There exists a partially differentiable function $p_i : \Delta_i \to \mathbb{R}$ *such that, for all* $\theta_i \in \Theta_i$ *and all* $\delta_i, \hat{\delta}_i \in \Delta_i$,

$$
\begin{aligned}
&(3) & p_i(\delta_i) - p_i(\hat{\delta}_i) &\geq \nabla p_i(\hat{\delta}_i) \cdot (\delta_i - \hat{\delta}_i), \\
&(4) & \bar{t}_{ij}(\theta_i, \delta_i) &= \frac{\partial p_i(\delta_i)}{\partial \delta_{ij}} - \bar{\pi}_{ij}(\theta_i), \\
&(5) & \bar{t}_{ii}(\theta_i, \delta_i) &= p_i(\delta_i) - \nabla p_i(\delta_i) \cdot \delta_i - \bar{\pi}_{ii}(\theta_i) + \int_{\theta_i^{\min}}^{\theta_i} \bar{v}_i(r)\, dr.
\end{aligned}
$$

Proof. See Appendix A.3. ■

By Proposition 1, incentive compatibility requires interim-expected transfers to be additively separable in an agent's payoff type and social type. Conditions (i), (4), and (5) are a immediate from Lemma 2; they ensure incentive compatibility with respect to payoff types. The distinctive feature of Proposition 1 is condition (3) which ensures incentive compatibility with respect to social types. When choosing interim-expected transfers independent from social types (by letting p_i affine), then the necessary conditions of Lemma 2 are also sufficient. Notice that condition (3) implies but is not equivalent to the convexity of p_i, since p_i might not be *continuously* partially differentiable.

4 Ex-Post Efficient Implementation

In this section, I prove and interpret the following Theorem.

Theorem 1 (Social-Preference Compatible Mechanisms: SPC-Mechanisms) *An ex-post efficient allocation function* $k^*(\theta) \in \arg\max_{k \in K} \sum_{i \in I} \pi_i(k \mid \theta_i)$ *can be Bayesian implemented through ex-post budget-balanced transfers if and only if transfers* $T^* = (t_i^*)_{i \in I}$ *take the following form: For reported types* $(\hat{\theta}, \hat{\delta})$,

$$
t_i^*(\hat{\theta}, \hat{\delta}) = \sum_{\ell \in I \setminus \{i\}} \Big[\mathbb{E}_{\theta_{-i}} \big[\pi_\ell(k^*(\hat{\theta}_i, \theta_{-i}) \mid \theta_\ell) \big] - \mathbb{E}_{\theta_{-\ell}} \big[\pi_i(k^*(\hat{\theta}_\ell, \theta_{-\ell}) \mid \theta_i) \big] \Big] + s_i(\hat{\theta}, \hat{\delta}),
$$

where the components $s_i : \Theta \times \Delta \to \mathbb{R}$ *satisfy the following conditions for all* $i, j \in I$:

(i) $\sum_{i \in I} s_i(\theta, \delta) = 0$ *on* $\Theta \times \Delta$.

(ii) $\mathbb{E}_{\theta_{-i},\delta_{-i}}[s_j(\theta,\delta)]$ *is independent from* θ_i *for all* δ_i.

(iii) $\left(\mathbb{E}_{\theta_{-i},\delta_{-i}}[s_j(\theta,\delta)]\right)_{j\neq i} = \nabla p_i(\delta_i) + a_i$ *for all* $\delta_i \in \Delta_i$, *where* $a_i \in \mathbb{R}^{n-1}$ *is constant and* $p_i : \Delta_i \to \mathbb{R}$ *is partially differentiable and satisfies* $p_i(\delta_i) + \nabla p_i(\delta_i) \cdot (\mathbf{1} - \delta_i) = c_i$ *and* $p_i(\delta_i) - p_i(\hat{\delta}_i) \geq \nabla p_i(\hat{\delta}_i) \cdot (\delta_i - \hat{\delta}_i)$ *for some* $c_i \in \mathbb{R}$ *and all* $\delta_i, \hat{\delta}_i \in \Delta_i$.[16]

Theorem 1 states that, besides the very constrained choice of $(s_i)_i$, there exists a unique mechanism that is ex-post efficient and incentive-compatible.[17] This mechanism belongs to the class of expected-externality mechanisms, or AGV-mechanisms (after Arrow, 1979, and d'Aspremont and Gérard-Varet, 1979), which I take to be defined broadly through transfers $t_i^*(\hat{\theta}) = \sum_{\ell\neq i} \mathbb{E}_{\theta_{-i}}\left[\pi_\ell(k^*(\hat{\theta}_i, \theta_{-i}) \,|\, \theta_\ell)\right] + h_i(\hat{\theta}_{-i})$ for some functions $(h_i)_i$. In what follows, I refer to the mechanisms of Theorem 1 as *social-preference compatible* mechanisms (*SPC-mechanisms*).[18]

Leaving the $(s_i)_i$ aside, SPC-transfers can be interpreted as follows: Each i makes a gross concession to every other ℓ amounting to what ℓ expects i's payoff to be when reporting $\hat{\theta}_\ell$; that is, i pays $\mathbb{E}_{\theta_{-\ell}}\left[\pi_i(k^*(\hat{\theta}_\ell, \theta_{-\ell}) \,|\, \theta_i)\right.$ to ℓ, and ℓ pays $\mathbb{E}_{\theta_{-i}}\left[\pi_\ell(k^*(\hat{\theta}_i, \theta_{-i}) \,|\, \theta_\ell)\right]$ to i. Loosely speaking, i pays ℓ the money equivalent of what ℓ believes to contribute to i's material well-being (when reporting $\hat{\theta}_\ell$), and vice versa.

Now consider the incentives to reveal social preferences: Obviously, any affine function $p_i(\delta_i) = b_i \cdot (\mathbf{1} - \delta_i)$, with $b_i \in \mathbb{R}^{n-1}$, satisfies the conditions of Theorem 1(iii), in which case interim-expected transfers are independent from social types, rendering them strategically irrelevant. In dyads, this strategic irrelevance of social types is even necessary.[19] Intuitively, since each agent assigns greater value to his own than to his opponent's payoff and since transfers are supposed to be budget-balanced, an agent will always report a social type that yields him the maximum (or: close-to-supremum) interim-expected

[16] **Notation:** $\mathbf{1} = (1, ..., 1) \in \mathbb{R}^{n-1}$.

[17] Furthermore, from an interim perspective and with respect to payoff types, this mechanism is in fact unique; this follows from condition (ii) of Theorem 1 and Lemma 3 below. Notice that the uniqueness is not caused by social preferences per se but rather by the asymmetry of information about them: I show in Daske (2017) for $n = 2$ that, if social preferences are common knowledge, there exist other than SPC-mechanisms that are efficient and incentive-compatible.

[18] The conventional AGV is defined through

$$h_i(\hat{\theta}_{-i}) \quad = \quad \frac{-1}{n-1} \cdot \sum_{j\neq i} \sum_{\ell\neq j} \mathbb{E}_{\theta_{-j}}\left[\pi_\ell(k^*(\hat{\theta}_j, \theta_{-j}) \,|\, \theta_\ell)\right].$$

The AGV is social-preference compatible *if and only if* $n = 2$.

[19] **Proof:** For $n = 2$, the partial differential equation in Theorem 1(iii) reads $p_i(\delta_i) + p_i'(\delta_i)(1 - \delta_i) = c_i$, where $p_i : (-1, 1) \to \mathbb{R}$. Hence, $p_i'' = 0$. (I obtained this result earlier in Daske, 2017.)

transfer whenever that transfer varies in social types. However, this intuition does no longer apply in larger networks: Suppose there exists a report $\hat{\delta}_i$ such that $\bar{t}_{ii}(\hat{\delta}_i) > \bar{t}_{ii}(\delta_i)$ for i's true type δ_i; then i might still report δ_i if doing so yields some agent a huge transfer at the expense of a third agent that i likes less. In fact, in networks of at least three agents, social preferences can be strategically operative.[20] In other words:

Corollary 1 *Ex-post efficient implementation can be social-preference robust; in dyads, it must be.*

The next two Propositions give proof of Theorem 1.

Proposition 2 *Ex-post efficient incentive-compatible mechanisms are of SPC-type.*

Proof. Suppose k^* can be Bayesian implemented through budget-balanced transfers $T = (t_i)_{i \in I}$. Notice that one can always write

$$t_j(\theta, \delta) \quad = \quad \sum_{\ell \in I \setminus \{j\}} \Big[\mathbb{E}_{\theta_{-j}} \big[\pi_\ell(k^*(\theta) \,|\, \theta_\ell) \big] - \mathbb{E}_{\theta_{-\ell}} \big[\pi_j(k^*(\theta) \,|\, \theta_j) \big] \Big] + s_j(\theta, \delta)$$

for appropriate functions $s_j : \Theta \times \Delta \to \mathbb{R}$ satisfying $\sum_{j \in I} s_j = 0$. Then, for all i,

$$\begin{aligned} \bar{t}_{ij}(\theta_i, \delta_i) \quad &\overset{j \neq i}{=} \quad \sum_{\ell \in I} \mathbb{E}_\theta \big[\pi_\ell(k^*(\theta) \,|\, \theta_\ell) \big] - (n-1) \cdot \mathbb{E}_\theta \big[\pi_j(k^*(\theta) \,|\, \theta_j) \big] \\ &\qquad - \mathbb{E}_{\theta_{-i}} \big[\pi_j(k^*(\theta) \,|\, \theta_j) \big] + \mathbb{E}_{\theta_{-i}, \delta_{-i}} \big[s_j(\theta, \delta) \big]. \end{aligned} \tag{6}$$

By Proposition 1(ii), there exists a partially differentiable function $p_i : \Delta_i \to \mathbb{R}$ such that

$$\frac{\partial p_i(\delta_i)}{\partial \delta_{ij}} \quad = \quad \mathbb{E}_{\theta_{-i}} \big[\pi_j(k^*(\theta) \,|\, \theta_j) \big] + \bar{t}_{ij}(\theta_i, \delta_i). \tag{7}$$

[20] An example is the "gamified" SPC-mechanism derived in Section 7. Another, simpler example is the following: The conditions of Theorem 1(iii) imply convexity of $p_i(\delta_i)$ and, due to Euler's formula, linear homogeneity of its transform $p_i(\mathbf{1} - \delta_i)$; the inequality is equivalent to the convexity of p_i if p_i is *continuously* partially differentiable. Each function $p_i(\delta_i) = ||\mathbf{1} - \delta_i||$, with some continuously partially differentiable *norm* $|| \cdot || : \mathbb{R}^{n-1} \to [0, \infty)$, satisfies these conditions; for those functions, the following transfer components $(s_i)_i$ are non-constant (also ex interim), ex-post budget-balanced, and even dominant-strategy incentive-compatible:

$$s_i(\hat{\delta}) \quad = \quad \sum_{j \neq i} \frac{\partial p_j(\hat{\delta}_j)}{\partial \delta_{ji}} - \sum_{j \neq i} \frac{\partial p_i(\hat{\delta}_i)}{\partial \delta_{ij}}.$$

Substituting (7) into (6) yields

$$\mathbb{E}_{\theta_{-i},\delta_{-i}}\big[s_j(\theta,\delta)\big] \overset{j\neq i}{=} \frac{\partial p_i(\delta_i)}{\partial \delta_{ij}} - \sum_{\ell\in I}\mathbb{E}_\theta\big[\pi_\ell(k^*(\theta)\,|\,\theta_\ell)\big] + (n-1)\cdot\mathbb{E}_\theta\big[\pi_j(k^*(\theta)\,|\,\theta_j)\big].$$

Hence, $\mathbb{E}_{\theta_{-i},\delta_{-i}}\big[s_j(\theta,\delta)\big]$ is independent from θ_i for all $j\neq i$. Due to budget balance, the same holds for $\mathbb{E}_{\theta_{-i},\delta_{-i}}\big[s_i(\theta,\delta)\big]$. In particular, there exists a constant vector $a_i\in\mathbb{R}^{n-1}$ such that $\big(\mathbb{E}_{\theta_{-i},\delta_{-i}}[s_j(\theta,\delta)]\big)_{j\neq i} = \nabla p_i(\delta_i)+a_i$. By identities (4) and (5) of Proposition 1, budget-balance requires

$$\begin{aligned} 0 &= \sum_{j\in I}\bar{t}_{ij}(\theta_i,\delta_i) \\ &= p_i(\delta_i)+\nabla p_i(\delta_i)\cdot(\mathbf{1}-\delta_i) - \sum_{j\in I}\mathbb{E}_{\theta_{-i}}\big[\pi_j(k^*(\theta)\,|\,\theta_j)\big] + \int_{\theta_i^{\min}}^{\theta_i}\bar{v}_i(r)\,dr. \end{aligned}$$

Hence, there must exist a constant $c_i\in\mathbb{R}$ such that $p_i(\delta_i)+\nabla p_i(\delta_i)\cdot(\mathbf{1}-\delta_i)=c_i$ for all δ_i. By condition (3) of Proposition 1, p_i must also satisfy $p_i(\delta_i)-p_i(\hat{\delta}_i)\geq\nabla p_i(\hat{\delta}_i)\cdot(\delta_i-\hat{\delta}_i)$ for all $\delta_i,\hat{\delta}_i$. ■

Proposition 3 *SPC-mechanisms are ex-post efficient and incentive-compatible.*[21]

