

Daske, Thomas

Working Paper

Efficient Incentives in Social Networks: Gamification and the Coase Theorem

Suggested Citation: Daske, Thomas (2020) : 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/10419/222527>

Standard-Nutzungsbedingungen:

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

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

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

Terms of use:

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

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

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

Efficient Incentives in Social Networks: Gamification and the Coase Theorem

Thomas Daske*

This Version: July 28, 2020

First Version: February 23, 2019

This study explores mechanism design for networks of interpersonal relationships. Agents' social (more or less altruistic or spiteful) preferences and private payoffs are all subject to asymmetric information. Remarkably, the asymmetry of information about agents' social preferences can be operationalized to satisfy agents' participation constraints. The main result is a constructive proof of the Coase theorem, in its typical mechanism-design interpretation, for networks of *at least three* agents: If endowments are sufficiently large, any such network can resolve any given allocation problem with an ex-post budget-balanced mechanism that is Bayesian incentive-compatible, interim individually rational, and ex-post Pareto-efficient. The endogenously derived solution concept is interpreted as *gamification*: Resolve the agents' allocation problem with an efficient social-preference robust mechanism; attract agents' participation by complementing this mechanism with a budget-balanced game that operates on their social preferences and provides them with a platform to live out their propensities to cooperate or compete.

JEL Classification: C72; C78; D62; D82

Keywords: Mechanism design; social preferences; gamification; joyful games;
Coase theorem

*TUM School of Management, Technical University of Munich, Arcisstr. 21, 80333 Munich, Germany. Fax: +49-89-289-25701. Phone: +49-89-289-25707. Email: thomas.daske@tum.de.

For their helpful comments and critical remarks, I want to thank Christian Feilcke, Aart Gerritsen, Christoph Gschnaidtner, Julian Hackinger, Kai Konrad, Christoph March, Salmai Qari, Matthias Regier, Felix Reuss, Marco Sahm, Klaus Schmidt, Johannes Schneider, Roland Strausz, and Robert von Weizsäcker. I am also grateful for inspiring comments by participants of the European Winter Meeting of the Econometric Society in Milan, the World Congress of the Game Theory Society in Maastricht, the Annual Meeting of the Association for Public Economic Theory in Strasbourg, the Annual Congress of the International Institute of Public Finance in Glasgow, the European Meeting of the Econometric Society in Manchester, the Annual Congress of the German Economic Association in Leipzig, the virtual Econometric Society World Congress, and seminars at the Universities of Augsburg, Berlin, and Leipzig. Financial support by the Max Planck Society is gratefully acknowledged. First version available under: <http://hdl.handle.net/10419/193148>.

“[T]he Coase theorem is much more than simplistic overoptimism or circular reasoning. Organized markets in standardized commodities are not the only institutions for economists to analyze. People can be ingenious in seeking to improve their lot, and even when markets fail some hope remains for cooperation and efficiency. . . . [E]conomists should not forget that people can be creative and can bypass unsatisfactory institutions.” [Farrell \(1987\)](#)

1 Introduction

How can people resolve their conflicts of interest in an efficient and mutually acceptable way? The Coase theorem ([Coase, 1960](#)), as synthesized by [Medema \(2020\)](#) from six decades of academic debate, asserts:

Efficiency claim: “If agents are rational and the costs of transacting are zero, resources will be allocated efficiently independent of how rights over those resources are initially distributed.” *Invariance claim:* “Moreover, if utility functions are uniformly affine and the registration of subjective values is not wealth-constrained, this allocation is independent of the initial rights structure.”¹ In particular, “[a] world of zero transaction costs . . . is characterized by fully specified property rights, transferable utility, and costless information.”

It is commonplace in contemporary economics that the Coase theorem cannot be established under asymmetric information and strategic interaction.² In its typical mechanism-design interpretation, the ‘theorem’ or, say, ‘Coasean implementation’ reads as follows:

Efficiency claim: In quasi-linear environments with transferable utility and endowment-unconstrained agents, any allocation problem can be resolved with an ex-post budget-balanced mechanism that is Bayesian incentive-compatible, interim individually rational, and ex-post Pareto-efficient.³

Invariance claim: This mechanism is invariant to changes in the status quo.

Based on asymmetric information about agents’ preferences for consumption, already the efficiency claim has been falsified for numerous allocation problems.⁴ And even if one is

¹See [Bergstrom \(2017\)](#) for details on “uniform affinity,” which appears here merely as quasi-linearity.

²Even under complete information, strategic interaction inevitably violates the theorem in its claim to universality; see, e.g., [Dixit and Olson \(2000\)](#), [Jackson and Wilkie \(2005\)](#), and [Ellingsen and Paltseva \(2016\)](#). For a general discussion, see [Bertrand \(2019\)](#).

³Ex-post budget balance plays a crucial role: It guarantees the *autonomy* of the group of agents, since the mechanism need not be subsidized from the outside, and it is a facet of *zero transaction costs*, since the mechanism does not run a deficit. (Another facet of zero transaction costs are the informational requirements for Bayesian implementation, which are implicitly assumed to be satisfied.)

⁴Counterexamples have been provided by, e.g., [Myerson and Satterthwaite \(1983\)](#) for bilateral trade, [Williams \(1999\)](#) for multilateral trade, [Mailath and Postlewaite \(1990\)](#) for the provision of public goods, and, in particular, [Segal and Whinston \(2016\)](#) in a rather general bargaining framework. For an early discussion of the Coase theorem in the light of incomplete information, see [Farrell \(1987\)](#).

satisfied with second-best solutions, there has not been a shred of status-quo invariance.⁵ The present study seeks to add some color to this dismal picture.

What do the counterexamples put forward against ‘Coasean implementation’ have in common? Certainly, two things. First, mechanisms that are both incentive-compatible and efficient are, in most cases, feasible;⁶ but those mechanisms unfold *distributive effects* that violate the participation constraints of some agents. Loosely speaking, the mechanisms designed to resolve the problem of interpersonal externalities inevitably create new externalities to which those who anticipate to be harmed would not consent.

And second, asymmetric information has always been associated with *allocation*, but (almost) never with *distribution*: Agents have always been assumed to be privately informed about how they perceive the externalities that others might impose on them (this is the *allocation* part), but how they perceive the externalities that they themselves impose on those same others (that’s the *distribution* part) has mostly been assumed *common knowledge*, typically coming along as agents’ selfishness, because ‘they wouldn’t mind.’⁷

This study explores mechanism design for other-regarding agents: In any interpersonal relationship, every one person is to some extent uncertain about every other person’s propensity to internalize the externalities that their actions might impose on others. That is, there is asymmetric information about the extent to which agents internalize the distributive effects of a mechanism. I show that this kind of information asymmetry does not necessarily have adverse effects but, instead, can be operationalized to satisfy agents’ participation constraints and, thereby, simplify the mechanism-design problem substantially. Specifically, in an otherwise standard model, I explore mechanism design for *social networks* of agents whose more or less altruistic or spiteful preferences as well as preferences for consumption are *all* subject to asymmetric information.

The main result, Theorem 1, is a constructive proof of the mechanism-design version of the Coase theorem, in its efficiency claim and, to large extent, even in its invariance

⁵See, e.g., [McKelvey and Page \(2002\)](#) and [Segal and Whinston \(2016\)](#).

⁶For the majority of allocation problems that have been invoked against the Coase theorem, a feasible mechanism is the renowned AGV-mechanism by [Arrow \(1979\)](#) and [d’Aspremont and Gérard-Varet \(1979\)](#). An exception has been provided by [Jehiel and Moldovanu \(2001\)](#) who show that allocation problems with *multi-dimensional* allocative and informational externalities do “generically” not even allow for efficiency plus incentive compatibility. I will refer to their study in more detail below.

⁷Exceptions are the studies of [Bierbrauer and Netzer \(2016\)](#), [Bartling and Netzer \(2016\)](#), and [Bierbrauer et al. \(2017\)](#). These studies, however, restrict attention to mechanisms that are *social-preference robust*; thereby, they neglect right from the outset the implementation *possibilities* stemming from this kind of information asymmetry. I will refer to [Bierbrauer and Netzer \(2016\)](#) in more detail below.

claim, for social networks of *at least three* agents. Hence, except for dyads, the standard counterexamples put forward against Coasean implementation are not robust with respect to *asymmetric information* about agents' interpersonal concerns.

One is inclined to think that extending the scope of asymmetric information must, generally, further confine the set of 'feasible' mechanisms. While this is clearly so for incentive-compatibility constraints, Theorem 1 shows in particular that this assumption is untenable when it comes to participation constraints. Notice that the requirement of budget balance changes character in the presence of social preferences; while *monetary* transfers can be ex-post budget-balanced, the corresponding utility transfers cannot.

The reasoning behind the possibility result is essentially the following. First, an ex-post *materially* efficient mechanism (which maximizes the sum of agents' private payoffs while being ex-post budget-balanced) is also ex-post *Pareto*-efficient if agents are 'moderately' altruistic or spiteful; second, such mechanism can be social-preference robust in that it operates exclusively on agents' preferences for consumption while leaving (the asymmetry of information about) agents' social preferences as a strategic degree of freedom for further interaction; and third, this strategic degree of freedom can be utilized by complementing the mechanism with an additional, ex-post budget-balanced game that attracts agents' participation in the mechanism at large. More specifically, I show that 'Coasean implementation' is feasible with a direct mechanism consisting of two parts, the *terms of trade* and a *gamification scheme*.

The *terms of trade* operate exclusively on agents' preferences for consumption and are dedicated to the resolution of the actual allocation problem. They specify an ex-post materially efficient social alternative and ex-post budget-balanced transfers that apply the principle of the *dyadical* AGV-mechanism (due to [Arrow, 1979](#), and [d'Aspremont and Gérard-Varet, 1979](#)) to each and every single dyad: Each agent i makes to every other j a gross monetary concession amounting to what j expects to contribute to i 's material well-being when announcing his payoff type. I show that the terms of trade are Bayesian incentive-compatible, ex-post Pareto-efficient whenever agents are moderately altruistic or spiteful, and social-preference robust in that they leave agents' social preferences strategically inoperative.

The *gamification scheme* is an additional transfer scheme, not at all affecting the choice of social alternatives, serving a single purpose: *to attract agents' participation in*

the terms of trade. A gamification scheme specifies ex-post budget-balanced transfers that operate exclusively on agents' social preferences, are strategy-proof, and implement a sufficiently strong, and strict, *interim-expected Pareto improvement upon a status quo of zero-transfers*. Strikingly, this is possible for social networks of *at least three* agents. (By contrast, no ex-post budget-balanced transfer scheme Pareto-dominates another if agents are moderately altruistic or spiteful.) Under the assumption that agents' endowments are sufficiently large, this interim-expected Pareto improvement can be amplified to satisfy any given collection of interim participation constraints, by scaling all the components of the gamification scheme with the same sufficiently large factor.

In other words, I show how mutual participation in a game that merely redistributes money between them yields agents an interim-expected Pareto improvement upon a status quo in which such game is not played. The intuition behind this possibility is the following: Under the assumption that all social types participate, relatively selfish types can be attracted with the interim expectation to earn money; that money can be collected from relatively pro- (anti-) social types by compensating those mentally with the interim-expected positive (negative) externalities that their actions in the game would impose on their opponents. These interim expectations about the distributive consequences of the game can be rendered *ex-ante* budget-balanced for networks of *arbitrary* size and even *ex-post* budget-balanced for *at least three* agents.

Gamification schemes can be implemented with various ex-post payoff-equivalent game forms that provide agents with either team- or relative-performance incentives, or hybrids of those. Under all those game forms, each agent imposes more or less pronounced positive (negative) effort externalities on every other agent, depending on whether he is more or less altruistic (spiteful) towards that same agent. Utility-wise, then, gamification results in games that, on every single bilateral level, are closely related to hawk-dove.

These implications can be interpreted as agents deriving 'joy' from playing (i.e., participating in) games that involve positive or negative effort externalities. While each agent benefits from gamification *ex interim*, some agents might regret *ex post* to have participated at all, depending on the composition of social types at play. Put simply, as it is typical for *parlor games*, everybody wants to play, but there will be winners and losers both in terms of ex-post payoffs and ex-post utilities. Notice that joy from playing emerges endogenously.

The variety of game forms that can be used to attract agents' participation in an otherwise unattractive mechanism justifies the term *gamification*: The abundance of characterizations of “gamification,” to be found in the vast non-economical literature on the topic, all surround the idea of applying “game design principles in non-gaming contexts” (Robson et al., 2015). Economically speaking, the vague idea is to complement contracts with additional incentives so as to increase *efficiency* and attract agents' *participation*. With its implicit appeal to the *autonomy* of the contracting parties, the very concept is, in my understanding, closely related to the Coase theorem. However, or perhaps even because of its relation to the Coase theorem, this literature is aware of lacking a clean theoretical foundation.⁸ This lack of theory is associated with microeconomic theory lacking a clean foundation of how agents derive joy from playing games.⁹ The present study proposes an incentive-theoretical foundation of both, how agents derive joy from paying games and gamification as such.

With a broader notion of monetary transfers, appealing to transferable utility, some figurative examples can be given: Think of a community organizing a fundraiser 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, if only second-best, is actually quite simple, realistically speaking: ‘Once you’re in, you have to give,’ as a matter of social norm. Events of this sort are often complemented with some soft-core incentive device, like awarding the best-dressed guest. The major purpose of such add-on contest is not to make guests dress well, but rather to suppress free-riding-at-the-doorstep by compensating participants for their monetary ‘losses’ (the lost returns from free-riding) with the ‘joy’ they derive from playing games. Awarding the best-dressed guest provides participants with a platform to live out their propensities to compete, and it is this attraction that helps pull them over the doorstep.¹⁰ Another example is the ‘peace-keeping process’ complemented with international sporting events such as the Olympic games, from which ‘defective’ countries are threatened to be excluded.¹¹

⁸See, e.g., the surveys of Seaborn and Fels (2015) and Koivisto and Hamari (2019).

⁹To the best of my knowledge, there is but one study that puts gamification into a game-theoretical perspective: Hamari, Huotari, and Tolvanen (2015) address several questions that arise with this concept, however, without taking a comprehensive incentive-theoretical view.

¹⁰A similar point is frequently made in conceptual research on how to organize fundraisers; see, e.g., Webber (2004) and Pelozo and Hassay (2007).

¹¹For instance, as many as 66 countries boycotted the Moscow Olympics in 1980 in response to the Soviet Union’s invasion of Afghanistan. South Africa was banned between 1964 and 1988 to amplify political pressure on the Apartheid regime.

The paper proceeds as follows. Section 2 outlines the model framework. I introduce *social networks* of altruistic (or spiteful) agents in the manner of [Bourlès, Bramoullé, and Perez-Richet \(2017\)](#);¹² here, however, agents' interpersonal concerns are their private information. Agents face an allocation problem with asymmetric information about their private payoffs from different social alternatives.

In Section 3, I introduce *gamification mechanisms* formally. Those consist of the payoff-type dependent *terms of trade* (which are Bayesian incentive-compatible, ex-post materially efficient, ex-post Pareto-efficient if agents are moderately altruistic or spiteful, and social-preference robust) and a social-type dependent *gamification scheme* (which, by definition, is strategy-proof, ex-post budget-balanced, and Pareto-improves ex interim upon a status quo of zero-transfers). If gamification schemes exist, they can be designed to satisfy any given interim participation constraints; however, as I show, gamification is not feasible for *dyads*.

The remainder of the paper is dedicated to showing that gamification schemes do exist for social networks of *more than two* agents, and how they can be constructed. I start out, in Section 4, by considering gamification schemes that are not ex-post but *ex-ante* budget-balanced. This simplification allows me to construct gamification schemes successively and make the economic intuition behind them transparent. Ex-ante budget balance can be interpreted as the participation constraint of a rent-seeking mediator who is not affected herself by the choice of social alternatives and who subsidizes or sanctions agents based on their actions in the corresponding game. I show that *ex-ante* budget-balanced gamification is feasible for networks of arbitrary size, *including dyads*.

In Section 5, I reimpose *ex-post* budget balance. The *ex-ante* budget-balanced gamification scheme derived in Section 4 can be adapted such that the role of the mediator can be taken on by some network member, who herself is now affected by the choice of social alternatives. Gamification is thus feasible with *at least three* agents. Based on this insight, I establish 'Coasean implementation' for social networks of at least three agents.

Sections 4 and 5 also show how gamification schemes can be implemented with either team- or relative-performance incentives, and even hybrids of those. Section 6 concludes.

¹²For evidence on altruism, see [Andreoni and Miller \(2002\)](#), [Charness and Rabin \(2002\)](#), and [Bruhin, Fehr, and Schunk \(2019\)](#). For evidence on spite, see [Saijo and Nakamura \(1995\)](#), [Fehr, Hoff, and Kshetramade \(2008\)](#), and [Prediger, Volland, and Herrmann \(2014\)](#).

2 The Model

2.1 Allocations, Utility, and Information

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

Agents exhibit interpersonal preferences in the form of altruism or spite: From the allocation of private payoffs, agent i derives ex-post *utility*

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

where the value δ_{ij} that i assigns to j 's payoff, $j \neq i$, belongs to a closed (proper) interval $\Delta_{ij} = [\delta_{ij}^{\min}, \delta_{ij}^{\max}] \subset (\frac{-1}{n-1}, 1)$, while $\delta_{ii} = 1$. Notice that $(\frac{-1}{n-1}, 1)$ is the maximum range of interpersonal altruism, or spite, for which agents care about overall material efficiency while still being selfish to the extent that every one of them prefers a dollar to be his own rather than having that same dollar distributed among the others, regardless how. The model allows in particular for all agents being selfish; i.e., $\delta_{ij} = 0$ for all i and all $j \neq i$.

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*.¹³ I denote the collection of social types by $\delta = (\delta_i, \delta_{-i})$, with $\delta_{-i} = (\delta_j)_{j \neq i}$, and Cartesian products of type sets by $\Theta = \prod_i \Theta_i$, $\Delta_i = \prod_{j \neq i} \Delta_{ij}$, and $\Delta = \prod_i \Delta_i$.

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 while edges are weighted by the degrees of altruism between agents. I refer to this structure as the agents' *social network*. Figure 1 gives an illustration.

Each agent is privately informed about his payoff type and social type, which realize independently according to strictly positive densities. The various degrees of altruism

¹³See Bergstrom (1999) and the discussion in Broulès, Bramoullé, and Perez-Richet (2017) for how agents' social types can be *derived* if their degrees of altruism are not directly attributed to other agents' private payoffs but, instead, to their *utilities*.

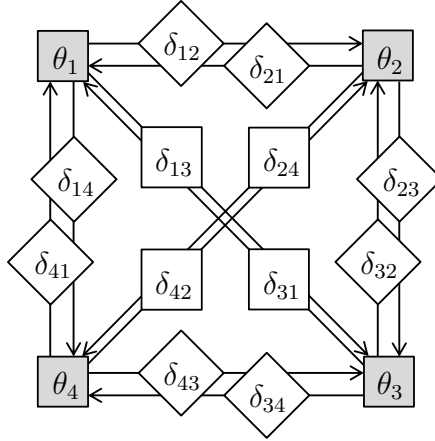


Figure 1: A social network of $n = 4$ agents with payoff types $(\theta_i)_i$ and more or less altruistic or spiteful interpersonal preferences $(\delta_{ij})_{i,j \neq i}$, all of which are private information.

determining i 's social type may correlate. At the interpersonal level, agents' types are independent.

Despite the asymmetry of information, it can still be common knowledge who is 'friends' and who is 'foes.' For instance, if $\delta_{kl}^{\max}, \delta_{\ell k}^{\max} < 0 < \delta_{ij}^{\min}, \delta_{ji}^{\min}$, then, in comparison, i and j are friends, whereas k and ℓ are foes. Similarly, it can be common knowledge that i likes j more than k , which is the case if $\delta_{ik}^{\max} < \delta_{ij}^{\min}$. The key assumption is that in any bilateral relationship there remains, to some extent, uncertainty about who (dis-)likes whom how much. Notice that the variance of every δ_{ij} , while strictly positive, is allowed to be arbitrarily small. In this respect, reciprocal interpersonal preferences can be captured by letting $\Delta_{ij} = \Delta_{ji}$ and $\delta_{ij}^{\min} \approx \delta_{ij}^{\max}$.

2.2 Material Efficiency and Pareto Efficiency

I assume (and relax this assumption only in Section 4) that agents do not have access to an outside source of money, such that transfers must satisfy $\sum_{i \in \mathcal{I}} t_i \leq 0$. The agents' problem is to choose a social alternative k and transfers $(t_i)_{i \in \mathcal{I}}$ such that the resulting allocation is ex-post Pareto-efficient.