Proof. Obviously, T^* is budget-balanced; thus, SPC-mechanisms are ex-post efficient by construction. By assumption, there exists a partially differentiable function $p_i:\Delta_i\to\mathbb{R}$, satisfying $p_i(\delta_i)+\nabla p_i(\delta_i)\cdot(\mathbf{1}-\delta_i)=c_i$ and $p_i(\delta_i)-p_i(\hat{\delta}_i)\geq\nabla p_i(\hat{\delta}_i)\cdot(\delta_i-\hat{\delta}_i)$, such that $\partial p_i(\delta_i)/\partial\delta_{ij}+a_{ij}=\mathbb{E}_{\theta_{-i},\delta_{-i}}[s_j(\theta,\delta)]$ for some $a_{ij}\in\mathbb{R}$. Suppose the agents other than i reveal their types truthfully. Then interim-expected transfers are given by

$$\begin{aligned} \mathbb{E}_{\theta_{-i}}[t_i(\hat{\theta}_i,\hat{\delta}_i,\theta_{-i},\delta_{-i})] &= \sum_{\ell\in I\setminus\{i\}}\mathbb{E}_{\theta_{-i}}\big[\pi_\ell(k^*(\hat{\theta}_i,\theta_{-i})\,|\,\theta_\ell)\big] - (n-1)\cdot\mathbb{E}_\theta\big[\pi_i(k^*(\theta)\,|\,\theta_i)\big] \\ &\quad - \Big(\nabla p_i(\hat{\delta}_i)+a_i\Big)\cdot\mathbf{1}, \end{aligned}$$

[21] For dyads, this result was obtained earlier by Bartling and Netzer (2016) in the range of altruism and spite, and by Bierbrauer and Netzer (2016) for intention-based preferences à la Rabin (1993).

$$
\begin{aligned}
\mathbb{E}_{\theta_{-i}}[t_j(\hat{\theta}_i, \hat{\delta}_i, \theta_{-i}, \delta_{-i})] \overset{j \neq i}{=} & \sum_{\ell \in I \setminus \{i,j\}} \mathbb{E}_{\theta_{-i}, \theta_{-j}} \big[\pi_\ell(k^*(\theta_j, \theta_{-j}) \,|\, \theta_\ell) \\
& - \sum_{\ell \in I \setminus \{i,j\}} \mathbb{E}_{\theta_{-i}, \theta_{-\ell}} \big[\pi_j(k^*(\theta_\ell, \theta_{-\ell}) \,|\, \theta_j) \big] \big] \\
& + \mathbb{E}_{\theta_{-i}, \theta_{-j}} \big[\pi_i(k^*(\theta_j, \theta_{-j}) \,|\, \theta_i) \big] \\
& - \mathbb{E}_{\theta_{-i}} \big[\pi_j(k^*(\hat{\theta}_i, \theta_{-i}) \,|\, \theta_j) \big] \\
& + \frac{\partial p_i(\hat{\delta}_i)}{\partial \delta_{ij}} + a_{ij} \\
= & \sum_{\ell \in I} \mathbb{E}_\theta \big[\pi_\ell(k^*(\theta) \,|\, \theta_\ell) \big] - (n-1) \cdot \mathbb{E}_\theta \big[\pi_j(k^*(\theta) \,|\, \theta_j) \big] \\
& - \mathbb{E}_{\theta_{-i}} \big[\pi_j(k^*(\hat{\theta}_i, \theta_{-i}) \,|\, \theta_j) \big] + \frac{\partial p_i(\hat{\delta}_i)}{\partial \delta_{ij}} + a_{ij}.
\end{aligned}
$$

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

$$
\begin{aligned}
U_i(\hat{\theta}_i, \hat{\delta}_i \,|\, \theta_i, \delta_i) &= \sum_{j \in I} \delta_{ij} \cdot \mathbb{E}_{\theta_{-i}} \big[\pi_j(k^*(\hat{\theta}_i, \theta_{-i}) \,|\, \theta_j) + t_j(\hat{\theta}_i, \theta_{-i}) \big] \\
&= \mathbb{E}_{\theta_{-i}} \left[\sum_{\ell \in I} \pi_\ell(k^*(\hat{\theta}_i, \theta_{-i}) \,|\, \theta_\ell) \right] \\
&\quad + \left(\sum_{j \in I \setminus \{i\}} \delta_{ij} \right) \cdot \mathbb{E}_\theta \left[\sum_{\ell \in I} \pi_\ell(k^*(\theta) \,|\, \theta_\ell) \right] \\
&\quad - (n-1) \cdot \mathbb{E}_\theta \left[\sum_{j \in I} \delta_{ij} \pi_j(k^*(\theta) \,|\, \theta_j) \right] \\
&\quad - \nabla p_i(\hat{\delta}_i) \cdot (\mathbf{1} - \delta_i) - a_i \cdot (\mathbf{1} - \delta_i).
\end{aligned} \tag{8}
$$

If truthful revelation of θ_i was inferior for i, then there would exist $\hat{\theta}_i$ and θ such that $\sum_{\ell \in I} \pi_\ell(k^*(\hat{\theta}_i, \theta_{-i}) \,|\, \theta_\ell) > \sum_{\ell \in I} \pi_\ell(k^*(\theta_i, \theta_{-i}) \,|\, \theta_\ell)$, which contradicts the definition of k^*. If truthful revelation of δ_i was inferior, then there would exist $\hat{\delta}_i$ such that $\nabla p_i(\hat{\delta}_i) \cdot (\mathbf{1} - \delta_i) < \nabla p_i(\delta_i) \cdot (\mathbf{1} - \delta_i)$; but since p_i satisfies $p_i(\delta_i) + \nabla p_i(\delta_i) \cdot (\mathbf{1} - \delta_i) = c_i$, this implies that

$$
\begin{aligned}
0 &> \nabla p_i(\hat{\delta}_i) \cdot (\mathbf{1} - \delta_i) - \nabla p_i(\delta_i) \cdot (\mathbf{1} - \delta_i) \\
&= \nabla p_i(\hat{\delta}_i) \cdot (\mathbf{1} - \delta_i) + p_i(\delta_i) - c_i - p_i(\hat{\delta}_i) - \nabla p_i(\hat{\delta}_i) \cdot (\mathbf{1} - \hat{\delta}_i) + c_i \\
&= \nabla p_i(\hat{\delta}_i) \cdot (\hat{\delta}_i - \delta_i) + p_i(\delta_i) - p_i(\hat{\delta}_i),
\end{aligned} \tag{9}
$$

which contradicts the assumption that $p_i(\delta_i) - p_i(\hat{\delta}_i) \geq \nabla p_i(\hat{\delta}_i) \cdot (\delta_i - \hat{\delta}_i)$ for all $\delta_i, \hat{\delta}_i$. ■

Proposition 4 *SPC-mechanisms are ex-post efficient and incentive-compatible even if private payoffs are quasi-linear, $\pi_i(k \mid \theta_i) = \hat{v}_i(k \mid \theta_i)$, as long as the $\hat{v}_i : K \times \Theta_i \to \mathbb{R}$ are such that SPC-mechanisms are well-defined (i.e., k^* and all expected values exist).*

Proof. The proof of Proposition 3 applies word by word. ■

5 Interim Participation Constraints

Suppose there exists a status-quo allocation k° which is always a social alternative, $k^\circ \in K$, and that implementing a different allocation requires mutual consent of all agents at the stage at which types are private information.

Denote status-quo private payoffs by $\pi_i^\circ(\theta_i) = \pi_i(k^\circ \mid \theta_i)$; agent i's *interim reservation utility* is thus given by $u_i^\circ(\theta_i, \delta_i) = \sum_{j \in I} \delta_{ij} \cdot \mathbb{E}_{\theta_j}\left[\pi_j^\circ(\theta_j)\right]$. For an ex-post efficient allocation function $k^* : \Theta \to \mathbb{R}$, write $\pi_i^*(\theta) = \pi_i(k^*(\theta) \mid \theta_i)$; hence, $\sum_{i \in I} \pi_i^*(\theta) \geq \sum_{i \in I} \pi_i^\circ(\theta)$ on Θ. I will say that *ex-post efficiency is incentive-feasible* in a given allocation problem if there exists an ex-post efficient mechanism that is Bayesian incentive-compatible and ex-interim individually rational, such that $U_i(\theta_i, \delta_i) \geq u_i^\circ(\theta_i, \delta_i)$ on $\Theta_i \times \Delta_i$ for all $i \in I$. By Theorem 1, only SPC-mechanisms need to be considered, in which case $U_i(\theta_i, \delta_i)$ is given by equation (8). I restrict attention to SPC-mechanisms that are *social-preference robust* which, by Corollary 1, is necessary for dyads. For networks of at least three agents, social-preference robustness yields a reference point for the gamification approach developed below. The following Lemma eases the exposition.

Lemma 3 *If SPC-transfer components $(s_i(\theta, \delta))_{i \in I}$ are social-preference robust, such that $\mathbb{E}_{\theta_{-j}, \delta_{-j}}[s_i(\theta, \delta)]$ is constant in both θ_i and δ_i, then there exist constants $(S_i)_{i \in I}$, with $\sum_{i \in I} S_i = 0$, such that $\mathbb{E}_{\theta_{-j}, \delta_{-j}}[s_i(\theta, \delta)] = S_i$ for all $i, j \in I$.*

Proof. See Appendix A.4. ■

Theorem 1 yields a clear-cut criterion for when ex-post efficiency is incentive-feasible in dyads.

Theorem 2 (A Generalized Myerson-Satterthwaite Theorem) *In dyads, ex-post efficiency is incentive-feasible if and only if, for all $\theta \in \Theta$,*

$$0 \;\geq\; \mathbb{E}_\theta\big[\pi_1^*(\theta)+\pi_2^*(\theta)\big] + \sum_{i=1,2} \frac{u_i^\circ(\theta_i,\delta_i^{\min}) - \mathbb{E}_{\theta_{-i}}\big[\pi_1^*(\theta)+\pi_2^*(\theta)\big]}{1-\delta_i^{\min}}. \tag{10}$$

Proof. By Theorem 1, ex-post efficient mechanisms are incentive-compatible if and only if transfers are of SPC-type. In this case, interim-expected utility from participation is given by

$$\begin{aligned} U_1(\theta_1,\delta_1) &= \mathbb{E}_{\theta_2}\big[\pi_1^*(\theta)+\pi_2^*(\theta)\big] - (1-\delta_1)\cdot\mathbb{E}_\theta\big[\pi_1^*(\theta)\big] + (1-\delta_1)\cdot\mathbb{E}_{\theta_2,\delta_2}\big[s(\theta,\delta)\big], \\ U_2(\theta_2,\delta_2) &= \mathbb{E}_{\theta_1}\big[\pi_1^*(\theta)+\pi_2^*(\theta)\big] - (1-\delta_2)\cdot\mathbb{E}_\theta\big[\pi_2^*(\theta)\big] - (1-\delta_2)\cdot\mathbb{E}_{\theta_1,\delta_1}\big[s(\theta,\delta)\big], \end{aligned}$$

where $\mathbb{E}_{\theta_i,\delta_i}[s(\theta,\delta)]$ must be constant on $\Theta_{-i}\times\Delta_{-i}$ for all i. By Lemma 3, there exists a constant S such that $\mathbb{E}_{\theta_2,\delta_2}\big[s(\theta,\delta)\big] = S = \mathbb{E}_{\theta_1,\delta_1}\big[s(\theta,\delta)\big]$. Hence, $U_i(\theta_i,\delta_i) \geq u_i^\circ(\theta_i,\delta_i)$ holds for both i if and only if there exists a constant S such that both

$$S \;\geq\; \mathbb{E}_\theta\big[\pi_1^*(\theta)\big] + \frac{u_1^\circ(\theta_1,\delta_1) - \mathbb{E}_{\theta_2}\big[\pi_1^*(\theta)+\pi_2^*(\theta)\big]}{1-\delta_1}, \tag{11}$$

$$S \;\leq\; -\mathbb{E}_\theta\big[\pi_2^*(\theta)\big] - \frac{u_2^\circ(\theta_2,\delta_2) - \mathbb{E}_{\theta_1}\big[\pi_1^*(\theta)+\pi_2^*(\theta)\big]}{1-\delta_2}, \tag{12}$$

which is equivalent to

$$0 \;\geq\; \mathbb{E}_\theta\big[\pi_1^*(\theta)+\pi_2^*(\theta)\big] + \sum_{i=1,2} \frac{u_i^\circ(\theta_i,\delta_i) - \mathbb{E}_{\theta_{-i}}\big[\pi_1^*(\theta)+\pi_2^*(\theta)\big]}{1-\delta_i} \tag{13}$$

for all (θ,δ). Denote the right hand-side of (13) by $G(\theta,\delta)$. Notice that $u_i^\circ(\theta_i,\delta_i) = \pi_i^\circ(\theta_i) + \mathbb{E}_{\theta_{-i}}\big[\pi_{-i}^\circ(\theta_{-i})\big] \leq \mathbb{E}_{\theta_{-i}}\big[\pi_1^*(\theta)+\pi_2^*(\theta)\big]$, since $\pi_1^\circ+\pi_2^\circ \leq \pi_1^*+\pi_2^*$. Hence, $G(\theta,\delta)$ is weakly decreasing in δ_i, such that (13) holds on $\Theta\times\Delta$ if and only if (10) holds on Θ. ■

Condition (10) resembles Williams' (1999) criterion for the existence of "desirable" mechanisms in allocation problems with $n \geq 2$ agents whose utilities are independent. Desirable mechanisms, by definition, Bayesian implement k^*, are ex-interim individually rational, and satisfy *ex-ante* budget balance. The criterion derived here (for $n = 2$) is stricter in that it builds on *ex-post* budget balance: As I show below (see Corollary 2),

ex-ante and ex-post budget balance are *not* equivalent under incentive-feasible implementation if agents' social preferences are private information, contrary to allocation problems with independent utilities. Based on his criterion, Williams (1999) establishes the Myerson-Satterthwaite theorem for more than two agents, a result that I show to be not robust with respect to asymmetric information about agents' social preferences (Theorem 5).

The famous non-existence of ex-post efficient incentive-feasible mechanisms for bilaterally trading an indivisible good (Myerson and Satterthwaite, 1983) is readily obtained, thus giving Theorem 2 its name, when applying condition (10) to the allocation problem specified by: $K = \{0, 1\}$; $k^\circ = 1$; $\pi_1 = \theta_1 k$; $\pi_2 = \theta_2(1-k)$; $\Theta_i \subset [0, \infty)$; $\Theta_1 \cap \Theta_2 \neq \emptyset$; and $\delta_i^{\min} = 0$. The Myerson-Satterthwaite theorem has been put into perspective through plenty of model variations, which typically concern the choice of K and k°.[22] I cannot confront this vast literature with Theorem 2 here, but I want to stress that this theorem provides a simple test to many bilateral-bargaining models, while Theorem 1 and Lemma 3 provide the desired mechanisms if they exist. All one would have to accept is that agents' social preferences (e.g., $\delta_i = 0$) are not common knowledge.

Theorem 1 helps explain how the feasibility of mutually beneficial agreements varies with the support of social-type distributions if only social-preference robust mechanisms are allowed.

Theorem 3 ('Worst-Type' Monotonicity) *The 'number' of allocation problems in which ex-post efficiency is incentive-feasible through social-preference robust mechanisms is in the following sense increasing in the lower bounds of agents' social preferences:*

(i) As $\delta_{ij}^{\min} \downarrow \frac{-1}{n-1}$ for all i and all $j \neq i$, ex-post efficiency is incentive-feasible if and only if $\pi_i^\circ(\theta_i) - \mathbb{E}_{\theta_i}\big[\pi_i^\circ(\theta_i)\big] = \mathbb{E}_{\theta_{-i}}\big[\sum_{j\in I} \pi_j^*(\theta)\big] - \mathbb{E}_\theta\big[\sum_{j\in I} \pi_j^*(\theta)\big]$ for all i and θ_i.

(ii) If ex-post efficiency is (not) incentive-feasible for given $(\delta_{ij}^{\min})_{i,j\neq i}$, then it is also (not) if $\delta_{ij}^{\min}$ is larger (smaller) for some i and some $j \neq i$.

(iii) As $\delta_{ij}^{\min} \uparrow 1$ for some i and all $j \neq i$, ex-post efficiency becomes incentive-feasible whenever i, regardless his payoff type, ex-interim expects a strict efficiency gain from participation: $\mathbb{E}_{\theta_{-i}}\big[\sum_{j\in I} \pi_j^*(\theta)\big] > \mathbb{E}_{\theta_{-i}}\big[\sum_{j\in I} \pi_j^\circ(\theta)\big]$ for all θ_i.

[22] E.g., $K = [0, 1]$ as in McAfee (1991), or $k^\circ = 1/2$ as in Cramton, Gibbons, and Klemperer (1987). Segal and Whinston (2011) even make k° contingent on payoff-type distributions.

Proof. Conditions (ii) and (iii) are trivial. Ad (i): Define $G_i(\theta_i, \delta_i) = U_i(\theta_i, \delta_i) - u_i^\circ(\theta_i, \delta_i)$. Notice that G_i is continuous in (θ_i, δ_i); hence, if $\delta_{ij}^{\min} \downarrow \frac{-1}{n-1}$ for all $j \neq i$, then $G_i(\theta_i, \delta_i) \geq 0$ holds for all θ_i if and only if $G_i(\theta_i, (\frac{-1}{n-1})_j) \geq 0$. Due to Lemma 3 and equation (8), this latter condition can be written as

$$
\begin{aligned}
(14) \qquad \frac{n}{n-1} \cdot S_i &= S_i + \sum_{j \in I \setminus \{i\}} \frac{-1}{n-1} \cdot S_j \\
&\geq \pi_i^\circ(\theta_i) - \frac{1}{n-1} \cdot \sum_{j \in I \setminus \{i\}} \mathbb{E}_{\theta_j}\big[\pi_j^\circ(\theta_j)\big] \\
&\qquad + n \cdot \mathbb{E}_\theta\big[\pi_i^*(\theta)\big] - \mathbb{E}_{\theta_{-i}}\left[\sum_{\ell \in I} \pi_\ell^*(\theta)\right],
\end{aligned}
$$

for constants $(S_j)_j$ satisfying $\sum_j S_j = 0$. Interim individual rationality requires condition (14) to hold for all i and all θ_i. Summing (14) over all i yields:

$$
0 \;\geq\; \sum_{j \in I} \pi_j^\circ(\theta_j) - \sum_{j \in I} \mathbb{E}_{\theta_j}\big[\pi_j^\circ(\theta_j)\big] + n \cdot \mathbb{E}_\theta\left[\sum_{\ell \in I} \pi_\ell^*(\theta)\right] - \sum_{j \in I} \mathbb{E}_{\theta_{-j}}\left[\sum_{\ell \in I} \pi_\ell^*(\theta)\right].
$$

Denote the right-hand side of this condition by $G^-(\theta)$ and observe that $\mathbb{E}_\theta\big[G^-(\theta)\big] = 0$, such that $0 \geq G^-(\theta)$ holds on Θ if and only if $G^-(\theta) = 0$ for all θ. Interim individual rationality thus requires that $0 = \mathbb{E}_{\theta_{-i}}\big[G^-(\theta)\big] = \pi_i^\circ(\theta_i) - \mathbb{E}_{\theta_i}\big[\pi_i^\circ(\theta_i)\big] + \mathbb{E}_\theta\big[\sum_{\ell \in I} \pi_\ell^*(\theta)\big] - \mathbb{E}_{\theta_{-i}}\big[\sum_{\ell \in I} \pi_\ell^*(\theta)\big]$ for all i and all θ_i. On the other hand, if $\mathbb{E}_{\theta_{-i}}\big[G^-(\theta)\big] = 0$ for all i, then condition (14) reads

$$
\frac{n}{n-1} \cdot S_i \;\geq\; \mathbb{E}_{\theta_i}\big[\pi_i^\circ(\theta_i)\big] - \mathbb{E}_\theta\big[\sum_{\ell \in I} \pi_\ell^*(\theta)\big] - \frac{1}{n-1} \cdot \sum_{j \in I \setminus \{i\}} \mathbb{E}_{\theta_j}\big[\pi_j^\circ(\theta_j)\big] + n \cdot \mathbb{E}_\theta\big[\pi_i^*(\theta)\big],
$$

both sides of which sum to zero when summing over all i; hence, when choosing the $(S_j)_j$ such that this latter condition holds with equality, then $G_i(\theta_i, (\frac{-1}{n-1})_j) \geq 0$ for all i. ■

Condition (ii) is trivial because eliminating (adding) types makes things easier (harder). Condition (iii) is trivial because a perfect altruist is willing to make *any* monetary concession as long as it serves the common goal; if this money is equally distributed among the others, they will participate, too. Theorem 3 gains substance from condition (i) which constitutes a meaningful anchor for a chain of banal implications. For instance, if private payoffs are continuously differentiable in k, and $w_i = 0$ for all i, then (i) implies that k^*

must satisfy $\sum_{i\in I}\pi_i^*(\theta) = \sum_{i\in I}\pi_i^\circ(\theta)$.[23] In this case, extremely spiteful types are stuck in the status quo.[24] Notice that $\delta_{ij}^{\min} \geq 0$ as $n \to \infty$; hence, in very large networks, *efficiency gains* tend to be not incentive-feasible under social-preference robust mechanisms even if agents are not spiteful but, at worst, selfish.[25]

6 Participation Through Gamification

In this section, I pursue an approach similar to the one in Myerson and Satterthwaite (1983) who introduce a rent-seeking intermediary to bilateral trade and show that ex-post efficient trade can be rendered incentive-feasible among those sellers and buyers for whom the gains from trade are sufficiently high. Here, however, a different strategic dimension will be utilized: agents' social preferences.[26]

A *mediator*, as conceived in this section, is an agent M *outside* the social network about whom the following is common knowledge: (i) no agent in the social network cares about her (i.e., $\delta_{iM} = 0$ for all $i \in I$); (ii) she is a pure-payoff maximizer, which means in particular that she does not care about any of the members of the social network; (iii) she knows the distribution of agents' types; (iv) she has the option to complement a social-preference robust SPC-mechanism with a payoff-type independent game the outcomes of which redistribute money between agents and herself and vary with the composition of social types at play; (v) she can commit to the ex-post outcomes of the complemented mechanism; and (vi) she is willing to offer such a complemented mechanism if and only if she ex-ante expects a non-negative profit from it.

I define *gamification* as follows.

Definition 1 (Gamification Mechanisms) *A gamification mechanism is specified by a social-choice function* (k^*, T^{**}) *consisting of an ex-post efficient allocation function*

[23] **Proof:** Differentiating the equation in (i) with respect to θ_i yields $v_i(k^\circ) = \mathbb{E}_{\theta_{-i}}\big[v_i(k^*)\big]$, such that $\mathbb{E}_\theta\big[\pi_i(k^\circ)\big] = \mathbb{E}_\theta\big[\pi_i(k^*)\big]$ if $w_i = 0$. Hence, $\mathbb{E}_\theta\big[\sum_{i\in I}\pi_i(k^\circ)\big] = \mathbb{E}_\theta\big[\sum_{i\in I}\pi_i(k^*)\big]$, while $\sum_{i\in I}\pi_i(k^\circ) \leq \sum_{i\in I}\pi_i(k^*)$, implying that $\sum_{i\in I}\pi_i(k^\circ) = \sum_{i\in I}\pi_i(k^*)$.

[24] Utility-wise, this result is obvious, and it holds generally: Since $\sum_{j\in I}\delta_{ij} \to 0$ as $(\delta_{ij})_{j\neq i} \to (\frac{-1}{n-1})_{j\neq i}$, equal shares of payoff-wise efficiency gains would vanish utility-wise. The result is interesting because extremely spiteful types are even *payoff-wise* stuck in the status quo.

[25] This observation resembles the result of Mailath and Postlewaite (1990) who show that the likelihood of ex-post efficient public-goods provision being incentive-feasible approaches zero as the number of (selfish) agents approaches infinity.

[26] This section builds on ideas developed in an unpublished working paper of mine; see Daske (2016).

$k^* : \Theta \to K$, *with* $k^*(\theta) \in \arg\max_{k \in K} \sum_{i \in I} \pi_i(k \,|\, \theta_i)$, *and a transfer scheme* $T^{**} = (t_i^{**})_{i \in I} : \Theta \times \Delta \to \mathbb{R}^n$ *given by*

$$(15) \qquad t_i^{**}(\hat{\theta}, \hat{\delta}) \;=\; \sum_{\ell \in I \setminus \{i\}} \Big[\mathbb{E}_{\theta_{-i}} \big[\pi_\ell(k^*(\hat{\theta}_i, \theta_{-i}) \,|\, \theta_\ell) \big] - \mathbb{E}_{\theta_{-\ell}} \big[\pi_i(k^*(\hat{\theta}_\ell, \theta_{-\ell}) \,|\, \theta_i) \big] \Big] + s_i^*(\hat{\delta}),$$

where the 'gamification scheme' $(s_i^*)_{i \in I}$, $s_i^* : \Delta \to \mathbb{R}$, *satisfies the following conditions:*

(i) $(s_i^*)_{i \in I}$ *is dominant-strategy incentive-compatible.*

(ii) $(s_i^*)_{i \in I}$ *is ex-ante budget-balanced: For all* $\delta \in \Delta$,

$$(16) \qquad \sum_{i \in I} \mathbb{E}_\delta \big[s_i^*(\delta) \big] \;=\; 0.$$

(iii) Ex interim, $(s_i^*)_{i \in I}$ *yields each agent* $i \in I$ *a positive utility level: For all* $\delta \in \Delta$,

$$(17) \qquad \sum_{\ell \in I} \delta_{i\ell} \cdot \mathbb{E}_{\delta_{-i}} \big[s_\ell^*(\delta) \big] \;>\; 0.$$

The *payoff*-type related transfer components of gamification mechanisms are those of SPC-mechanisms; these components are ex-post budget-balanced and ensure the implementation of k^*, due to Theorem 1. By condition (16), gamification replaces ex-post budget balance by the weaker notion of ex-ante budget-balance. When interpreting gains and losses on the network level as subsidies or sanctions through a mediator, then the *mediator's ex-post payoff* is given by

$$s_M(\delta) \;=\; -\sum_{i \in I} s_i^*(\delta),$$

such that her own (interim) participation constraint is satisfied through condition (16). Condition (17) requires that agents *desire* gamification, since it implies an interim-expected Pareto improvement upon choosing the $(s_i^*)_{i \in I}$ constantly zero. In order to cope with Wilson's (1987) call for avoiding common knowledge assumptions (here: about social-type distributions), I require gamification schemes to be dominant-strategy rather than Bayesian incentive-compatible.[27]

[27] This serves the purpose of not assuming common knowledge of social-type distributions right from the outset, even though interim individual rationality will require common knowledge of *some* higher moments of those distributions.

The purpose of the following subsections is to show that gamification mechanisms do actually exist, the central result being Proposition 5. Anticipating this result, the following Theorem states what gamification is good for.