The following Lemma gives rise to focusing on allocations that are *ex-post materially efficient*: $k^*(\theta) \in \arg \max_{k \in K} \sum_{i \in \mathcal{I}} \pi_i(k | \theta_i)$ while transfers $(t_i)_{i \in \mathcal{I}}$ are ex-post budget-balanced: $\sum_{i \in \mathcal{I}} t_i = 0$.

Lemma 1 *An allocation is ex-post Pareto-efficient only if transfers are budget-balanced. If $|\delta_{ij}| < \frac{1}{2n-3}$ for all i and all $j \neq i$, then an ex-post materially efficient allocation, consisting of a social alternative $k^*(\theta) \in \arg \max_{k \in K} \sum_{i \in \mathcal{I}} \pi_i(k | \theta_i)$ and ex-post budget-balanced transfers, is also ex-post Pareto-efficient; in particular, no ex-post Pareto-efficient transfer scheme Pareto-dominates another.*

Proof. See Appendix A. ■

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

Notice that the Pareto frontier can be indefinite for combinations of social types satisfying $|\delta_{ij}| \geq \frac{1}{2n-3}$, in which case a subgroup of agents might be willing to transfer arbitrary amounts of money to their joint favorite agent.¹⁴

2.3 Revelation Mechanisms

A *direct* revelation mechanism involves the agents in a strategic game of incomplete information in which they are asked to report their types truthfully; based on their reports, a social alternative will be chosen and transfers will be made.¹⁵

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

Then $\langle k, T \rangle$ is *Bayesian incentive-compatible* if, for all $i \in \mathcal{I}$ and all $(\theta_i, \delta_i) \in \Theta_i \times \Delta_i$, $U_i(\theta_i, \delta_i) = \max_{(\hat{\theta}_i, \hat{\delta}_i) \in \Theta_i \times \Delta_i} U_i(\hat{\theta}_i, \hat{\delta}_i | \theta_i, \delta_i)$. In this case, k is *Bayesian implementable*.¹⁶

¹⁴An example is the network of three agents with $\delta_{13} = \delta_{23} > 1/3$, $\delta_{12} = \delta_{21} = -1/3$, and $\delta_{31} = \delta_{32} = 0$, in which agents 1 and 2 are willing to *jointly* transfer arbitrary individual amounts of $t > 0$ to agent 3.

¹⁵By the *revelation principle*, which applies to the present setup (Myerson, 1979), there is no loss of generality in considering only direct mechanisms.

¹⁶Bayesian implementation has been criticized for assuming that the ex-ante distribution of agents' types is common knowledge. As an alternative, Bergemann and Morris (2005) have proposed *ex-post im-*

Formally, the mechanism-design problem is one of one-dimensional allocative and informational externalities.¹⁷

2.4 Interim Participation Constraints

The application of a mechanism requires all agents' approval at the interim stage. Agents' interim-expected utilities from participating in that mechanism must therefore exceed their interim-expected utilities from the consequences of this mechanism not finding unanimous approval. If *property rights*, determining a status-quo social alternative, and *liability rules*, determining interpersonal transfers if bargaining fails, are well-defined, then agents can form rational expectations about those consequences: Depending on the state of the world, (θ, δ) , a status-quo social alternative would be implemented and (e.g., as a result of subsequent bargaining), interpersonal transfers would be made.

Following [Segal and Whinston \(2016\)](#), a property-rights and liability-rules regime can be captured by a direct revelation mechanism $\langle k^\circ, T^\circ \rangle$, with $k^\circ : \Theta \times \Delta \rightarrow K$ and $T^\circ = (t_i^\circ)_{i \in \mathcal{I}} : \Theta \times \Delta \rightarrow \mathbb{R}^n$, that is Bayesian incentive-compatible. Then the mechanism $\langle k, T \rangle$ finds unanimous approval at the interim stage among all types and agents if $U_i(\theta_i, \delta_i) > U_i^\circ(\theta_i, \delta_i) = \sum_{j \in \mathcal{I}} \delta_{ij} [\bar{\pi}_{ij}^\circ(\theta_i, \delta_i) + \bar{t}_{ij}^\circ(\theta_i, \delta_i)]$ on $\Theta_i \times \Delta_i$ for all $i \in \mathcal{I}$, where $\bar{\pi}_{ij}^\circ(\theta_i, \delta_i) = \mathbb{E}_{\theta_{-i}, \delta_{-i}} [\pi_j(k^\circ(\theta, \delta) | \theta_j)]$ and $\bar{t}_{ij}^\circ(\theta_i, \delta_i) = \mathbb{E}_{\theta_{-i}, \delta_{-i}} [t_j^\circ(\theta, \delta)]$. In this case, $\langle k, T \rangle$ is *interim individually rational*.

3 Gamification Mechanisms

The idea of 'gamification,' as conceived in this paper, is to attract agents' participation in an otherwise not individually rational mechanism by complementing this mechanism with

plementation for model economies with interdependent utilities (in which dominant-strategy implementation is not feasible; [Williams and Radner, 1988](#)), requiring that truthful revelation of types constitutes a Nash equilibrium under the respective mechanism. However, as shown by [Jehiel et al. \(2006\)](#), ex-post implementation is 'generically' not feasible in the presence of informational externalities.

¹⁷[Jehiel and Moldovanu \(2001\)](#) show that, in allocation problems of allocative and informational externalities, efficient incentive-compatible mechanisms are 'generically' feasible if and only if agents' privately known payoff externalities are one-dimensional. The here conceived model differs from theirs in two aspects: First, while agents impose *one*-dimensional payoff-related externalities on each other, the extent to which an agent *perceives* those externalities is now his private information, determined by his interpersonal preferences; and second, those perceptions now extend to other agents' transfers, as 'social' agents internalize the distributive effects of a mechanism. While the impossibility result of [Jehiel and Moldovanu \(2001\)](#) on *multi*-dimensional informational externalities reinforces the asymmetric-information argument against 'Coasean implementation,' the 'gamification' approach presented here might still be relevant to attracting agents' participation in second-best solutions.

a ‘joyful’ game: Only if agents participate in the basic mechanism they are allowed to play the ‘joyful’ game. In what follows, I give this notion an incentive-theoretical foundation. I require in particular that ‘joy from playing games’ must emerge endogenously within the model framework of Section 2. I show then that ‘gamification works’ in that it renders ‘Coasean implementation’ possible.

Consider the *dyadical* AGV-mechanism (due to [Arrow, 1979](#), and [d’Aspremont and Gérard-Varet, 1979](#)): For $n = 2$ agents, this mechanism is specified by an ex-post materially efficient allocation function $k^*(\theta) \in \arg \max_{k \in K} \pi_1(k | \theta_1) + \pi_2(k | \theta_2)$ and ex-post budget-balanced transfers $t_i^*(\hat{\theta}) = \mathbb{E}_{\theta_{-i}} [\pi_{-i}(k^*(\hat{\theta}_i, \theta_{-i}) | \theta_{-i})] - \mathbb{E}_{\theta_i} [\pi_i(k^*(\theta_i, \hat{\theta}_{-i}) | \theta_i)]$.

Under this mechanism, each agent $i \in \{1, 2\}$ makes a gross monetary concession to agent $-i$ amounting to what $-i$ believes to contribute to i ’s material well-being when reporting some payoff type $\hat{\theta}_{-i}$, namely $\mathbb{E}_{\theta_i} [\pi_i(k^*(\theta_i, \hat{\theta}_{-i}) | \theta_i)]$.

[Bierbrauer and Netzer \(2016\)](#) have shown that, in quasi-linear environments with two other-regarding agents, the AGV-mechanism is not only Bayesian incentive-compatible but also *social-preference robust* in that it leaves agents’ other-regarding concerns strategically irrelevant; that is, agents are incentivized to behave *as if* they were selfish.¹⁸ The reason is that the material externality that i expects to impose on $-i$ when reporting $\hat{\theta}_i$ is rendered strategically irrelevant to i : If $-i$ reports his payoff type truthfully, then $\mathbb{E}_{\theta_{-i}} [\pi_{-i}(k^*(\hat{\theta}_i, \theta_{-i}) | \theta_{-i}) + t_{-i}^*(\hat{\theta}_i, \theta_{-i})] = \mathbb{E}_{\theta} [\pi_i(k^*(\theta) | \theta_i)]$. Thereby, also i ’s degree of altruism is rendered strategically irrelevant.

While some authors deem social-preference robustness a desirable feature (because common-knowledge assumptions about social-type distributions can thus be avoided), the standard caveat prevails: For most allocation problems, social-preference robust mechanisms are not interim individually rational. Therefore, the present study takes the exact opposite view, by asking: Can the asymmetry of information about agents’ social preferences be operationalized, by complementing an efficient social-preference robust mechanism with a social-type dependent transfer scheme, so as to satisfy agents’ interim participation constraints (while keeping common-knowledge assumptions about social-type distributions as weak as possible)?

The idea is, first, to generalize the robustness result of [Bierbrauer and Netzer \(2016\)](#) to social networks of arbitrary size. This will be achieved by applying the mutual-concessions

¹⁸[Bierbrauer and Netzer \(2016\)](#) coin this property the ‘insurance property,’ as it insures agents against the other-regarding concerns of other agents.

principle of the dyadical AGV to each and every single dyad: For ex-post materially efficient social alternatives $k^*(\theta)$, every agent i pays every other j the money equivalent of what j believes to contribute to i 's material well-being when reporting $\hat{\theta}_j$; that is, i transfers $\mathbb{E}_{\theta_{-j}}[\pi_i(k^*(\hat{\theta}_j, \theta_{-j}) | \theta_i)]$ to j and receives $\mathbb{E}_{\theta_{-i}}[\pi_j(k^*(\hat{\theta}_i, \theta_{-i}) | \theta_j)]$ from j .

I refer to the resulting mechanism as the *terms of trade*, for it prescribes how agents should trade in the tangible externalities they expect to impose on each other when choosing one social alternative or another. Notice that the terms of trade are ex-post budget-balanced and operate exclusively on agents' reported payoff types.

I show that the terms of trade implement ex-post materially efficient allocations (which, by Lemma 1, are 'often' ex-post Pareto-efficient) and are social-preference robust. Thereby, they preserve agents' social preferences as a strategic degree of freedom.

Second, this strategic degree of freedom can be utilized by complementing the terms of trade with an *ex-post budget balanced* transfer scheme, independent of social alternatives k , that yields agents a sufficiently strong *interim-expected Pareto improvement upon a status quo of zero-transfers*. (That this is actually possible is what this paper is all about. Recall from Lemma 1 that an *ex-post* Pareto improvement through ex-post budget-balanced transfers is never feasible if agents are moderately altruistic or spiteful.) Under this requirement, agents' interim-expected utilities from unanimous participation will outweigh their interim reservation utilities. I refer to such transfer schemes as *gamification schemes*.

The term 'gamification' is appropriate because an interim-expected Pareto improvement upon a status quo of zero-transfers can be interpreted as agents deriving 'joy' from playing the game induced by such transfer scheme. Notice that 'joy' is required to emerge endogenously.

A *gamification mechanism*, thus, complements the terms of trade with a gamification scheme. Only if agents participate in the terms of trade, they are allowed to play the 'joyful' game induced by the gamification scheme. Formally:¹⁹

¹⁹With respect to private payoffs, gamification mechanisms belong to the class of *expected-externality mechanisms*, defined by materially efficient allocations k^* and transfers $t_i(\hat{\theta}) = \sum_{j \neq i} \mathbb{E}_{\theta_{-i}}[\pi_j(k^*(\hat{\theta}_i, \theta_{-i}) | \theta_j)] + h_i(\hat{\theta}_{-i})$, for arbitrary functions $h_i : \Theta_{-i} \rightarrow \mathbb{R}$. Notice that the AGV, defined through $h_i(\hat{\theta}_{-i}) = \frac{-1}{n-1} \sum_{j \neq i} \sum_{\ell \neq j} \mathbb{E}_{\theta_{-j}}[\pi_\ell(k^*(\hat{\theta}_j, \theta_{-j}) | \theta_\ell)]$, coincides with the *terms of trade* if and only if $n = 2$. While the AGV subsidizes or sanctions the *average* externalities that an agent imposes on the rest of the group, gamification mechanisms treat interpersonal externalities on the *bilateral* level.

Definition 1 (Gamification Mechanisms) A gamification mechanism $\langle k^*, T^* \rangle$ consists of the ex-post materially efficient allocation function $k^* : \Theta \rightarrow K$, with $k^*(\theta) \in \arg \max_{k \in K} \sum_{i \in \mathcal{I}} \pi_i(k | \theta_i)$, and an ex-post budget-balanced transfer scheme $T^* = (t_i^*)_{i \in \mathcal{I}} : \Theta \times \Delta \rightarrow \mathbb{R}^n$ defined by

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

where the gamification scheme $s^* = (s_i^*)_{i \in \mathcal{I}} : \Delta \rightarrow \mathbb{R}^n$ satisfies the following conditions:

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

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

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

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

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

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

The next two Propositions are the basis for establishing ‘Coasean implementation’ through ‘gamification’ (Theorem 1 in Section 5.1).

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

Proof. Efficiency: Gamification mechanisms are ex-post materially efficient by construction; hence, by Lemma 1 (with a proof in Appendix A), they are also ex-post Pareto-efficient if $|\delta_{ij}| < \frac{1}{2n-3}$ for all i and all $j \neq i$.

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

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

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

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

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

By equation (1), the terms of trade are social-preference robust: Agents' social preferences are rendered strategically irrelevant when it comes to implementing the materially efficient allocation function k^* . Consequently, the asymmetry of information about agents' social preferences can (for the moment: hypothetically) be operationalized to satisfy agents' interim participation constraints. The idea is simply to amplify the interim-expected

Pareto improvement induced by s^* through Definition 1(iii) by scaling all the s_i^* with the same sufficiently large factor:

Proposition 2 *If gamification schemes exist and agents' endowments are sufficiently large, then there exist gamification mechanisms that satisfy any given interim participation constraints.*

Proof. Suppose a gamification scheme $s^* = (s_i^*)_{i \in \mathcal{I}}$ exists. Conditions (i) to (iii) of Definition 1 are invariant under scaling all the components s_i^* with the same $\alpha^* > 0$, while the assumption of sufficiently large individual endowments guarantees that each agent i is able to make the respective payments for any given α^* . Now consider the gamification mechanism $\langle k^*, T^* \rangle$ with modified gamification scheme $(\alpha^* \cdot s_i^*)_{i \in \mathcal{I}}$. By equation (1), and since $\langle k^*, T^* \rangle$ is Bayesian incentive-compatible, agent i 's interim-expected utility from unanimous participation in $\langle k^*, T^* \rangle$ is given by

$$(2) \quad U_i(\theta_i, \delta_i) = \mathbb{E}_{\theta_{-i}} \left[\sum_{\ell \in \mathcal{I}} \pi_\ell(k^*(\theta) | \theta_\ell) \right] + \left(\sum_{j \neq i} \delta_{ij} \right) \mathbb{E}_\theta \left[\sum_{\ell \in \mathcal{I}} \pi_\ell(k^*(\theta) | \theta_\ell) \right] \\ - (n-1) \mathbb{E}_\theta \left[\sum_{j \in \mathcal{I}} \delta_{ij} \pi_j(k^*(\theta) | \theta_j) \right] + \alpha^* \cdot \sum_{j \in \mathcal{I}} \delta_{ij} \mathbb{E}_{\delta_{-i}} [s_j^*(\delta)].$$

Notice that $U_i(\theta_i, \delta_i)$ is independent of the status-quo property-rights and liability-rules regime $\langle k^\circ, T^\circ \rangle$, as defined in Section 2.4. By Definition 1(iii), $\sum_{j \in \mathcal{I}} \delta_{ij} \mathbb{E}_{\delta_{-i}} [s_j^*(\delta)] > 0$. Hence, if α^* is chosen sufficiently large, then $U_i(\theta_i, \delta_i) > U_i^\circ(\theta_i, \delta_i)$ for any given interim reservation utility $U_i^\circ(\theta_i, \delta_i)$, for all $i \in \mathcal{I}$ and all $(\theta_i, \delta_i) \in \Theta_i \times \Delta_i$. ■

Whether gamification schemes exist, and how to construct them if they do, is not a trivial question. The intuitive suspicion is that conditions (ii) and (iii) of Definition 1 are mutually exclusive, which is (only) true for dyads:

Proposition 3 *Gamification schemes do not exist if $n = 2$.*

Proof. Suppose the opposite is true. Then an interim-expected Pareto improvement upon a status quo of zero-transfers requires that $0 < \mathbb{E}_{\delta_{-i}} [s_i^*(\delta)] + \delta_i \mathbb{E}_{\delta_{-i}} [s_{-i}^*(\delta)]$ for both $i \in \{1, 2\}$ and all $\delta_i \in (-1, 1)$, while $s_{-i}^*(\delta) = -s_i^*(\delta)$, due to ex-post budget balance. Hence, $0 < (1 - \delta_i) \mathbb{E}_{\delta_{-i}} [s_i^*(\delta)]$, implying that $0 < \mathbb{E}_{\delta_{-i}} [s_i^*(\delta)]$ for all i, δ_i . But then, $0 < \mathbb{E}_\delta [s_i^*(\delta)]$ for both i , in contradiction to ex-post budget balance. ■

The next two Sections construct gamification schemes for social networks of $n \geq 3$ agents. The conclusive existence result is Proposition 8 in Section 5.1.

4 Ex-Ante Budget Balance: Gamification with the Help of an Outside Mediator

I construct gamification schemes successively. For this purpose, I first consider gamification schemes that are not ex-post but *ex-ante* budget-balanced:²⁰ $\sum_{i \in \mathcal{I}} \mathbb{E}_\delta [s_i^*(\delta)] = 0$ for all $\delta \in \Delta$. (I reimpose ex-post budget balance in Section 5 below.)

An ex-ante budget-balanced gamification scheme can be interpreted as follows. A *mediator* M is an agent *outside* the social network about whom the following is common knowledge: No network member cares about her; she is a pure-payoff maximizer, implying that she does not care about network members either; she is not affected by the choice of social alternatives k ; she knows the distribution of agents' types; she can offer agents to play a game the outcomes of which merely redistribute money between agents and herself; she can commit to the ex-post outcomes of that game; and she is willing to offer the game if and only if she expects a non-negative payoff from it.

When interpreting the gamification scheme s^* as *subsidies* or *sanctions* by that mediator M , then M 's ex-post payoff is given by $s_M(\delta) = -\sum_{i \in \mathcal{I}} s_i^*(\delta)$, such that M 's participation constraint is just satisfied if $\sum_{i \in \mathcal{I}} \mathbb{E}_\delta [s_i^*(\delta)] = 0$.

This Section proceeds as follows. I start out with an example of an ex-ante budget-balanced gamification scheme in a simple two-agents environment. Section 4.2 then derives ex-ante budget-balanced gamification schemes for arbitrary dyads. These schemes are adapted to arbitrary social networks in Section 4.3. Finally, Section 4.4 justifies the term 'gamification' by showing how ex-ante budget-balanced gamification schemes can be implemented with a multitude of qualitatively different game forms.

4.1 An Example

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

²⁰For first steps in this direction see my unpublished paper [Daske \(2016\)](#).

monetary outcomes of which are determined by the transfer scheme $(s_i)_{i=1,2}$ depicted on the left-hand side of Figure 2 (a prisoners' dilemma, from a pure-payoff perspective); M 's payoff is given by $s_M = -s_1 - s_2$.²¹

Maximizing utility $u_i = s_i + \delta_i s_{-i}$, a selfish type ($\delta_i = 0$) has the dominant strategy to behave 'hawkish' in that he opts for playing H , whereas an altruist ($\delta_i = 1/2$) has the dominant strategy to behave 'dovish' in that he opts for playing D . The game offered by M is thus *strategy-proof* and separates hawks from doves behaviorally.²² The respective utility levels are depicted on the right-hand side of Figure 2: Dominant-strategy utility levels, considered jointly for both social types, are those of *hawk-dove*, in the strict game-theoretical sense.

Consider next a status quo of zero-transfers, and assume that offering and playing the game requires unanimous approval among the three parties. Given that types are i.i.d. and equally likely, $(s_i)_i$ is *ex-ante budget-balanced*, because the eight payoffs specified by $(s_i)_i$ sum up to zero. M 's ex-ante expected payoff is thus zero, and her participation constraint is satisfied.

If the other agent, whether hawk or dove, is willing to participate, a hawk interim-expects the payoff $(5x - 3x)/2 = x > 0$ and is thus willing to participate.