Theorem 4 (Gamification) *For any collection of reservation utilities $(u_i^\circ)_{i\in I}$, there exists a gamification mechanism that is incentive-feasible, ex-ante budget-balanced, and implements the ex-post efficient allocation function k^*. The result extends to allocation problems with quasi-linear private payoffs.*

Proof. Consider a gamification mechanism (k^*, T^{**}) with transfers $(s_i^*)_{i\in I}$ from Proposition 5 below. This mechanism is ex-ante budget-balanced and Bayesian implements k^*, due to Theorem 1 and Propositions 4 and 5. Notice that the conditions of Definition 1 are invariant under scaling all the $(s_i^*)_{i\in I}$ with the same $\alpha > 0$. Due to condition (17), agents' interim participation constraints can thus be satisfied for given $(u_i^\circ)_{i\in I}$ by choosing $\alpha > 0$ sufficiently large. ■

The proof of Theorem 4 makes clear that my gamification approach builds on (arbitrary) interim-expected Pareto improvements through social-preference dependent transfers that redistribute money between agents and mediator. This would be impossible in generic social networks if agents' social types were common knowledge and participation required ex-post individual rationality; an insight immediate from Lemma 1.

Corollary 2 *In dyads, ex-ante and ex-post budget balance are not 'equivalent': There (do not) exist ex-interim individually rational mechanisms that Bayesian implement k^* and satisfy ex-ante (ex-post) budget balance.*[28]

Proof. Immediate from Theorems 2 and 4. ■

Börgers and Norman (2009) prove the equivalence of ex-ante and ex-post budget balance in a model that allows for informational externalities. In terms of the model conceived here, they prove the equivalence for some sort of 'semi-altruism', one under which each agent i internalizes others' valuations of social alternatives k but disregards the distributive effects of transfers; formally, as if $u_i = \sum_{j\in I} \delta_{ij} \cdot \pi_j(k \,|\, \theta_j) + t_i$. The here

[28]Regarding the validity of this statement for $n \geq 3$, the results of this paper are inconclusive.

obtained non-equivalence of ex-ante and ex-post budget balance illustrates the incentive-theoretical impact of agents internalizing fully the distributive effects of a mechanism.

The mediator can even reap profits from gamification:

Corollary 3 *For any collection of reservation utilities $(u_i^\circ)_{i\in I}$, a rent-seeking mediator can offer agents a mechanism that is incentive-feasible, implements the efficient allocation function k^*, and yields the mediator a positive profit ex ante.*

Proof. Consider a gamification mechanism in the manner of Theorem 4, with gamification scheme $(s_i^*)_{i\in I}$. Then the gamification scheme $(s_i^{**}(\delta))_{i\in I} = (s_i^*(\delta) - \epsilon)_{i\in I}$, with

$$0 \;<\; \epsilon \;<\; \frac{\sum_{\ell\in I} \delta_{i\ell} \cdot \mathbb{E}_{\delta_{-i}}\left[s_\ell^*(\delta)\right]}{\sum_{\ell\in I} \delta_{i\ell}},$$

is dominant-strategy incentive-compatible, satisfies $\sum_{\ell\in I} \delta_{i\ell} \cdot \mathbb{E}_{\delta_{-i}}\left[s_\ell^{**}(\delta)\right] > 0$, and yields the mediator an ex-ante profit of $n \cdot \epsilon$. When scaling all the $s_i^{**}(\delta)$ with the same, sufficiently large $\alpha > 0$, each agent's participation constraint will be satisfied and the mediator's ex-ante profit amounts to $\mathbb{E}_\delta\left[s_M(\delta)\right] = n \cdot \alpha \cdot \epsilon > 0$. ■

When leaving the choice of $k \in K$ out of account, Corollary 3 suggests a rationale for casino gambling (in the manner of competitive games like poker):

Corollary 4 (Rational Gambling) *Under asymmetric information about agents' social preferences, a rent-seeking 'casino' can offer a game that redistributes money between visitors and the 'bank' such that visitors play voluntarily and the casino reaps profits in expectation. The game is a 'gamble' in that its outcomes depend on the unknown social types of (other) players and effectively determine their competitiveness.*[29]

The conventional economic stance on casino gambling, taking an expected-payoff perspective, has been that it is irrational but that it might be rationally addictive (e.g., Becker and Murphy, 1988, and Mobilia, 1993). A behavioral rationale, based on "stake-dependent" utility, has recently been put forward by Eisenhuth, Murphy, and Neuhierl (2018). Another, more naïve, view is to pronounce the entertainment aspect of gambling

[29] Game outcomes may also have an exogenous random component (like rolling a dice) in which case gamification transfers can be interpreted as expected values. For instance, ex-post transfers might be $\hat{s}_i(\delta, X) = X \cdot s_i^{**}(\delta)$, for some real-valued random variable X satisfying $\mathbb{E}_X[X] = 1$ and the gamification scheme $(s_i^{**}(\delta))_{i\in I}$ in the proof of Corollary 3; in this case, $s_i^{**}(\delta) = \mathbb{E}_X[\hat{s}_i(\delta, X)]$.

(e.g., Marfels, 2001), simply acknowledging that gambling might be 'fun', that gamblers receive some sort of mental compensation for their expected material losses. The rationale put forward here is more subtle than that, for it provides an explanation for the *origin* of such mental compensation, namely, agents' social preferences which entail both agents' distributive preferences and their propensities to cooperate or compete. To make transparent how these aspects of social preferences can be exploited to produce a sort of mental compensation that actually outweighs the disutility from an anticipated material loss is another purpose of the following subsections, next to showing that gamification mechanisms do actually exist.

This section proceeds as follows: Without loss of generality, I focus on gamification schemes $(s_i^*)_{i \in I}$ and suppress the payoff-type related transfer components of gamification mechanisms. I start out (in 6.1) with a simple example that provides some basic intuition for how arbitrary interim-expected Pareto improvements among mediator and agents are feasible. I then derive a gamification scheme (in 6.2) for a simplified version of the model framework of Section 2. This mechanism can be adjusted (in 6.3) to match the general model framework.

6.1 A Simple Example

Consider a mediator M and two agents $i \in \{1, 2\}$ whose social types are i.i.d., taking values $\delta_i = 0$ or $\delta_i = 1/2$, each with probability of $1/2$.

The mediator offers agents to play a game the material outcomes of which are determined by the transfer scheme $(s_i)_{i=1,2}$ depicted on the left-hand side of Figure 2; the ex-post payoff of M is given by $s_M = -s_1 - s_2$. From a pure-payoff perspective, this is a prisoners' dilemma, which could be overcome through mutual pro-social behavior among agents. Maximizing utility $u_i = s_i + \delta_i \cdot s_{-i}$, a pure-payoff maximizer ($\delta_i = 0$) has the dominant strategy to behave *hawkish* in that he opts for playing H; an altruist ($\delta_i = 1/2$) has the dominant strategy to behave *dovish* in that he opts for playing D; the respective utility levels are depicted on the right-hand side of Figure 2. The game offered by M is thus dominant-strategy incentive-compatible, distinguishing hawks from doves behaviorally.

Reservation payoffs and, thus, reservation utilities are assumed to be zero for all.

s_i \ s_{-i}	D		H	
D	$4x$	$4x$	$-8x$	$8x$
H	$8x$	$-8x$	$-4x$	$-4x$

u_{Dove} \ u_{-i}	D		H	
D	$6x$	...	$-4x$	...
H	$4x$	...	$-6x$	...

u_{Hawk} \ u_{-i}	D		H	
D	$4x$	...	$-8x$	...
H	$8x$	...	$-4x$	...

Figure 2: *On the left: The transfer scheme* $(s_i)_{i=1,2}$ *underlying the strategic game that a mediator* M *offers two agents to play, where* $x > 0$ *and where* M*'s ex-post payoff is given by* $s_M = -s_1 - s_2$*. At the upper right: An altruist, with* $u_i = s_i + s_{-i}/2$*, acts dovish by playing his dominant strategy* D*. At the lower right: A pure-payoff maximizer, with* $u_i = s_i$*, acts hawkish by playing his dominant strategy* H*.*

Figure 3 depicts the payoffs s_ℓ, for $\ell \in \{D, H, M\}$, and utilities u_ℓ, for $\ell \in \{D, H\}$, that a hawk (H, with $\delta_i = 0$), a dove (D, with $\delta_i = 1/2$), and the mediator would derive from each of the feasible type combinations. Figure 3 also depicts the probabilities ($\Pr_\ell$) that the mediator and *privately informed* social types do rationally assign to each of those type combinations. The mediator's interim-expected payoff, assuming hawks and doves participate, is zero; her participation constraint as stated by Definition 1(ii) is thus satisfied. A hawk, assuming the other agent participates, interim-expects a payoff of $2x > 0$ and would thus participate, too.

The decisive role is played by the dovish type: Assuming the other agent participates whatever his type may be, he expects a *material loss* of $1/2 \cdot (4x - 8x) = -2x$, for which he is *mentally compensated* by the positive externality his appearance would have on both a hawk (who would obtain $8x$) or another dove (who would obtain $4x$); the resulting interim-expected positive externality, amounting to $1/2 \cdot (8x + 4x) = 6x$, weighted with his degree of altruism, provides a dovish type with an interim-expected *mental compensation* of $\delta_D \cdot 6x = 3x$, which outweighs his interim-expected *material loss* of $-2x$.[30]

[30] Leaving all else equal, the above results can also be obtained, qualitatively, when letting $\delta_D \in (\frac{1}{2}, 1)$. In this case, the respective ex-post utility levels from playing hawk or dove (colored red and blue in Figure 2) are those of a *hawk-dove game* in the strict, game-theoretical sense.

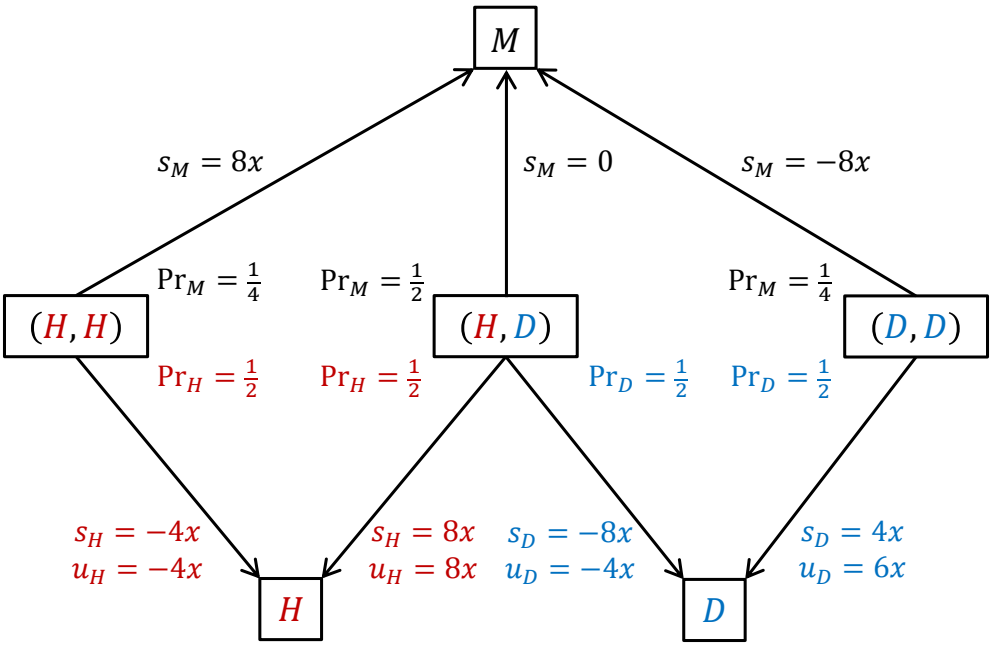

Figure 3: *The payoffs and utilities that a hawk (H), a dove (D), and the mediator (M) would obtain under each of the feasible combinations of social types, and the probabilities* ($\Pr_\ell$, $\ell \in \{D, H, M\}$) *they rationally assign to each of those type combinations.*

The transfer scheme $(s_i)_{i=1,2}$ thus satisfies the conditions of Definition 1, such that mutual participation in the respective game yields an interim-expected Pareto improvement upon the status quo. Letting x approach infinity even allows for arbitrary interim-expected Pareto improvements.[31] I show in Appendix B how similar examples can be constructed for arbitrary i.i.d. types and discrete social-type sets. The example presented here is a special case of those.

[31] To make the example a little more lively, the transfer scheme on the left-hand side of Figure 2 can be interpreted as resulting from *relative* or *team performance incentives*. For instance, M could offer an all-pay auction in which agents can either invest or abstain, $e_i \in \{0, 1\}$, to obtain a net return of

$$s_i(e_i, e_{-i}) \quad = \quad -F - C \cdot e_i + \begin{cases} P, & e_i > e_{-i} \\ P/2, & e_i = e_{-i} \\ 0, & e_i < e_{-i} \end{cases}$$

where $F = 8x$ is a participation fee, $C = 8x$ are investment costs, and $P = 24x$ is the prize; in this case, the hawkish strategy H is to invest, whereas the dovish strategy D is to abstain. Payoff-equivalently, M could offer a public-good game in which agents do either invest or free-ride, $e_i \in \{0, 1\}$, to obtain a net return of $s_i(e_i, e_{-i}) = -F - C \cdot e_i + B \cdot (e_i + e_{-i})$, at participation fee $F = 4x$, investment costs $C = 16x$, and team bonus per unit of effort $B = 12x$; here, strategy H is to free-ride, whereas D is to invest.