The decisive role is played by the dovish type. If the other agent, whether hawk or dove, is willing to participate, a dove's interim-expected *monetary loss* amounts to $(3x - 5x)/2 = -x$. This loss serves to attract the hawkish type. For her monetary loss, a dove is *mentally compensated* by the positive monetary externality that her appearance would have on both a hawk (who would obtain $5x$) or another dove (who would obtain $3x$). Her interim-expected externality amounts to $(5x + 3x)/2 = 4x$ and yields her an interim-

²¹This example is obtained from the gamification scheme $(s_i^*)_i$ of equation (7) below when letting $s_i = 8x \cdot s_i^*$, where $s_i^*(\delta) = 2\delta_{-i} - \delta_i^2 - 3/8$ under the distributional assumptions made right here.

²²The transfer scheme can be interpreted as resulting from relative- or team-performance incentives. For instance, M can offer an *all-pay auction* in which investments (or efforts) $e_i \in \{0, 1\}$ result in payoffs

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

where $F = 5x$ is a participation fee, $C = 6x$ are investment costs, and $P = 16x$ is the winner's prize. Here, the hawkish strategy H is to invest, whereas the dovish strategy D is to abstain. *Payoff-equivalently*, M can offer a (local) *public-good game* in which investments $e_i \in \{0, 1\}$ result in payoffs $s_i(e_i, e_{-i}) = -F - Ce_i + B(e_i + e_{-i})$, at participation fee $F = 3x$, investment costs $C = 10x$, and team bonus per unit of effort $B = 8x$. In this case, strategy H is to free-ride, whereas D is to invest.

		s_{-i}		
		D	H	
s_i	D	$3x$	$3x$	$-5x$ $5x$
	H	$5x$	$-5x$	$-3x$ $-3x$

		u_{-i}		
		D	H	
u_{Dove}	D	$\frac{9}{2}x$	\dots	$-\frac{5}{2}x$ \dots
	H	$\frac{5}{2}x$	\dots	$-\frac{9}{2}x$ \dots

		u_{-i}		
		D	H	
u_{Hawk}	D	$3x$	\dots	$-5x$ \dots
	H	$5x$	\dots	$-3x$ \dots

Figure 2: On the left: The transfer scheme $(s_i)_{i=1,2}$ underlying the strategic game that a mediator offers two agents to play, where $x > 0$. 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 selfish type, with $u_i = s_i$, acts ‘hawkish’ by playing his dominant strategy H . If types are i.i.d. and equally likely, then transfers are ex-ante budget-balanced, and a comparison of the dominant-strategy related utility levels from meeting either hawk or dove reveals that mutual participation yields agents an interim-expected Pareto improvement.

expected mental compensation of $\delta_D \cdot 4x = 2x$, which clearly outweighs her interim-expected monetary loss.

Hence, ex interim, offering and participating in $(s_i)_i$ are mutually consistent decisions among the three parties. In particular, mutual participation provides both social types with a *strict interim-expected utility gain*. Consequently, agents have a willingness to pay for their participation: M , not knowing agents’ types, can demand a *uniform* participation fee amounting to $f < \frac{2}{3}x$.²³ In this case, M realizes herself a strict ex-ante expected utility gain.²⁴

Their interim-expected utility gains can be interpreted as agents deriving ‘joy’ from playing the game, where the source of that joy are the interim-expected distributive

²³**Proof:** A hawk, seeking money only, is willing to pay x at most, which is his interim-expected payoff from participation. A dove is willing to pay and participate if and only if her own interim-expected monetary loss, which amounts to $(3x - 5x)/2 - f = -(x + f)$, is compensated for by the mental compensation that she receives from her interim-expected monetary externality on the other agent, given by $\delta_D \cdot (4x - f) = 2x - f/2$. Hence, $f < 2x/3$.

²⁴This insight suggests that the same implications would hold if M was part of a social network of $n = 3$ agents: M would want to collect the fee whenever she valued her own material wealth more than that of any other agent, and agents would be willing to pay if the fee was not too high and their (potential) spite towards M was not too strong. This intuition will be the basis for constructing gamification schemes that are *ex-post* budget-balanced, in Section 5 below.

effects of the game. Ex post, however, some party must regret to have participated in, or offered, the game: Collective joy from playing is an interim phenomenon.

With endowments being arbitrarily large, every increase in x allows for even further interim-expected Pareto improvements.

4.2 Ex-Ante Budget-Balanced Gamification: Two Agents

With all else equal to the model setup of Section 2.1, assume here that there are only two agents. A reasonable approach is to look for an ex-ante budget-balanced gamification scheme $s^* = (s_1^*, s_2^*)$ with twice continuously partially differentiable functions $s_i^* : \Delta_i \rightarrow \mathbb{R}$. Hence, s^* must be strategy-proof, implying that

$$(3) \quad \frac{\partial s_i^*(\delta)}{\partial \delta_i} + \delta_i \frac{\partial s_{-i}^*(\delta)}{\partial \delta_i} = 0 \quad \text{for both } i \text{ and all } \delta,$$

ex-ante budget-balanced,

$$(4) \quad \mathbb{E}_\delta [s_1^*(\delta) + s_2^*(\delta)] = 0,$$

and yield agents an interim-expected Pareto improvement from unanimous participation,

$$(5) \quad \mathbb{E}_{\delta_{-i}} [s_i^*(\delta)] + \delta_i \mathbb{E}_{\delta_{-i}} [s_{-i}^*(\delta)] = g_i(\delta_i) \quad \text{for both } i \text{ and all } \delta_i,$$

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

The idea is to derive s^* from $(g_i)_{i=1,2}$ for appropriate functions $(g_i)_{i=1,2}$.²⁵

Proposition 4 *Let $i \in \{1, 2\}$. For positive, twice continuously differentiable functions $g_i : \Delta_i \rightarrow (0, \infty)$ satisfying $g_i'' > 0$ and*

$$(6) \quad \mathbb{E}_{\delta_i} [g_i'(\delta_i)] = 0 = \mathbb{E}_{\delta_i} [g_i(\delta_i) - \delta_i g_i'(\delta_i)],$$

²⁵The sufficient conditions of Proposition 4 below can be obtained from conditions (3) to (5) by the following reasoning: By differentiating (3) with respect to δ_{-i} one obtains that $\partial^2 s_i^* / \partial \delta_1 \partial \delta_2 = 0$, implying that s_i^* is additively separable: $s_i^*(\delta) = a_i(\delta_i) + b_i(\delta_{-i})$ for appropriate functions $a_i : \Delta_i \rightarrow \mathbb{R}$ and $b_i : \Delta_{-i} \rightarrow \mathbb{R}$. Hence, by condition (3) again, $a_i'(\delta_i) + \delta_i b_{-i}'(\delta_i) = 0$, such that partial integration yields $a_i(\delta_i) = -\delta_i b_{-i}(\delta_i) + \int_{\delta_{-i}^{\min}}^{\delta_i} b_{-i}(x) dx + C$, for a constant C . Define $g_i(\delta_i) = \int_{\delta_{-i}^{\min}}^{\delta_i} b_{-i}(x) dx + C$. Then, $a_i(\delta_i) = g_i(\delta_i) - \delta_i g_i'(\delta_i)$ and $b_i(\delta_{-i}) = g_{-i}'(\delta_{-i})$. Condition (6) of Proposition 4 can then be imposed in order to comply with conditions (4) and (5).

define the transfer scheme $s^* = (s_1^*, s_2^*)$ by $s_i^*(\delta) = g_i(\delta_i) - \delta_i g_i'(\delta_i) + g_{-i}'(\delta_{-i})$. Then the following holds:

(i) s^* is an ex-ante budget-balanced gamification scheme.

(ii) From unanimous participation in s^* , agent i derives an interim-expected utility gain of $g_i(\delta_i) > 0$ while interim-expecting a transfer of $\mathbb{E}_{\delta_{-i}}[s_i^*(\delta)] = g_i(\delta_i) - \delta_i g_i'(\delta_i)$ to himself and a transfer of $\mathbb{E}_{\delta_{-i}}[s_{-i}^*(\delta)] = g_i'(\delta_i)$ to agent $-i$.

Proof. Ad (i): *Strategy-proofness:* $d[s_i^*(\hat{\delta}_i, \delta_{-i}) + \delta_i s_{-i}^*(\hat{\delta}_i, \delta_{-i})]/d\hat{\delta}_i = (\delta_i - \hat{\delta}_i)g_i''(\hat{\delta}_i)$; as $g_i'' > 0$ by assumption, i has the strictly dominant strategy to report $\hat{\delta}_i = \delta_i$. *Ex-ante budget balance:* $\mathbb{E}_\delta[s_i^*(\delta)] = 0$ for both i , due to (6). *Interim-expected utility:* By condition (6), $\mathbb{E}_{\delta_{-i}}[s_i^*(\delta)] + \delta_i \mathbb{E}_{\delta_{-i}}[s_{-i}^*(\delta)] = [g_i(\delta_i) - \delta_i g_i'(\delta_i)] + \delta_i [g_i'(\delta_i)] = g_i(\delta_i) > 0$. Ad (ii): Obvious from the arguments on (i). ■

Under the gamification scheme of Proposition 4, the transfer that an agent interim-expects for himself is maximal, and positive, if that agent is a pure-payoff maximizer ($\delta_i = 0$), since $d\mathbb{E}_{\delta_{-i}}[s_i^*(\delta)]/d\delta_i = -\delta_i g_i''(\delta_i)$, while $g_i'' > 0$. Money is thus redistributed to those agents who ‘care least’ about others. On the other hand, the transfer that an agent interim-expects for his opponent increases in his own social type, since $d\mathbb{E}_{\delta_{-i}}[s_{-i}^*(\delta)]/d\delta_i = g_i''(\delta_i) > 0$, and is zero ex ante, since $\mathbb{E}_{\delta_i}[g_i'(\delta_i)] = 0$. Hence, least (most) altruistic types interim-expect to impose a negative (positive) externality on their opponent. This interim-expected externality, weighted with an agent’s social type, *mentally overcompensates* for interim-expected *monetary losses*: $\mathbb{E}_{\delta_{-i}}[s_i^*(\delta)] + \delta_i \mathbb{E}_{\delta_{-i}}[s_{-i}^*(\delta)] = g_i(\delta_i) > 0$.

From an ex-post perspective, mediator M prefers hawkish (less altruistic) over dovish (more altruistic) players; by contrast, agents prefer dovish over hawkish opponents.²⁶

In order to conform with Wilson’s (1987) call for avoiding stark common-knowledge assumptions, the functions $(g_i)_{i=1,2}$ should be chosen such that common-knowledge assumptions about agents’ social-type distributions are as weak as possible. In fact, it suffices to assume common knowledge about the mean and variance of an agent’s social type: The functions

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

²⁶**Proof:** On the one hand, $\partial s_M(\delta)/\partial \delta_i = -\partial[s(\delta_1, \delta_2) + s(\delta_2, \delta_1)]/\partial \delta_i = -(1 - \delta_i)g''(\delta_i) < 0$ for all i and δ_i ; on the other hand, $\partial[s(\delta_i, \delta_{-i}) + \delta_i s(\delta_{-i}, \delta_i)]/\partial \delta_{-i} = (1 - \delta_i \delta_{-i})g''(\delta_i) > 0$ for all i and all δ .

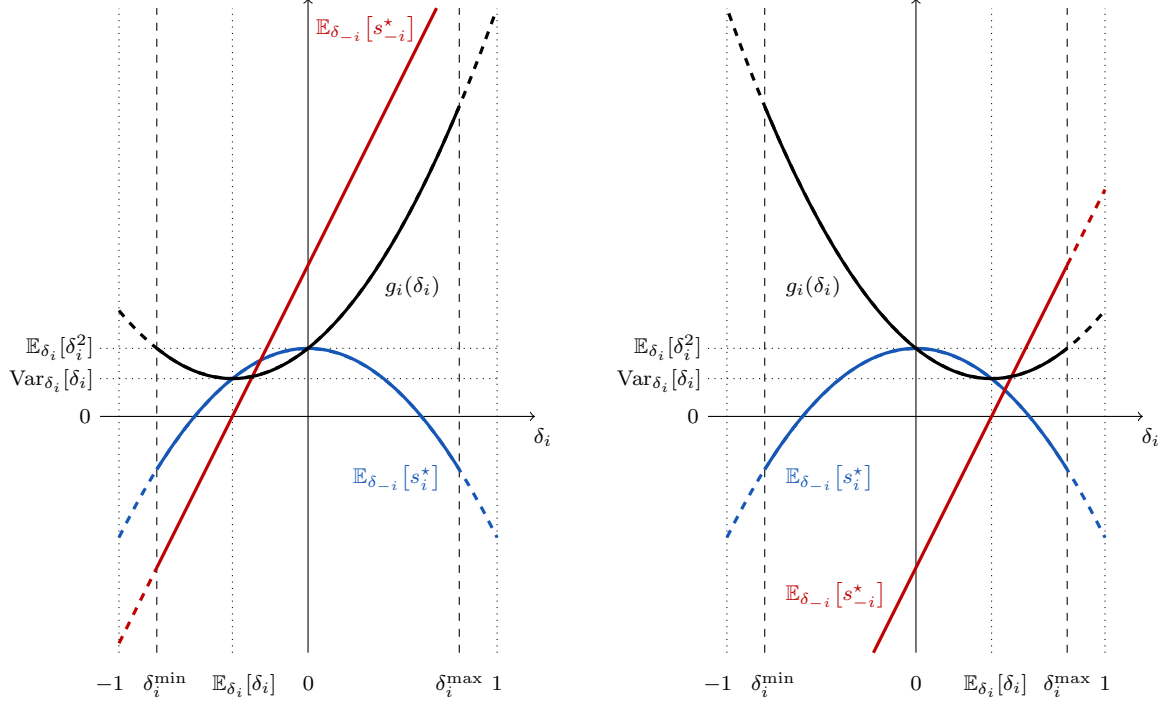


Figure 3: The utility gain $g_i(\delta_i) = \mathbb{E}_{\delta_{-i}}[s_i^*(\delta)] + \delta_i \mathbb{E}_{\delta_{-i}}[s_{-i}^*(\delta)] > 0$ that a social type δ_i interim-expects under the gamification scheme $(s_i^*)_{i=1,2}$ of equation (7), for two different type distributions: $\delta_i \in [\delta_i^{\min}, \delta_i^{\max}] = [-4/5, 4/5]$, $\mathbb{E}_{\delta_i}[\delta_i] = \mp 2/5$, and $\text{Var}_{\delta_i}[\delta_i] = 1/5$, such that $\mathbb{E}_{\delta_{-i}}[s_i^*(\delta)] = 9/25 - \delta_i^2$, $\mathbb{E}_{\delta_{-i}}[s_{-i}^*(\delta)] = 2\delta_i \pm 4/5$, and $g_i(\delta_i) = (\delta_i \pm 2/5)^2 + 1/5$.

satisfy the conditions of Proposition 4.²⁷ The corresponding gamification scheme, $s_i^*(\delta) = g_i(\delta_i) - \delta_i g_i'(\delta_i) + g_{-i}'(\delta_{-i})$, is given by

$$(7) \quad s_i^*(\delta) = \mathbb{E}_{\delta_i}[\delta_i^2] - \delta_i^2 + 2(\delta_{-i} - \mathbb{E}_{\delta_{-i}}[\delta_{-i}]).$$

The interim-expected distributive effects of $(s_i^*)_{i=1,2}$ are illustrated in Figure 3: Social types satisfying $|\delta_i| > \sqrt{\mathbb{E}_{\delta_i}[\delta_i^2]}$ incur interim-expected monetary losses, $\mathbb{E}_{\delta_{-i}}[s_i^*(\delta)] < 0$, for which they are mentally overcompensated through sufficiently strong interim-expected externalities, $\mathbb{E}_{\delta_{-i}}[s_{-i}^*(\delta)] = 2\delta_i - 2\mathbb{E}_{\delta_i}[\delta_i]$. These interim-expected monetary losses of relatively ‘strong’ social types are the source for attracting relatively selfish agents with interim-expected monetary gains: $\mathbb{E}_{\delta_{-i}}[s_i^*(\delta)] > 0$ for social types $|\delta_i| < \sqrt{\mathbb{E}_{\delta_i}[\delta_i^2]}$.

²⁷**Proof:** Obviously, $g_i'' > 0$ and $\mathbb{E}_{\delta_i}[g_i'(\delta_i)] = 0$. On the other hand, as $\text{Var}_{\delta_i}[\delta_i] = \mathbb{E}_{\delta_i}[\delta_i^2] - \mathbb{E}_{\delta_i}[\delta_i]^2$, one has $g_i(\delta_i) - \delta_i g_i'(\delta_i) = \mathbb{E}_{\delta_i}[\delta_i^2] - \mathbb{E}_{\delta_i}[\delta_i]^2 + \delta_i^2 - 2\delta_i \mathbb{E}_{\delta_i}[\delta_i] + \mathbb{E}_{\delta_i}[\delta_i]^2 - 2\delta_i(\delta_i - \mathbb{E}_{\delta_i}[\delta_i]) = \mathbb{E}_{\delta_i}[\delta_i^2] - \delta_i^2$; hence, $\mathbb{E}_{\delta_i}[g_i(\delta_i) - \delta_i g_i'(\delta_i)] = 0$.

4.3 Ex-Ante Budget-Balanced Gamification: The General Case

As utility is linear in transfers, the sufficiency result of Proposition 4 suggests the following approach to ex-ante budget-balanced gamification in networks of arbitrary size. Define transfers $s^* = (s_i^*)_{i \in \mathcal{I}} : \Delta \rightarrow \mathbb{R}^n$ by

$$(8) \quad s_i^*(\delta) = g_i(\delta_i^S) - \delta_i^S g_i'(\delta_i^S) + \sum_{j \neq i} g_j'(\delta_j^S),$$

$$(9) \quad g_i(\delta_i^S) = \text{Var}_{\delta_i}[\delta_i^S] + (\delta_i^S - \mathbb{E}_{\delta_i}[\delta_i^S])^2,$$

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

Under s^* , as I show below, agent i has the *strictly* dominant strategy to report

$$(10) \quad \delta_i^S = \sum_{j \neq i} \delta_{ij} \in (-1, n-1),$$

which I refer to as i 's *degree of pro-sociality*. When substituting for (10), then the mechanism induced by (8) and (9) is a *direct* mechanism under which each agent i has the *weakly* dominant strategy to reveal his social type δ_i truthfully.

For the mappings $(\delta_i^S)_i$, the functions $(g_i)_i$ of (9) satisfy the necessary conditions of Proposition 4:

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

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

Proposition 5 *The transfer scheme s^* defined by (8) to (10) is an ex-ante budget-balanced gamification scheme.*

Proof. *Strategy proofness:* Under s^* , each $i \in \mathcal{I}$ reports a social type $\hat{\delta}_i$, which is strategically equivalent to reporting some signal $\hat{\delta}_i^S \in \mathbb{R}$. His ex-post utility is given by

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

Hence, i maximizes $g_i(\hat{\delta}_i^S) + (\delta_i^S - \hat{\delta}_i^S) g'_i(\hat{\delta}_i^S)$ over the choice of $\hat{\delta}_i^S$. As g_i is strictly convex, i has the strictly dominant strategy to report his degree of pro-sociality: $\hat{\delta}_i^S = \delta_i^S$.

Ex-ante budget balance: By Lemma 2, $\mathbb{E}_{\delta_j} [g'_j(\delta_j^S)] = 0 = \mathbb{E}_{\delta_j} [g_j(\delta_j^S) - \delta_j^S g'_j(\delta_j^S)]$ for all $j \in \mathcal{I}$. Hence, $\mathbb{E}_\delta [s_i^*(\delta)] = 0$ and, thus, $\sum_{i \in \mathcal{I}} \mathbb{E}_\delta [s_i^*(\delta)] = 0$.

Interim-expected Pareto improvement: By Lemma 2 again, with $\hat{\delta}_i^S = \delta_i^S$ in Bayesian equilibrium, each $i \in \mathcal{I}$ derives positive interim-expected utility from unanimous participation: $\sum_{j \in \mathcal{I}} \delta_{ij} \mathbb{E}_{\delta_{-i}} [s_j^*(\delta)] = g_i(\delta_i^S) - \delta_i^S g'_i(\delta_i^S) + \left(\sum_{j \neq i} \delta_{ij} \right) g'_i(\delta_i^S) = g_i(\delta_i^S) > 0$.