6.2 Sufficient Conditions in a Simplified Model

With all else equal to the model setup of Section 2.1, suppose here that there are only two agents, $i \in \{1, 2\}$, and their social types are i.i.d.. Ease notation by rewriting $\delta_{i,-i} = \delta_i$.

A reasonable approach is to look for a gamification scheme $(s_i^*)_i$ that is *symmetric* in the sense that $s_1^*(\delta) = s(\delta_1, \delta_2)$ and $s_2^*(\delta) = s(\delta_2, \delta_1)$ for some twice continuously partially differentiable function $s : \Delta \to \mathbb{R}$. The $(s_i^*)_i$ are supposed to be dominant-strategy incentive-compatible. Due to i.i.d. social types, ex-ante budget balance requires

$$\mathbb{E}_{\delta_1,\delta_2}\big[s(\delta_1, \delta_2)\big] \quad = \quad 0. \tag{18}$$

Due to condition (17),

$$\mathbb{E}_{\delta_{-i}}\big[s(\delta_i, \delta_{-i})\big] + \delta_i \cdot \mathbb{E}_{\delta_{-i}}\big[s(\delta_{-i}, \delta_i)\big] \quad = \quad g(\delta_i) \tag{19}$$

for some function $g : \Delta_i \to (0, 1)$ determining i's strictly positive utility gain from gamification as compared to a status quo of zero-transfers. I refer to g as the *gain function*. Notice that the gain function is the same for both agents, due to i.i.d. social types, and that it is twice continuously differentiable, due to the assumptions on s.

The idea is to derive s from g for an appropriate g.

Lemma 4 *The following holds necessarily:*

(i) The gain function g is convex and satisfies $\mathbb{E}_{\delta_i}\big[g'(\delta_i)\big] = 0 = \mathbb{E}_{\delta_i}\big[g(\delta_i) - \delta_i \cdot g'(\delta_i)\big]$.

(ii) Agent i of type δ_i expects a transfer of $\mathbb{E}_{\delta_{-i}}\big[s(\delta_i, \delta_{-i})\big] = g(\delta_i) - \delta_i \cdot g'(\delta_i)$ to himself.

(iii) Agent i of type δ_i expects a transfer of $\mathbb{E}_{\delta_{-i}}\big[s(\delta_{-i}, \delta_i)\big] = g'(\delta_i)$ to agent $-i$.

(iv) The gamification scheme $s(\delta_i, \delta_{-i})$ is additively separable.

(v) The gamification scheme is given by $s(\delta_i, \delta_{-i}) = g(\delta_i) - \delta_i \cdot g'(\delta_i) + g'(\delta_{-i})$.

Proof. For each i, the gain function g is exactly the function p_i from Proposition 1(ii) when letting there $n = 2$, $t_i = s_i^*$, and $v_i, w_i = 0$; conditions (ii) and (iii) of Lemma 4 are thus immediate from conditions (5) and (4), respectively. The equation in condition (i) follows from (ii) and (iii), due to (18), and g is convex due to Proposition 1(ii).

Due to strategy proofness, agent i's first-order condition $0 = \frac{\partial}{\partial \delta_i} s(\delta_i, \delta_{-i}) + \delta_i \frac{\partial}{\partial \delta_i} s(\delta_{-i}, \delta_i)$ must be satisfied for all $\delta_i, \delta_{-i} \in (-1, 1)$; differentiating this with respect to δ_{-i} yields $0 = \frac{\partial^2}{\partial \delta_2 \partial \delta_1} s(\delta_1, \delta_2) + \delta_1 \frac{\partial^2}{\partial \delta_2 \partial \delta_1} s(\delta_2, \delta_1)$ and $0 = \frac{\partial^2}{\partial \delta_1 \partial \delta_2} s(\delta_2, \delta_1) + \delta_2 \frac{\partial^2}{\partial \delta_1 \partial \delta_2} s(\delta_1, \delta_2)$, implying that $0 = (1 - \delta_1 \delta_2) \frac{\partial^2}{\partial \delta_1 \partial \delta_2} s(\delta_1, \delta_2)$; hence, s is additively separable. Condition (v) is easily obtained from conditions (i) to (iv). ■

The convexity of g implies that the transfer an agent expects to himself is maximal (and positive) for a pure-payoff maximizer ($\delta_i = 0$) and (locally) minimal for (at least one of) the extreme social types ($\delta_i^{\min}$ and $\delta_i^{\max}$). That is, money is redistributed toward those who 'care least' about others. Since the monetary externality that an agent imposes on his opponent, $\mathbb{E}_{\delta_{-i}}\big[s(\delta_{-i}, \delta_i)\big]$, increases in his social type and is zero ex ante, least (most) altruistic types impose a negative (positive) externality on their opponents. This externality, weighted with an agent's social type, *mentally over-compensates* those agents who expect a *material loss*: $\mathbb{E}_{\delta_{-i}}\big[s(\delta_i, \delta_{-i})\big] + \delta_i \cdot \mathbb{E}_{\delta_{-i}}\big[s(\delta_{-i}, \delta_i)\big] = g(\delta_i) - \delta_i \cdot g'(\delta_i) + \delta_i \cdot g'(\delta_i) = g_i(\delta_i) > 0$. It is easy to see that, from an ex-post perspective, the mediator prefers less altruistic over more altruistic types, while agents prefer more altruistic over less altruistic types.

In order to cope with the 'Wilson doctrine', the gain function should be chosen such that common knowledge assumptions about the social-type distribution are as weak as possible. In fact, it suffices to assume common knowledge of its mean and variance:

Lemma 5 *Let* $\delta^* = -1 + \sqrt{\mathrm{Var}_{\delta_i}[\delta_i] + (1 + \mathbb{E}_{\delta_i}[\delta_i])^2}$. *Then the function* $g : (-1, 1) \to (0, \infty)$, *with* $g(\delta_i) = (\delta_i - \delta^*)^2 + 2(1 + \delta_i)(\delta^* - \mathbb{E}_{\delta_i}[\delta_i])$, *satisfies the conditions of Lemma 4(i).*

Proof. Obviously, $g'' > 0$; and $\mathbb{E}_{\delta_i}[g'(\delta_i)] = 0$, since $g'(\delta_i) = 2(\delta_i - \mathbb{E}_{\delta_i}[\delta_i])$. Furthermore, we have $\mathbb{E}_{\delta_i}[g(\delta_i)] = \mathbb{E}_{\delta_i}[\delta_i \cdot g'(\delta_i)]$ if and only if $(1 + \delta^*)^2 = \mathbb{E}_{\delta_i}[\delta_i^2] + 2\mathbb{E}_{\delta_i}[\delta_i] + 1$. The latter equality is satisfied for $\delta^* = -1 + \sqrt{\mathbb{E}_{\delta_i}[\delta_i^2] + 2\mathbb{E}_{\delta_i}[\delta_i] + 1} = -1 + \sqrt{\mathrm{Var}_{\delta_i}[\delta_i] + (1 + \mathbb{E}_{\delta_i}[\delta_i])^2}$. Since $\mathrm{Var}_{\delta_i}[\delta_i] > 0$ by assumption, $\delta^* > \mathbb{E}_{\delta_i}[\delta_i]$, such that $g(\delta_i) > 0$ for all $\delta_i \in (-1, 1)$. ■

By construction, the transfer scheme defined by Lemmas 4(v) and 5 is a gamification scheme.

6.3 Gamification in Arbitrary Social Networks

As utility is linear in transfers, Lemmas 4 and 5 suggest the following approach to deriving gamification schemes $(s_i^*)_{i\in I}$ for arbitrary social networks. Consider transfers

$$s_i^*(\delta) = g_i(\delta_i^S) - \delta_i^S \cdot g_i'(\delta_i^S) + \sum_{j\in I\setminus\{i\}} g_j'(\delta_j^S), \tag{20}$$

for appropriate functions $\delta_i^S : \Delta_i \to \mathbb{R}$, mapping an agent's $(n-1)$-dimensional social type to a one-dimensional signal (e.g., an effort level), and functions

$$g_i(\delta_i^S) = (\delta_i^S - \delta_i^*)^2 + 2(1+\delta_i^S)(\delta_i^* - \mathbb{E}[\delta_i^S]), \tag{21}$$

$$\delta_i^* = -1 + \sqrt{\mathrm{Var}_{\delta_i}[\delta_i^S] + (1+\mathbb{E}_{\delta_i}[\delta_i^S])^2}. \tag{22}$$

By Lemma 5, the functions $(g_i)_i$ each satisfy the conditions of Lemma 4(i), provided $\mathbb{E}_{\delta_i}[\delta_i^S]$ and $\mathrm{Var}_{\delta_i}[\delta_i^S]$ exist. How must the functions δ_i^S be chosen?

If each agent i is asked to report *some* real-valued signal δ_i^S, then, as I show below, i has the strictly dominant strategy to report his aggregate degree of altruism,

$$\delta_i^S(\delta_i) = \sum_{j\in I\setminus\{i\}} \delta_{ij}, \tag{23}$$

which I refer to as i's *degree of pro-sociality.*

Proposition 5 *Suppose the expected value and variance of each agent's degree of pro-sociality are common knowledge. Then the transfer scheme determined by equations (20) to (22) is a gamification scheme in terms of Definition 1.*

Proof. Under $(s_i^*)_{i\in I}$, each agent i reports a social type $\hat{\delta}_i$ so as to maximize his (interim-expected) utility, which is strategically equivalent to reporting some degree $\hat{\delta}_i^S$ of pro-sociality. His ex-post utility is given by

$$\begin{aligned}\sum_{j\in I} \delta_{ij} \cdot s_j^*(\hat{\delta}) &= g_i(\hat{\delta}_i^S) - \hat{\delta}_i^S \cdot g_i'(\hat{\delta}_i^S) + \sum_{j\neq i} g_j'(\hat{\delta}_j^S) + \sum_{j\neq i} \delta_{ij} \cdot \left[g_j(\hat{\delta}_j^S) - \hat{\delta}_j^S \cdot g_j'(\hat{\delta}_j^S)\right] \\ &\quad + \sum_{j\neq i} \delta_{ij} \cdot \sum_{\ell\neq j,i} g_\ell'(\hat{\delta}_\ell^S) + \left(\sum_{j\neq i} \delta_{ij}\right) \cdot g_i'(\hat{\delta}_i^S).\end{aligned}$$

Hence, i maximizes the function $g_i(\hat{\delta}_i^S) - \hat{\delta}_i^S \cdot g_i'(\hat{\delta}_i^S) + \delta_i^S \cdot g_i'(\hat{\delta}_i^S)$ over the choice of $\hat{\delta}_i^S$. Since g_i is strictly convex, i has the strictly dominant strategy to truthfully report $\hat{\delta}_i^S = \delta_i^S$. When substituting for $\mathbb{E}_{\delta_j}\left[g_j'(\delta_j^S)\right] = 0 = \mathbb{E}_{\delta_j}\left[g_j(\delta_j^S) - \delta_j^S \cdot g_j'(\delta_j^S)\right]$, due to Lemma 5, agent i's interim-expected utility satisfies $\sum_{j \in I} \delta_{ij} \cdot \mathbb{E}_{\delta_{-i}}[s_j^*(\delta)] = g_i(\delta_i^S)$, which is positive. By Lemma 5 again, ex-ante budget balance is satisfied, since $\mathbb{E}_\delta[s_i^*(\delta)] = 0$ for all i. ■

The distributive effects of gamification in the manner of Proposition 5 are similar to those discussed in the last subsection. From an ex-post perspective, the mediator prefers anti-social over pro-social types, while the opposite is true for agents.

7 Gamification and Network Autonomy

The previous section shows that when ex-post budget balance is replaced by ex-ante budget balance, then an ex-post efficient allocation rule can always be Bayesian implemented with a mechanism that is individually rational ex interim. I have interpreted ex-ante budget balance as the participation constraint of a mediator who subsidizes or sanctions agents based on the outcomes of a game that serves as an add-on to a social-preference robust SPC-mechanism. Consequently, that mediator is an outsider to the agents' network in that interpersonal preferences between her and network members are absent. In this section, I reintroduce ex-post budget balance by asking: Can the role of the mediator be taken by a network member? In other words: Can all social networks resolve all their allocation problems *autonomously*? By Theorem 2, the answer is 'no' for dyads; I show that 'yes' for larger networks.

Consider some agent $M \in I$ who is to serve as mediator. In line with the model setup of Section 2.1, this agent's utility from budget-balanced transfers $(s_i)_{i \in I}$ is given by

$$u_M = -\sum_{j \neq M} s_j + \sum_{j \neq M} \delta_{Mj} \cdot s_j, \tag{24}$$

where $\delta_{Mj} \in [\delta_{Mj}^{\min}, \delta_{Mj}^{\max}] \subset \left(\frac{-1}{n-1}, 1\right)$ for all $j \in I \setminus \{M\}$.

The results of Section 6.3 suggest the following approach to gamification in autonomous social networks: Define transfers $(s_i^{**}(\delta))_{i\in I}$ by

$$s_j^{**}(\delta) = -F + g_j(\delta_j^M) - \delta_j^M \cdot g_j'(\delta_j^M) + \sum_{\ell\neq j,M} g_\ell'(\delta_\ell^M), \quad \text{for } j \neq M, \text{ and} \tag{25}$$

$$s_M^{**}(\delta) = -\sum_{j\neq M} s_j^{**}(\delta), \tag{26}$$

for an appropriate fee $F > 0$ and functions $\delta_j^M : \Delta_j \to \mathbb{R}$ and

$$g_j(\delta_j^M) = (\delta_j^M - \delta_j^{**})^2 + 2(1+\delta_j^M)(\delta_j^{**} - \mathbb{E}[\delta_j^M]), \tag{27}$$

$$\delta_j^{**} = -1 + \sqrt{\mathrm{Var}_{\delta_j}[\delta_j^M] + (1+\mathbb{E}_{\delta_j}[\delta_j^M])^2}. \tag{28}$$

By Lemma 5, the functions $(g_j)_{j\neq M}$ each satisfy the conditions of Lemma 4(i), provided $\mathbb{E}_{\delta_j}[\delta_j^M]$ and $\mathrm{Var}_{\delta_j}[\delta_j^M]$ exist. How must the functions δ_j^M be chosen?

If each agent $j \neq M$ is asked to report *some* real-valued signal δ_j^M, then, as I show below, j has the strictly dominant strategy to report

$$\delta_j^M(\delta_j) = \frac{\sum_{\ell\neq j,M}(\delta_{j\ell} - \delta_{jM})}{\delta_{jj} - \delta_{jM}}, \tag{29}$$

which can be interpreted as j's relative marginal utility from a redistribution of M's money either to the others (who obtain equal shares) or to j himself; recall that $\delta_{jj} = 1$. I refer to δ_j^M as j's *relative spite towards* M, because δ_j^M decreases in j's altruism towards M and increases in j's aggregate altruism toward the others. Notice that the expected value and variance of δ_j^M exist, since $[\delta_{jM}^{\min}, \delta_{jM}^{\max}] \subset \left(\frac{-1}{n-1}, 1\right)$.

Proposition 6 *Suppose a social network of at least three agents contains an agent M such that the expected value and variance of every other j's relative spite towards M, given by (29), are common knowledge. Then the transfer scheme $(s_i^{**})_{i\in I}$ as given by equations (25) to (28), with a fee $F = \min_{j\neq M}\left(\delta_j^{**} - \mathbb{E}_{\delta_j}[\delta_j^M]\right)$, is ex-post budget-balanced, dominant-strategy incentive-compatible, and yields each agent $i \in I$ positive interim-expected utility: $\sum_{\ell\in I} \delta_{i\ell} \cdot \mathbb{E}_{\delta_{-i}}\left[s_\ell^{**}(\delta)\right] > 0$.*

Proof. Notice first that $F > 0$, since $\mathrm{Var}_{\delta_j}[\delta_j^M] > 0$ for all $j \neq M$, and that M has no strategic role to play. Under $(s_i^{**})_{i\in I}$, each agent $j \neq M$ reports a social type $\hat{\delta}_j$ so as

to maximize his (interim-expected) utility, which is strategically equivalent to reporting some degree $\hat{\delta}_j^M$ of relative spite towards M; his ex-post utility is given by

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

where only the second and last line are strategically relevant to j. Since $\delta_{jj} = 1 > \delta_{jM}$, agent j maximizes $g_j(\hat{\delta}_j^M) - \hat{\delta}_j^M \cdot g_j'(\hat{\delta}_j^M) + \delta_j^M \cdot g_j'(\hat{\delta}_j^M)$ over the choice of $\hat{\delta}_j^M$, where $\delta_j^M = \sum_{\ell \neq j,M} (\delta_{j\ell} - \delta_{jM}) / (\delta_{jj} - \delta_{jM})$. Since g_j is strictly convex, j has the strictly dominant strategy to truthfully report $\hat{\delta}_j^M = \delta_j^M$. When substituting for $\mathbb{E}_{\delta_\ell}\left[g_\ell'(\delta_\ell^M)\right] = 0 = \mathbb{E}_{\delta_\ell}\left[g_\ell(\delta_\ell^M) - \delta_\ell^M \cdot g_\ell'(\delta_\ell^M)\right]$, due to Lemma 5, agent j's interim-expected utility is obtained as

$$\begin{aligned}
\sum_{\ell \neq M} (\delta_{j\ell} - \delta_{jM}) \cdot \mathbb{E}_{\delta_{-j}}[s_\ell^{**}(\delta)] \;=\; & - F \cdot \sum_{\ell \neq M} (\delta_{j\ell} - \delta_{jM}) \\
& + (1 - \delta_{jM}) \cdot \left[g_j(\delta_j^M) - \delta_j^M \cdot g_j'(\delta_j^M) \right] \\
& + \left[\sum_{\ell \neq j,M} (\delta_{j\ell} - \delta_{jM}) \right] \cdot g_j'(\delta_j^M) \\
=\; & - F \cdot \sum_{\ell \neq M} (\delta_{j\ell} - \delta_{jM}) \\
& + (1 - \delta_{jM}) \cdot \left[g_j(\delta_j^M) - \delta_j^M \cdot g_j'(\delta_j^M) + \delta_j^M \cdot g_j'(\delta_j^M) \right] \\
=\; & (1 - \delta_{jM}) \cdot g_j(\delta_j^M) - F \cdot \sum_{\ell \neq M} (\delta_{j\ell} - \delta_{jM}) \\
=\; & (1 - \delta_{jM}) \cdot \left[g_j(\delta_j^M) - F \cdot (1 + \delta_j^M) \right],
\end{aligned}$$

where $1-\delta_{jM}>0$, and $g_j(\delta_j^M)>0$ for any δ_j^M. If $1+\delta_j^M \leq 0$, then j's interim-expected utility is obviously positive. Suppose, $1+\delta_j^M>0$; then F must satisfy $0<F<\frac{g_j(\delta_j^M)}{1+\delta_j^M}$. Since $g_j(\delta_j^M)$ is minimal at $\delta_j^M=\mathbb{E}_{\delta_j}[\delta_j^M]$, and $\delta_j^{**}>\mathbb{E}_{\delta_j}[\delta_j^M]$,

$$\begin{aligned}\frac{g_j(\delta_j^M)}{1+\delta_j^M} &\geq \frac{\left(\delta_j^{**}-\mathbb{E}_{\delta_j}[\delta_j^M]\right)^2+2\left(1+\delta_j^M\right)\left(\delta_j^{**}-\mathbb{E}_{\delta_j}[\delta_j^M]\right)}{1+\delta_j^M}\\ &> \delta_j^{**}-\mathbb{E}_{\delta_j}[\delta_j^M].\end{aligned}$$

Hence, choosing $F=\min_{j\neq M}\left(\delta_j^{**}-\mathbb{E}_{\delta_j}[\delta_j^M]\right)$ ensures that each agent j's interim-expected utility is positive. Finally, due to Lemma 5 again, M's interim-expected utility is positive:

$$\begin{aligned}\mathbb{E}_{\delta_{-M}}[u_M(\delta)] &= \sum_{j\neq M}(\delta_{Mj}-1)\cdot\left[-F+\mathbb{E}_{\delta_j}\left[g_j(\delta_j^M)-\delta_j^M\cdot g_j'(\delta_j^M)\right]+\sum_{\ell\neq j,M}\mathbb{E}_{\delta_\ell}[g_\ell'(\delta_\ell^M)]\right]\\ &= F\cdot\sum_{j\neq M}(1-\delta_{Mj}).\end{aligned}$$

This completes the proof.[32] ■

With this result at hand, gamification can be deployed to ensure the incentive-feasibility of ex-post efficiency.

Theorem 5 (Gamification in Autonomous Networks) *Suppose a social network of at least three agents contains an agent M such that the expected value and variance of every other j's relative spite towards M, given by (29), are common knowledge. Then, for any collection of reservation utilities, this network can resolve any allocation problem with a mechanism that is Bayesian incentive-compatible, ex-interim individually rational, and ex-post Pareto-efficient (in particular, ex-post budget-balanced). The result extends to allocation problems with quasi-linear private payoffs.*

[32] As suggested by Lemma 2, the functions $p_i:\Delta_i\to\mathbb{R}$ of Theorem 1 are simply given by $p_j(\delta_j)=(1-\delta_{jM})\cdot\left[g_j(\delta_j^M)-F\cdot(1+\delta_j^M)\right]$ for $j\neq M$ and $p_M(\delta_M)=F\cdot\sum_{j\neq M}(1-\delta_{Mj})$, such that $\partial p_j(\delta_j)/\partial\delta_{j\ell}=\mathbb{E}_{\delta_{-j}}[s_\ell^{**}(\delta)]$ and $\partial p_M(\delta_M)/\partial\delta_{Mj}=\mathbb{E}_{\delta_{-M}}[s_j^{**}(\delta)]$. Each p_i is smooth and convex, and its transform $p_i(\mathbf{1}-\delta_i)$ is homogeneous of degree one. If $F=\min_{j\neq M}\left(\delta_j^{**}-\mathbb{E}_{\delta_j}[\delta_j^M]\right)$, then $p_i(\delta_i)>0$ for each $i\in I$.

Proof. A feasible mechanism is specified by the social-choice function (k^*, T^{**}) consisting of an ex-post efficient allocation function $k^* : \Theta \to K$, $k^*(\theta) \in \arg\max_{k \in K} \sum_{i \in I} \pi_i(k \,|\, \theta_i)$, and the transfer scheme $T^{**} = (t_i^{**})_{i \in I} : \Theta \times \Delta \to \mathbb{R}^n$,

$$(30) \qquad t_i^{**}(\hat{\theta}, \hat{\delta}) \;=\; \sum_{\ell \in I \setminus \{i\}} \Big[\mathbb{E}_{\theta_{-i}} \big[\pi_\ell(k^*(\hat{\theta}_i, \theta_{-i}) \,|\, \theta_\ell) \big] - \mathbb{E}_{\theta_{-\ell}} \big[\pi_i(k^*(\hat{\theta}_\ell, \theta_{-\ell}) \,|\, \theta_i) \big] \Big] + s_i^{**}(\hat{\delta}),$$

with $(s_i^{**})_{i \in I}$ as defined by equations (25) to (28) for $F = \min_{j \neq M} \big(\delta_j^{**} - \mathbb{E}_{\delta_j}[\delta_j^M] \big)$. This mechanism is ex-post Pareto-efficient and Bayesian incentive-compatible, due to Theorem 1, Proposition 4, and Proposition 6. When scaling the $(s_i^{**})_{i \in I}$ with the same, sufficiently large $\alpha > 0$, each agent's participation constraint will be satisfied, due to Proposition 6. ■

Theorem 5 gives proof to the Coase theorem, in its mechanism design interpretation, for social networks of at least three agents.

8 Limitations to Gamification

There are some limitations to the practical scope of gamification.

The agent who is to serve as mediator, whether she belongs or is an outsider to the network, must be able to *commit* to the ex-post outcomes of gamification. This commitment might be enforced by an external institution. In the casino example (Corollary 4), for instance, the state can enforce the casino's commitment as long as gambling is legal. In this respect, the ability to commit can be interpreted as being part of the implicit assumption made throughout this paper: that property rights are well-defined and contracts are enforceable.

I have assumed that agents are not exposed to any *endowment constraints*, such that the redistributive effects of gamification can be scaled up to an arbitrary extent. I show in Daske (2016) for the case of two agents with i.i.d social types (as considered here in Section 6.2) how gamification can be deployed to incentivize the *production* of those endowments (the market returns from production) that are then to be distributed among the mediator (employer) and the agents (workers). That is, gamification schemes can take the form of labor contracts. These contracts payoff-equivalently provide agents with

relative or team performance incentives, both of which result in a hawk-dove like game between co-workers.

In order to cope with Wilson's (1987) call for avoiding *common-knowledge assumptions*, I have derived gamification schemes that merely require common knowledge of the expected values and variances of specific network determinants; those are agents' degrees of *relative spite* towards an *inside* mediator or their *pro-sociality* if the mediator is an *outsider* to the network (in which case interpersonal preferences between the mediator and network members are absent). Realistically, these assumptions are not met in large networks, if any. Furthermore, the computation of agents' relative spite from their interpersonal preferences actually requires common knowledge of their entire social-type distributions. By contrast, the expected values and variances of agents' pro-sociality are immediate from the expected values and variances of each of their interpersonal preferences, which might be easier to observe. From a practical point of view, gamification might thus be easier to deploy with the help of an outside mediator.

Risk aversion, finally, is another serious issue, for taking risks is essential to gamification as I have defined it.[33] The model conceived in this paper assumes that agents are risk-neutral with regard to 'money'. While there seems to be no reason to believe that gamification could deliver risk-averse agents with *arbitrary* interim-expected Pareto improvements upon a non-gamified status quo, I show in Appendix C that gamification with an outside mediator does always yield a slight interim-expected Pareto improvement; the mediator can even be privately informed about her social preferences towards the network. The observation that an interim-expected social-preference related *mental compensation* can outweigh an interim-expected *material loss*: it persists.

[33]The mediator takes the risk of ending up with dovish players who would collaborate against her. Players take the risk of ending up with hawkish opponents who make competition more severe and undermine collaboration.