■

As ex-ante budget-balanced gamification schemes provide agents with *strict* interim-expected utility gains, the mediator can even demand a uniform participation fee:

Corollary 1 *A rent-seeking mediator $M \notin \mathcal{I}$ can offer the members of \mathcal{I} to play a game that is ex-post budget-balanced among $\mathcal{I} \cup \{M\}$, strategy-proof and interim individually rational for all $i \in \mathcal{I}$, and yields M a positive ex-ante expected profit.*

Proof. For the ex-ante budget-balanced gamification scheme $(s_i^*)_{i \in \mathcal{I}}$ of Proposition 5, define the modified transfer scheme $(s_j^{**}(\delta))_{j \in \mathcal{I} \cup \{M\}}$ by $s_M^{**}(\delta) = nF - \sum_{i \in \mathcal{I}} s_i^{**}(\delta)$ and $(s_i^{**}(\delta))_{i \in \mathcal{I}} = (s_i^*(\delta) - F)_{i \in \mathcal{I}}$, at participation fee $0 < F < \sum_{j \in \mathcal{I}} \delta_{ij} \mathbb{E}_{\delta_{-i}} [s_j^*(\delta)] / \sum_{j \in \mathcal{I}} \delta_{ij}$. Then the result is immediate from the proof of Proposition 5, while M ex-ante expects a profit of $nF > 0$. ■

In the context of Corollary 1, the implicit common-knowledge assumption of Proposition 5 regarding mean and variance of every δ_i^S is sufficient but not necessary: As the gamification scheme is strategy-proof while the resulting interim-expected Pareto improvement is strict, it suffices to assume that agents and mediator have sufficiently good estimates of those means and variances.

4.4 Variants of Ex-Ante Budget-Balanced Gamification

Gamification in the manner of Proposition 5 can be implemented with qualitatively different, while ex-post payoff-equivalent, transfer schemes. The corresponding mechanisms are *indirect* in that agents do not report their social types directly but take actions (e.g., make investments, or exert efforts) from which their social types can be concluded.

Those indirect gamification schemes can be obtained by *reversing* the logic of the *revelation principle*: Suppose agent i has the strictly dominant strategy to report $\hat{\delta}_i^S = \delta_i^S$ so as to maximize utility $u_i(s^*(\hat{\delta}_i^S))$ under the *composite* game form $s^* = s[x] \circ x$, for functions $x : \mathbb{R} \rightarrow \mathbb{R}$ and $s[x] : \mathbb{R} \rightarrow \mathbb{R}^n$ with x being strictly monotone. Then i has also the strictly dominant strategy to ‘invest’ $\hat{x}_i = x(\delta_i^S)$ so as to maximize $u_i(s[x](\hat{x}_i))$ under the game form $s[x]$, with ex-post utility being left unchanged.

If x is increasing (decreasing), then ‘investments’ increase (decrease) in an agent’s pro-sociality $\delta_i^S = \sum_{j \neq i} \delta_{ij}$, suggesting that agents impose positive (negative) investment externalities on each other when playing the game induced by $s[x]$. Either scenario is feasible, as a special case of Proposition 7 below, such that gamification can be implemented with both team- or relative-performance incentives.

More generally, hybrids of team- and relative-performance incentives are feasible, too. In order to obtain (some of) those, consider a network partition into two teams: $\mathcal{I} = \mathcal{T}^1 \cup \mathcal{T}^2 = \mathcal{T}_i \cup \mathcal{T}_{-i}$, where $\mathcal{T}_i \in \{\mathcal{T}^1, \mathcal{T}^2\}$ denotes the team of agent i . Teams can be singletons, as long as $|\mathcal{T}^1| + |\mathcal{T}^2| = n$. Define transfers s^* by

$$(11) \quad s_i^*(\delta) = g_i(\delta_i^{\mathcal{T}}) - \delta_i^{\mathcal{T}} g'_i(\delta_i^{\mathcal{T}}) + \sum_{j \in \mathcal{T}_i \setminus \{i\}} g'_j(\delta_j^{\mathcal{T}}) - \sum_{j \in \mathcal{T}_{-i}} g'_j(\delta_j^{\mathcal{T}}),$$

$$(12) \quad g_i(\delta_i^{\mathcal{T}}) = \text{Var}_{\delta_i}[\delta_i^{\mathcal{T}}] + (\delta_i^{\mathcal{T}} - \mathbb{E}_{\delta_i}[\delta_i^{\mathcal{T}}])^2,$$

$$(13) \quad \delta_i^{\mathcal{T}} = \sum_{j \in \mathcal{T}_i \setminus \{i\}} \delta_{ij} - \sum_{j \in \mathcal{T}_{-i}} \delta_{ij} \in (-n, n).$$

Notice that s^* reduces to the scheme of Proposition 5 if \mathcal{T}^2 is chosen empty.

Under s^* , as I show below, each agent i has the strictly dominant strategy to report $\hat{\delta}_i^{\mathcal{T}} = \delta_i^{\mathcal{T}}$, a term that can be interpreted as i ’s relative pro-sociality towards his own team.

Proposition 6 *The transfer scheme s^* defined by (11) to (13) is an ex-ante budget-balanced gamification scheme.*

Proof. *Strategy proofness:* In $(s_j^*)_{j \in \mathcal{I}}$, the transfer components strategically relevant to $i \in \mathcal{I}$ are given by $g_i(\hat{\delta}_i^{\mathcal{T}}) - \hat{\delta}_i^{\mathcal{T}} g'_i(\hat{\delta}_i^{\mathcal{T}})$ for $j = i$, by $g'_i(\hat{\delta}_i^{\mathcal{T}})$ for $j \in \mathcal{T}_i \setminus \{i\}$, and by $-g'_i(\hat{\delta}_i^{\mathcal{T}})$ for $j \in \mathcal{T}_{-i}$. Hence, for $\delta_i^{\mathcal{T}} = \sum_{j \in \mathcal{T}_i \setminus \{i\}} \delta_{ij} - \sum_{j \in \mathcal{T}_{-i}} \delta_{ij}$, agent i maximizes $g_i(\hat{\delta}_i^{\mathcal{T}}) + (\delta_i^{\mathcal{T}} - \hat{\delta}_i^{\mathcal{T}}) g'_i(\hat{\delta}_i^{\mathcal{T}})$ over the choice of $\hat{\delta}_i^{\mathcal{T}}$. As $g''_i > 0$, agent i has the strictly dominant strategy to report $\hat{\delta}_i^{\mathcal{T}} = \delta_i^{\mathcal{T}}$.

Ex-ante budget balance: By Lemma 2, $\mathbb{E}_{\delta_j} [g'_j(\delta_j^{\mathcal{T}})] = 0 = \mathbb{E}_{\delta_j} [g_j(\delta_j^{\mathcal{T}}) - \delta_j^{\mathcal{T}} g'_j(\delta_j^{\mathcal{T}})]$. Hence, $\mathbb{E}_{\delta} [s_i^*(\delta)] = 0$ and, thus, $\sum_{i \in \mathcal{I}} \mathbb{E}_{\delta} [s_i^*(\delta)] = 0$.

Interim-expected Pareto improvement: By Lemma 2 again, each $i \in \mathcal{I}$ derives positive interim-expected utility from unanimous participation: $\sum_{j \in \mathcal{I}} \delta_{ij} \mathbb{E}_{\delta_{-i}} [s_j^*(\delta)] = g_i(\delta_i^{\mathcal{T}}) - \delta_i^{\mathcal{T}} g'_i(\delta_i^{\mathcal{T}}) + \left(\sum_{j \in \mathcal{T}_i \setminus \{i\}} \delta_{ij} - \sum_{j \in \mathcal{T}_{-i}} \delta_{ij} \right) g'_i(\delta_i^{\mathcal{T}}) = g_i(\delta_i^{\mathcal{T}}) > 0$. ■

The *reversed* logic of the *revelation principle* yields the following hybrids of team- and relative-performance incentives. (The parameter μ in Proposition 7 is introduced to facilitate the exposition in Section 5.2.)

Proposition 7 *Let $\mu = n$. The gamification scheme of Proposition 6 can be implemented with a team competition under which individual investments $x_i \geq 0$ result in payoffs $\hat{s}_i((x_j)_{j \in \mathcal{I}}) = \hat{c}_i - x_i + 2\mu\sqrt{x_i} + 2 \sum_{j \in \mathcal{T}_i \setminus \{i\}} \sqrt{x_j} - 2 \sum_{j \in \mathcal{T}_{-i}} \sqrt{x_j}$, with an appropriate constant \hat{c}_i . In the resulting game, dominant strategy investments $\hat{x}_i = (\mu + \delta_i^{\mathcal{T}})^2$ increase in $\delta_i^{\mathcal{T}}$. Payoff-equivalently, agents can compete with the members of their ‘teams’ while contributing to a public good on the network level: $\tilde{s}_i((x_j)_{j \in \mathcal{I}}) = \tilde{c}_i - x_i + 2(\mu - 1)\sqrt{x_i} - 4 \sum_{j \in \mathcal{T}_i \setminus \{i\}} \sqrt{x_j} + 2 \sum_{j \in \mathcal{I}} \sqrt{x_j}$, with an appropriate constant \tilde{c}_i . In this case, dominant-strategy investments $\tilde{x}_i = (\mu - \delta_i^{\mathcal{T}})^2$ decrease in $\delta_i^{\mathcal{T}}$.²⁸*

Proof. Observe that

$$\begin{aligned}
s_i^*(\delta) &= 2 \sum_{j \in \mathcal{T}_i \setminus \{i\}} (\delta_j^{\mathcal{T}} - \mathbb{E}_{\delta_j} [\delta_j^{\mathcal{T}}]) - 2 \sum_{j \in \mathcal{T}_{-i}} (\delta_j^{\mathcal{T}} - \mathbb{E}_{\delta_j} [\delta_j^{\mathcal{T}}]) + \mathbb{E}_{\delta_i} [(\delta_i^{\mathcal{T}})^2] - (\delta_i^{\mathcal{T}})^2 \\
&= \hat{c}_i - (\mu + \delta_i^{\mathcal{T}})^2 + 2\mu(\mu + \delta_i^{\mathcal{T}}) + 2 \sum_{j \in \mathcal{T}_i \setminus \{i\}} (\mu + \delta_j^{\mathcal{T}}) - 2 \sum_{j \in \mathcal{T}_{-i}} (\mu + \delta_j^{\mathcal{T}}) \\
&= \hat{c}_i - \hat{x}_i + 2\mu\sqrt{\hat{x}_i} + 2 \sum_{j \in \mathcal{T}_i \setminus \{i\}} \sqrt{\hat{x}_j} - 2 \sum_{j \in \mathcal{T}_{-i}} \sqrt{\hat{x}_j} \\
&= \hat{s}_i((\hat{x}_j)_j)
\end{aligned}$$

²⁸When choosing \mathcal{T}^2 empty, then $(\hat{s}_i)_i$ and $(\tilde{s}_i)_i$, respectively, provide agents with pure team- and pure relative-performance incentives. Notice that the alternative substitutions $\hat{x}_j = \mu + \delta_j^{\mathcal{T}}$ and $\tilde{x}_j = \mu - \delta_j^{\mathcal{T}}$ result in payoff-equivalent game forms, composed of linear returns and quadratic costs of ‘effort.’