A Omitted Proofs

A.1 Proof of Lemma 1

Having required weak budget balance, Pareto efficiency implies strict budget balance: Suppose $\sum_{i\in I} t_i = -\epsilon$ for some $\epsilon > 0$; then a Pareto improvement can be achieved through transfers $(t_i + \epsilon/n)_{i\in I}$, since $\sum_{j\in I} \delta_{ij} > 0$. Provided transfers are budget-balanced, the *payoff*-possibility set can be written as

$$\left\{ \begin{pmatrix} \vdots \\ \pi_i(k \,|\, \theta_i) + t_i \\ \vdots \\ \pi_n(k \,|\, \theta_n) - \sum_i t_i \end{pmatrix} \right\}_{\substack{(t_i)_{i<n}\in\mathbb{R}^{n-1}, \\ k\in K}} = \left\{ \frac{\sigma(k)}{n} \cdot \begin{pmatrix} \vdots \\ 1 \\ \vdots \\ 1 \end{pmatrix} + \begin{pmatrix} \vdots \\ t'_i \\ \vdots \\ -\sum_i t'_i \end{pmatrix} \right\}_{\substack{(t'_i)_{i<n}\in\mathbb{R}^{n-1}, \\ k\in K}},$$

where $\sigma(k) = \sum_{i\in I} \pi_i(k \,|\, \theta_i)$: For any pair $(k, (t_i)_i)$, with $t_n = -\sum_{i=1}^{n-1} t_i$, one can choose $t'_i = -\frac{\sigma(k)}{n} + \pi_i(k \,|\, \theta_i) + t_i$ for $i \leq n-1$, such that $-\sum_{i=1}^{n-1} t'_i = \frac{n-1}{n}\sigma(k) - \sum_{i=1}^{n-1} \pi_i(k \,|\, \theta_i) + t_n = -\frac{\sigma(k)}{n} + \pi_n(k \,|\, \theta_n) + t_n$. The *utility*-possibility set thus satisfies

$$\left\{ \begin{pmatrix} \vdots \\ u_i \\ \vdots \end{pmatrix} \right\}_{\substack{(t_i)_{i<n}\in\mathbb{R}^{n-1}, \\ t_n = -\sum_i t_i, \\ k\in K}} = \left\{ \frac{\sigma(k)}{n} \cdot \begin{pmatrix} \vdots \\ \sum_{j\in I} \delta_{ij} \\ \vdots \end{pmatrix} + \begin{pmatrix} \vdots \\ \sum_{j\in I} \delta_{ij} t'_j \\ \vdots \end{pmatrix} \right\}_{\substack{(t'_i)_{i<n}\in\mathbb{R}^{n-1}, \\ t'_n = -\sum_i t'_i, \\ k\in K}}.$$

Since $\sum_{j\in I} \delta_{ij} > 0$ for all i, every increase in $\sigma(k)$ yields a Pareto improvement (through an adequate change in budget-balanced transfers.)

Let $|\delta_{ij}| < 1/(2n-3)$ for all i and all $j \neq i$, and suppose there exist budget-balanced transfer schemes T and T' such that T' Pareto improves upon T. Then there exists an agent i and types (θ, δ) such that i makes the greatest monetary loss: For $(d_j)_j = (t'_j - t_j)_j$, let $i \in \arg\min_j d_j$, where $\min_j d_j = -\epsilon < 0$. Indicate by j_- those agents other than i who would also lose money when switching from T to T'; i.e., $d_{j_-} \in [-\epsilon, 0)$ for all j_-. Then i would realize the maximum utility feasible under T' if his favorite agent,

$j^* \in \arg\max_{j \neq i} \delta_{ij}$, obtained all redistribution mass: $d_{j^*} = \epsilon - \sum_{j_-} d_{j_-} > 0$. Therefore, i's change in utility when switching from T to T' satisfies

$$\begin{aligned}
u_i' - u_i &\leq -\epsilon + \sum_{j_-} \delta_{ij_-} \cdot d_{j_-} + \delta_{ij^*} \cdot \left(\epsilon - \sum_{j_-} d_{j_-} \right) \\
&< -\epsilon + \sum_{j_-} \frac{-1}{2n-3} \cdot (-\epsilon) + \delta_{ij^*} \cdot (1 + |\{j_-\}|) \cdot \epsilon \\
&< \left(-1 + \frac{2|\{j_-\}| + 1}{2n-3} \right) \cdot \epsilon \\
&\leq \left(-1 + \frac{2(n-2)+1}{2n-3} \right) \cdot \epsilon \\
&= 0.
\end{aligned}$$

Hence, i is worse off under T'; a contradiction.

A.2 Proof of Lemma 2

Suppose (k, T), with $k : \Theta \to K$ and $T : \Theta \times \Delta \to \mathbb{R}^n$, is incentive-compatible. Then the following must hold for all $\theta_i, \hat{\theta}_i \in \Theta_i$ and all $\delta_i, \hat{\delta}_i \in \Delta_i$:

$$U_i(\theta_i, \delta_i) \geq U_i(\hat{\theta}_i, \delta_i \,|\, \theta_i, \delta_i) = U_i(\hat{\theta}_i, \delta_i) + (\theta_i - \hat{\theta}_i) \cdot \bar{v}_i(\hat{\theta}_i), \tag{31}$$

$$U_i(\hat{\theta}_i, \delta_i) \geq U_i(\theta_i, \delta_i \,|\, \hat{\theta}_i, \delta_i) = U_i(\theta_i, \delta_i) + (\hat{\theta}_i - \theta_i) \cdot \bar{v}_i(\theta_i), \tag{32}$$

$$\begin{aligned}
U_i(\theta_i, \delta_i) &\geq U_i(\theta_i, \hat{\delta}_{ij}, \delta_i^{-j} \,|\, \theta_i, \delta_i) \\
&= U_i(\theta_i, \hat{\delta}_{ij}, \delta_i^{-j}) + (\delta_{ij} - \hat{\delta}_{ij}) \cdot \left[\bar{\pi}_{ij}(\theta_i) + \bar{t}_{ij}(\theta_i, \hat{\delta}_i, \delta_i^{-j}) \right],
\end{aligned} \tag{33}$$

$$\begin{aligned}
U_i(\theta_i, \hat{\delta}_{ij}, \delta_i^{-j}) &\geq U_i(\theta_i, \delta_i \,|\, \theta_i, \hat{\delta}_i, \delta_i^{-j}) \\
&= U_i(\theta_i, \delta_i) + (\hat{\delta}_{ij} - \delta_{ij}) \cdot \left[\bar{\pi}_{ij}(\theta_i) + \bar{t}_{ij}(\theta_i, \delta_i) \right].
\end{aligned} \tag{34}$$

Without loss of generality, let $\hat{\theta}_i > \theta_i$. Then (31) and (32) imply that

$$\bar{v}_i(\hat{\theta}_i) \geq \frac{U_i(\hat{\theta}_i, \delta_i) - U_i(\theta_i, \delta_i)}{\hat{\theta}_i - \theta_i} \geq \bar{v}_i(\theta_i). \tag{35}$$

Hence, $\bar{v}_i(\theta_i)$ is non-decreasing in θ_i and therefore, by Lebesgue's Theorem, partially differentiable in θ_i *almost everywhere*. Letting $\hat{\theta}_i$ approach θ_i in (35) implies that $U_i(\theta_i, \delta_i)$ is partially differentiable in θ_i *almost everywhere*, with $\partial U_i(\theta_i, \delta_i)/\partial \theta_i = \bar{v}_i(\theta_i)$. Monotonicity also ensures that $\bar{v}_i(\theta_i)$ is bounded around any $\theta_i \in \Theta_i$, such that letting $\hat{\theta}_i$ approach θ_i

in (35) implies that $U_i(\theta_i, \delta_i)$ is continuous on Θ_i. Hence, $U_i(\theta_i, \delta_i)$ satisfies condition (1) for some function $p_i : \Delta_i \to \mathbb{R}$. Similarly, let $\hat{\delta}_i > \delta_i$. By (33) and (34),

$$\begin{aligned} \bar{\pi}_{ij}(\theta_i) + \bar{t}_{ij}(\theta_i, \hat{\delta}_{ij}, \delta_i^{-j}) &\geq \frac{U_i(\theta_i, \hat{\delta}_{ij}, \delta_i^{-j}) - U_i(\theta_i, \delta_i)}{\hat{\delta}_{ij} - \delta_{ij}} \\ &\geq \bar{\pi}_{ij}(\theta_i) + \bar{t}_{ij}(\theta_i, \delta_i). \end{aligned} \tag{36}$$

Hence, $\bar{t}_{ij}(\theta_i, \delta_i)$ is non-decreasing in δ_{ij} for each $j \neq i$. By the same reasoning as above, letting $\hat{\delta}_{ij}$ approach δ_{ij} in (36) implies that (2) must hold for some function $q_{ij} : \Theta_i \times \Pi_{\ell \in I \setminus \{i,j\}} \Delta_{i\ell} \to \mathbb{R}$. By comparison of (1) and (2), p_i and q_{ij} are partially differentiable in δ_i and θ_i, respectively.

A.3 Proof of Proposition 1

Suppose (k, T), with $k : \Theta \to K$ and $T : \Theta \times \Delta \to \mathbb{R}^n$, is incentive-compatible. Condition (i) is necessary due to Lemma 2(i). By Lemma 2(iii), there exists a partially differentiable function $p_i : \Delta_i \to \mathbb{R}$ such that

$$\frac{\partial p_i(\delta_i)}{\partial \delta_{ij}} = \bar{\pi}_{ij}(\theta_i) + \bar{t}_{ij}(\theta_i, \delta_i), \tag{37}$$

which yields identity (4). Due to identities (1) and (4),

$$\begin{aligned} p_i(\delta_i) + \int_{\theta_i^{\min}}^{\theta_i} \bar{v}_i(r)\, dr &= \bar{\pi}_{ii}(\theta_i) + \bar{t}_{ii}(\theta_i, \delta_i) + \sum_{j \in I \setminus \{i\}} \delta_{ij} \cdot \left[\bar{\pi}_{ij}(\hat{\theta}_i) + \bar{t}_{ij}(\hat{\theta}_i, \hat{\delta}_i) \right] \\ &= \bar{\pi}_{ii}(\theta_i) + \bar{t}_{ii}(\theta_i, \delta_i) + \nabla p_i(\delta_i) \cdot \delta_i, \end{aligned}$$

which yields identity (5). Due to identities (4) and (5),

$$\begin{aligned} U_i(\hat{\theta}_i, \hat{\delta}_i \,|\, \theta_i, \delta_i) &= \theta_i \cdot \bar{v}_i(\hat{\theta}_i) + \bar{w}_i(\hat{\theta}_i) + \bar{t}_{ii}(\hat{\theta}_i, \hat{\delta}_i) + \sum_{j \in I \setminus \{i\}} \delta_{ij} \cdot \left[\bar{\pi}_{ij}(\hat{\theta}_i) + \bar{t}_{ij}(\hat{\theta}_i, \hat{\delta}_i) \right] \\ &= (\theta_i - \hat{\theta}_i) \cdot \bar{v}_i(\hat{\theta}_i) + \int_{\theta_i^{\min}}^{\hat{\theta}_i} \bar{v}_i(r)\, dr + p_i(\hat{\delta}_i) + \nabla p_i(\hat{\delta}_i) \cdot (\delta_i - \hat{\delta}_i), \end{aligned}$$

implying that interim-expected utility is additively separable in an agent's payoff type and social type. As $\bar{v}_i(\hat{\theta}_i)$ is non-decreasing by condition *(i)*, $U_i(\theta_i, \hat{\delta}_i \,|\, \theta_i, \delta_i) \geq U_i(\hat{\theta}_i, \hat{\delta}_i \,|\, \theta_i, \delta_i)$ for all $\theta_i, \hat{\theta}_i, \delta_i, \hat{\delta}_i$. Hence, with respect to *payoff* types, conditions (i), (4), and (5) are

		s_{-i}			
		D		H	
s_i	D	w	w	x	y
	H	y	x	z	z

Figure 4: *The payoff scheme $(s_i)_{i=1,2}$ underlying the strategic game that a mediator offers agents $i \in \{1,2\}$ to play.*

necessary and sufficient for the incentive compatibility of (k, T). On the other hand, due to (4) and (5), $U_i(\theta_i, \delta_i) \geq U_i(\hat{\theta}_i, \hat{\delta}_i \,|\, \theta_i, \delta_i)$ for all $\theta_i, \delta_i, \hat{\delta}_i$ if and only if $p_i(\hat{\delta}_i) + \nabla p_i(\hat{\delta}_i) \cdot (\delta_i - \hat{\delta}_i) \leq p_i(\delta_i)$ for all $\delta_i, \hat{\delta}_i$. Hence, with respect to *social* types, condition (ii) is necessary and sufficient for the incentive compatibility of (k, T).

A.4 Proof of Lemma 3

If $\mathbb{E}_{\theta_{-j},\delta_{-j}}[s_i(\theta, \delta)]$ is constant for all $i, j \in I$, then $\alpha_i = \sum_{j \in I} \mathbb{E}_{\theta_{-j},\delta_{-j}}[s_i(\theta, \delta)]$ is a constant. Taking expectations over $(\theta_{-\ell}, \delta_{-\ell})$ yields $\alpha_i = \mathbb{E}_{\theta_{-\ell},\delta_{-\ell}}[s_i(\theta, \delta)] + (n-1) \cdot \mathbb{E}_{\theta,\delta}[s_i(\theta, \delta)]$. Hence, $\mathbb{E}_{\theta_{-\ell},\delta_{-\ell}}[s_i(\theta, \delta)] = S_i$ for all $\ell \in I$, where $S_i = \alpha_i - (n-1) \cdot \mathbb{E}_{\theta,\delta}[s_i(\theta, \delta)]$ is constant.

B Gamification with Discrete Social-Type Sets

Consider a mediator M and two agents $i \in \{1, 2\}$ whose social types are i.i.d., taking the value $\delta_i = \delta_H$ with probability $p \in (0, 1)$ and the value $\delta_i = \delta_D$ with probability $1 - p$, where $-1 < \delta_H < \delta_D < 1$. The mediator offers agents to play a game the material outcomes of which are determined by the payoff scheme $(s_i)_{i=1,2}$ depicted in Figure 4. The parameters $w, x, y, z \in \mathbb{R}$ are supposed to be such that the following conditions hold simultaneously: agents reveal their social types in dominant-strategy equilibrium, with type δ_D playing D (condition C.1) and type δ_H playing H (C.2); interim-expected utilities from play are positive for both types (C.3); and M's payoff $s_M = -s_1 - s_2$ is zero ex ante (C.4). Formally,

$$
\begin{aligned}
\text{(C.1)} \qquad 0 &< (w+\delta_D w)-(y+\delta_D x),\\
0 &< (x+\delta_D y)-(z+\delta_D z);\\
\text{(C.2)} \qquad 0 &< (y+\delta_H x)-(w+\delta_H w),\\
0 &< (z+\delta_H z)-(x+\delta_H y);\\
\text{(C.3)} \qquad 0 &< (1-p)\cdot(w+\delta_D w)+p\cdot(x+\delta_D y),\\
0 &< (1-p)\cdot(y+\delta_H x)+p\cdot(z+\delta_H z);\\
\text{(C.4)} \qquad 0 &= -(1-p)^2\cdot(w+w)-2\cdot p\cdot(1-p)\cdot(x+y)-p^2\cdot(z+z).
\end{aligned}
$$

When substituting for $z=-\frac{1-p}{p}\cdot(x+y)-\left(\frac{1-p}{p}\right)^2\cdot w$, as implied by condition (C.4), and letting $x=(1-\frac{1-p}{p})\cdot w-y$, the system (C.1)–(C.3) reduces to only two conditions:

$$
\text{(38)} \qquad (1-\delta_D)\cdot y \;<\; \left(1+\frac{1-p}{p}\cdot\delta_D\right)\cdot w,
$$

$$
\text{(39)} \qquad \left(1+\frac{1-p}{p}\cdot\delta_H\right)\cdot w \;<\; (1-\delta_H)\cdot y.
$$

Observe that the term $\left(1+\frac{1-p}{p}\cdot\delta\right)/(1-\delta)$ strictly increases in $\delta\in(-1,1)$. Since $\delta_H<\delta_D$, conditions (38) and (39) are thus satisfied jointly if and only if $w>0$ and

$$
\frac{1+\frac{1-p}{p}\cdot\delta_H}{1-\delta_H}\cdot w \;<\; y \;<\; \frac{1+\frac{1-p}{p}\cdot\delta_D}{1-\delta_D}\cdot w.
$$

Hence, the system (C.1)–(C.4) is satisfied when letting $w>0$,

$$
\begin{aligned}
z &= -\frac{1-p}{p}\cdot w,\\
y &= \frac{1}{2}\cdot\left[\frac{1+\frac{1-p}{p}\cdot\delta_H}{1-\delta_H}+\frac{1+\frac{1-p}{p}\cdot\delta_D}{1-\delta_D}\right]\cdot w,\\
x &= \left(1-\frac{1-p}{p}\right)\cdot w-y.
\end{aligned}
$$

The example of Section 6.1 is obtained when letting $\delta_H=0$, $\delta_D=1/2$, $p=1/2$, and $w=4\hat{x}>0$, such that $z=-w=-4\hat{x}$, $y=2w=8\hat{x}$, and $x=-y=-8\hat{x}$.

C Gamification and Risk Aversion

This Appendix addresses the question of whether gamification in the manner of Section 6 is feasible if mediator and agents are risk-averse. The mediator may exhibit social preferences towards the agents.

Consider a group $I = \{1, ..., n\}$ of $n \geq 2$ agents in which agent i's utility from transfers $(s_j^{**})_{j \in I}$ is given by $u_i = \sum_{j \in I} \delta_{ij} \cdot v_j(s_j^{**})$, for continuously differentiable functions $v_j : \mathbb{R} \to \mathbb{R}$ satisfying $v_j' > 0$ and $v_j(0) = 0$ for all j. Again, $\delta_{ii} = 1$, and $\delta_{ij} \in [\delta_{ij}^{\min}, \delta_{ij}^{\max}] \subset \left(\frac{-1}{n-1}, 1\right)$ for all $j \neq i$. Consider also a mediator M, an outsider to the social network in the sense that no agent cares about her, whose utility from $(s_j^{**})_{j \in I}$ is given by

$$u_M = v_M \left(- \sum_{j \in I} s_j^{**} \right) + \sum_{j \in I} \delta_{Mj} \cdot v_j(s_j^{**}),$$

where $\delta_{Mj} \in [\delta_{Mj}^{\min}, \delta_{Mj}^{\max}] \subset \left(\frac{-1}{n-1}, 1\right)$ for all $j \in I$, and where $v_M : \mathbb{R} \to \mathbb{R}$ is continuously differentiable and satisfies $v_M' > 0$ and $v_M(0) = 0$. That is, transfers are assumed to be budget-balanced among mediator and agents, while M's valuation of a transfer $s_M^{**} = -\sum_{j \in I} s_j^{**}$ to herself is just $v_M(s_M^{**})$. Suppose further that agents' and the mediator's social types are private information but that the expected value and variance of each agent's degree of pro-sociality, $\delta_i^S = \sum_{\ell \in I \setminus \{i\}} \delta_{i\ell}$, are common knowledge. Define

$$(s_j^{**})_{j \in I} = \left[v_j^{-1} \left(\alpha \cdot \left(s_j^* - \epsilon \right) \right) \right]_{j \in I},$$

with the inverse v_j^{-1} of v_j, the gamification scheme $(s_j^*)_{j \in I}$ of Proposition 5, and constants $\alpha \geq 0$ and $\epsilon > 0$. By Proposition 5, $(s_j^{**})_{j \in I}$ is dominant-strategy incentive-compatible in agents' social types. As $\sum_{j \in I} \delta_{ij} \cdot \mathbb{E}_{\delta_{-i}} \left[s_j^*(\delta) \right] > 0$, transfers $(s_j^{**})_{j \in I}$ yield each agent i a positive interim-expected utility level as long as

$$\epsilon \quad < \quad \frac{\sum_{j \in I} \delta_{ij} \cdot \mathbb{E}_{\delta_{-i}} \left[s_j^*(\delta) \right]}{\sum_{j \in I} \delta_{ij}}.$$

Recall that $\mathbb{E}_\delta\big[s_j^*(\delta)\big] = 0$ for all j, where $\delta = (\delta_i)_{i\in I} = \delta_{-M}$. The mediator's interim-expected utility from $(s_j^{**})_{j\in I}$, as a function of α, is thus given by

$$\mathbb{E}_{\delta_{-M}}\big[u_M\big](\alpha) \;=\; \mathbb{E}_{\delta_{-M}}\left[v_M\left(-\sum_{j\in I} v_j^{-1}\left(\alpha\cdot\left(s_j^*(\delta_{-M})-\epsilon\right)\right)\right)\right] - \alpha\cdot\epsilon\cdot\left(\sum_{j\in I}\delta_{Mj}\right).$$

Observe that $\mathbb{E}_{\delta_{-M}}\big[u_M\big](0) = 0$ and

$$\begin{aligned}
&\frac{\partial}{\partial\alpha}\mathbb{E}_{\delta_{-M}}\big[u_M\big](\alpha)\Big|_{\alpha=0}\\
=\;& \mathbb{E}_{\delta_{-M}}\left[-v_M'\left(-\sum_{j\in I} v_j^{-1}\left(\alpha\cdot\left(s_j^*(\delta_{-M})-\epsilon\right)\right)\right)\cdot\sum_{j\in I}\frac{s_j^*(\delta_{-M})-\epsilon}{v_j'\left(v_j^{-1}\left(\alpha\cdot\left(s_j^*(\delta_{-M})-\epsilon\right)\right)\right)}\right]\Bigg|_{\alpha=0}\\
&\quad -\epsilon\cdot\left(\sum_{j\in I}\delta_{Mj}\right)\\
=\;& -v_M'(0)\cdot\mathbb{E}_{\delta_{-M}}\left[\sum_{j\in I}\frac{s_j^*(\delta_{-M})-\epsilon}{v_j'(0)}\right] - \epsilon\cdot\left(\sum_{j\in I}\delta_{Mj}\right)\\
=\;& \epsilon\cdot\sum_{j\in I}\left(\frac{v_M'(0)}{v_j'(0)} - \delta_{Mj}\right),
\end{aligned}$$

which is positive if M is not too altruistic towards the agents. In this case, participation in the game implementing the transfers scheme $(s_j^{**})_{j\in I}$ yields agents and the mediator an interim-expected Pareto improvement. Hence, gamification is also feasible if agents and the mediator are risk-averse, regardless how their valuations of money are shaped.

References

Arrow, Kenneth J. 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.

Bartling, Björn and Nick Netzer. 2016. "An externality-robust auction: Theory and experimental evidence." *Games and Economic Behavior* 97:186–204.

Becker, Gary S and Kevin M Murphy. 1988. "A theory of rational addiction." *Journal of Political Economy* 96 (4):675–700.

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.

Bierbrauer, Felix, Axel Ockenfels, Andreas Pollak, and Désirée Rückert. 2017. "Robust mechanism design and social preferences." *Journal of Public Economics* 149:59–80.

Börgers, Tilman and Peter Norman. 2009. "A note on budget balance under interim participation constraints: The case of independent types." *Economic Theory* 39 (3):477–489.

Bourlès, Renaud, Yann Bramoullé, and Eduardo Perez-Richet. 2017. "Altruism in networks." *Econometrica* 85 (2):675–689.

Bruhin, Adrian, Ernst Fehr, and Daniel Schunk. In press. "The many faces of human sociality: Uncovering the distribution and stability of social preferences." *Journal of the European Economic Association* .

Carlsson, Hans and Eric Van Damme. 1993. "Global games and equilibrium selection." *Econometrica* 61 (5):989–1018.

Coase, Ronald H. 1960. "The problem of social cost." *Journal of Law and Economics* 3:1–44.

Cramton, Peter, Robert Gibbons, and Paul Klemperer. 1987. "Dissolving a partnership efficiently." *Econometrica* 55 (3):615–632.

Daske, Thomas. 2016. "Pooling hawks and doves: Interim-efficient labor contracts for other-regarding agents." *Mimeo.* Available under https://www.econstor.eu/handle/10419/145951.

———. 2017. "Externality Assessments, Welfare Judgments, and Mechanism Design." *Mimeo.* Available under https://www.econstor.eu/handle/10419/172494.

d'Aspremont, Claude and Louis-André Gérard-Varet. 1979. "Incentives and incomplete information." *Journal of Public Economics* 11 (1):25–45.

Eisenhuth, Roland, Dermot Murphy, and Andreas Neuhierl. 2018. "Casino game markets." In *Handbook of Behavioral Industrial Organization*. Edward Elgar Publishing, 257–290.

Farrell, Joseph. 1987. "Information and the Coase theorem." *Journal of Economic Perspectives* 1 (2):113–129.

Franke, Jörg and Wolfgang Leininger. 2018. "Lotteries and Lindahl prices in public good provision." *Journal of Public Economic Theory* 20 (6):840–848.

Jehiel, Philippe, Moritz Meyer-ter Vehn, Benny Moldovanu, and William R 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.

Kucuksenel, Serkan. 2012. "Behavioral mechanism design." *Journal of Public Economic Theory* 14 (5):767–789.

Mailath, George J and Andrew Postlewaite. 1990. "Asymmetric information bargaining problems with many agents." *Review of Economic Studies* 57 (3):351–367.

Marfels, Christian. 2001. "Is gambling rational? The utility aspect of gambling." *Gaming Law Review* 5 (5):459–466.

McAfee, R Preston. 1991. "Efficient allocation with continuous quantities." *Journal of Economic Theory* 53 (1):51–74.

Mobilia, Pamela. 1993. "Gambling as a rational addiction." *Journal of Gambling Studies* 9 (2):121–151.

Myerson, Roger B. 1979. "Incentive compatibility and the bargaining problem." *Econometrica* 47 (1):61–73.

———. 1981. "Optimal auction design." *Mathematics of Operations Research* 6 (1):58–73.

Myerson, Roger B and Mark A Satterthwaite. 1983. "Efficient mechanisms for bilateral trading." *Journal of Economic Theory* 29 (2):265–281.

Peloza, John and Derek N Hassay. 2007. "A typology of charity support behaviors: Toward a holistic view of helping." *Journal of Nonprofit & Public Sector Marketing* 17 (1-2):135–151.

Rabin, Matthew. 1993. "Incorporating fairness into game theory and economics." *American Economic Review* 83 (5):1281–1302.

Robson, Karen, Kirk Plangger, Jan H Kietzmann, Ian McCarthy, and Leyland Pitt. 2015. "Is it all a game? Understanding the principles of gamification." *Business Horizons* 58 (4):411–420.

Segal, Ilya and Michael D Whinston. 2011. "A simple status quo that ensures participation (with application to efficient bargaining)." *Theoretical Economics* 6 (1):109–125.

———. 2016. "Property rights and the efficiency of bargaining." *Journal of the European Economic Association* 14 (6):1287–1328.

Webber, Daniel. 2004. "Understanding charity fundraising events." *International Journal of Nonprofit and Voluntary Sector Marketing* 9 (2):122–134.

Williams, Steven R. 1999. "A characterization of efficient, Bayesian incentive compatible mechanisms." *Economic Theory* 14 (1):155–180.

Wilson, Robert. 1987. "Game-theoretic analyses of trading processes." In *Advances in Economic Theory: Fifth World Congress*, edited by T. Bewley, chap. 2. Cambridge U.K.: Cambridge University Press, 33–70.

Zik, Boaz. 2019. "Ex-post implementation with social preferences." *Mimeo.* Available under http://boazzik.com/pdf/pdf3.pdf.