when letting $\hat{c}_i = 2\mu(|\mathcal{T}_{-i}| - |\mathcal{T}_i| + 1) - \mu^2 + \mathbb{E}_{\delta_i}[(\delta_i^T)^2] - 2\sum_{j \in \mathcal{T}_i \setminus \{i\}} \mathbb{E}_{\delta_j}[\delta_j^T] + 2\sum_{j \in \mathcal{T}_{-i}} \mathbb{E}_{\delta_j}[\delta_j^T]$ and $\hat{x}_j = (\mu + \delta_j^T)^2$. Notice that $d\hat{x}_j/d\delta_j^T > 0$, since $\delta_j^T > -n = -\mu$. Likewise,

$$\begin{aligned}
s_i^*(\delta) &= 2 \sum_{j \in \mathcal{T}_i \setminus \{i\}} (\delta_j^T - \mathbb{E}_{\delta_j}[\delta_j^T]) - 2 \sum_{j \in \mathcal{T}_{-i}} (\delta_j^T - \mathbb{E}_{\delta_j}[\delta_j^T]) + \mathbb{E}_{\delta_i}[(\delta_i^T)^2] - (\delta_i^T)^2 \\
&= \tilde{c}_i - (\mu - \delta_i^T)^2 + 2\mu(\mu - \delta_i^T) - 2 \sum_{j \in \mathcal{T}_i \setminus \{i\}} (\mu - \delta_j^T) + 2 \sum_{j \in \mathcal{T}_{-i}} (\mu - \delta_j^T) \\
&= \tilde{c}_i - \tilde{x}_i + 2(\mu - 1)\sqrt{\tilde{x}_i} - 4 \sum_{j \in \mathcal{T}_i \setminus \{i\}} \sqrt{\tilde{x}_j} + 2 \sum_{j \in \mathcal{I}} \sqrt{\tilde{x}_j} \\
&= \tilde{s}_i((\tilde{x}_j)_j)
\end{aligned}$$

when letting $\tilde{c}_i = 2\mu(|\mathcal{T}_i| - |\mathcal{T}_{-i}| - 1) - \mu^2 + \mathbb{E}_{\delta_i}[(\delta_i^T)^2] - 2\sum_{j \in \mathcal{T}_i \setminus \{i\}} \mathbb{E}_{\delta_j}[\delta_j^T] + 2\sum_{j \in \mathcal{T}_{-i}} \mathbb{E}_{\delta_j}[\delta_j^T]$ and $\tilde{x}_j = (\mu - \delta_j^T)^2$. Notice that $d\hat{x}_j/d\delta_j^T < 0$, since $\delta_j^T < n = \mu$. ■

The many variants presented above suggest that there is barely a limitation to the specific forms that gamification can take. The key feature is that agents, through their efforts or investments, impose positive or negative externalities on others, while the sign of those externalities can be chosen irrespective of who (dis-)likes whom how much.

5 Ex-Post Budget Balance: Gamification in Autonomous Networks

The previous Section shows that when ex-post budget balance is replaced with ex-ante budget balance, then an ex-post materially efficient allocation function can always be Bayesian implemented with a mechanism that is interim individually rational. I have interpreted ex-ante budget balance as the participation constraint of an *outside* mediator who subsidizes or sanctions network members based on the outcomes of a game that serves as an add-on to the *terms of trade*. In this Section, I reimpose ex-post budget balance by asking: Can the role of the mediator be taken on by network members? That is, can social networks resolve their allocation problems efficiently and *autonomously*? I show that ‘yes’ for social networks of at least three agents and establish ‘Coasean implementation’ on these grounds.²⁹

²⁹With regard to the feasibility of ‘Coasean implementation’ in dyads, the results of this paper are inconclusive, although Proposition 3 shows that at least the gamification approach must fail. However, as I show in a working paper (see [Daske, 2017](#), Theorem 1), under more restrictive assumptions, any

5.1 A Coase Theorem

Consider any network of at least three agents, $n \geq 3$, and denote by $M \in \mathcal{I}$ some network member who is to take on the role of a mediator. Consequently, that mediator is now herself affected by the choice of social alternatives k .

Proposition 5 and Corollary 1 suggest the following approach to gamification in autonomous social networks: Offer the agents other than M a gamification scheme that yields each of them a zero *ex-ante* transfer; then *interim*-expected utility to M , who has no strategic role to play, is zero; as the agents other than M each obtain a strict *interim*-expected utility gain from gamification, M can extract a monetary rent by demanding a uniform participation fee from each of them, and can thereby obtain herself an *interim*-expected utility gain; if this fee is sufficiently small, the agents other than M will be willing to pay, even if they are spiteful towards M .

Therefore, define ex-post budget-balanced transfers $s^* = (s_i^*)_{i \in \mathcal{I}} : \Delta \rightarrow \mathbb{R}^n$ by

$$(14) \quad s_M^*(\delta) = - \sum_{j \neq M} s_j^*(\delta),$$

$$(15) \quad s_j^*(\delta) = -F + g_j(\delta_j^*) - \delta_j^* g'_j(\delta_j^*) + \sum_{\ell \neq j, M} g'_\ell(\delta_\ell^*), \quad \text{for } j \neq M,$$

$$(16) \quad g_j(\delta_j^*) = \text{Var}_{\delta_j}[\delta_j^*] + (\delta_j^* - \mathbb{E}_{\delta_j}[\delta_j^*])^2,$$

for a uniform participation fee $F > 0$ and appropriate functions $\delta_j^* : \Delta_j \rightarrow \mathbb{R}$.

Under s^* , as I show below, each $j \neq M$ has the *strictly* dominant strategy to report

$$(17) \quad \delta_j^* = \frac{\sum_{\ell \neq j, M} (\delta_{j\ell} - \delta_{jM})}{\delta_{jj} - \delta_{jM}}.$$

This term gives 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 $\delta_{jj} = 1$. It can be referred to as j 's *relative spite towards* M , because δ_j^* decreases in j 's degree of altruism towards M and increases in j 's pro-sociality toward the others, $\sum_{\ell \neq j, M} \delta_{j\ell}$. Notice that s^* is independent of $(\delta_{Mj})_{j \neq M}$, such that M has no strategic role to play under s^* .

Proposition 8 *The transfer scheme s^* defined by (14) to (17) is a gamification scheme in the manner of Definition 1 if $F > 0$ is chosen sufficiently small.*

Bayesian incentive-compatible, ex-post Pareto-efficient mechanism must be a gamification mechanism. Jointly with Proposition 3, this result shows that Coasean implementation is impossible in dyads.

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

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

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

Ex-post budget balance: Obvious from equation (14).

Interim-expected Pareto improvement: When substituting for δ_j^* and $\mathbb{E}_{\delta_\ell} [g'_\ell(\delta_\ell^*)] = 0 = \mathbb{E}_{\delta_\ell} [g_\ell(\delta_\ell^*) - \delta_\ell^* g'_\ell(\delta_\ell^*)]$, due to Lemma 2, then j 's interim-expected utility from s^* is

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

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

By Proposition 8, gamification schemes in the manner of Definition 1 do exist if $n \geq 3$.

The implicit common-knowledge assumption of Proposition 8 regarding mean and variance of every δ_j^* is sufficient but not necessary: As the gamification scheme is strategy-proof, while the resulting interim-expected Pareto improvement is strict, it suffices to assume that agents have sufficiently good estimates of those means and variances.

With Proposition 8 at hand, the main result of this paper can be established:

Theorem 1 (Gamification in Autonomous Networks: A Coase Theorem)

Efficiency: If endowments are sufficiently large, any social network of $n \geq 3$ agents can resolve any given allocation problem with an ex-post budget-balanced mechanism $\langle k^*, T^* \rangle$ that is Bayesian incentive-compatible, interim individually rational, and ex-post materially efficient. If $|\delta_{ij}| < \frac{1}{2n-3}$ for all i and all $j \neq i$, then $\langle k^*, T^* \rangle$ is also ex-post Pareto-efficient.

Invariance: This mechanism is invariant to changes in the status-quo social alternative k° .

Proof. Consider the gamification mechanism $\langle k^*, T^* \rangle$ consisting of the ex-post materially efficient allocation function $k^* : \Theta \rightarrow K$, $k^*(\theta) \in \arg \max_{k \in K} \sum_{i \in \mathcal{I}} \pi_i(k | \theta_i)$, and the transfer scheme $T^* = (t_i^*)_{i \in \mathcal{I}} : \Theta \times \Delta \rightarrow \mathbb{R}^n$ given by

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

where $(s_i^*)_{i \in \mathcal{I}}$ is defined by equations (14) to (17) while $\alpha^* > 0$.

Efficiency: By Propositions 1 and 8, this mechanism is Bayesian incentive-compatible. It is ex-post budget-balanced and ex-post materially efficient by construction. By Lemma 1, it is ex-post Pareto-efficient if $|\delta_{ij}| < \frac{1}{2n-3}$ for all i and all $j \neq i$. By equation (2) in the proof of Proposition 2, agent i 's interim-expected utility from unanimous participation in $\langle k^*, T^* \rangle$ is given by

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

where $\sum_{j \in \mathcal{I}} \delta_{ij} \mathbb{E}_{\delta_{-i}} [s_j^*(\delta)] > 0$ due to Proposition 8. Hence, if α^* is chosen sufficiently large, agents' interim participation constraints are satisfied for any given collection of interim reservation utilities $(U_i^\circ(\theta_i, \delta_i))_{i \in \mathcal{I}}$, as specified in Section 2.4. The assumption of sufficiently large endowments guarantees that agents can afford the respective transfers $(t_i^*(\theta, \delta))_{i \in \mathcal{I}}$ whenever those are negative.

Invariance: Except for α^* , all components of $\langle k^*, T^* \rangle$ are independent of k° . Recall that $k^\circ \in K$, while K is bounded. Due to Proposition 8, every increase in α^* allows for a strict interim-expected Pareto improvement. For given individual endowments, a maximum α^* can be chosen such that the resulting, possibly negative, transfers $(\alpha^* s_i^*(\delta))_{i \in \mathcal{I}}$ are still affordable for all agents. Hence, if endowments are sufficiently large, this maximum α^* is independent of K and, thus, independent of $k^\circ \in K$. ■

The proof of Theorem 1 does not require any specific assumptions about agents' interim participation constraints: Those can reflect any kind of consequences (that are in line with the assumptions of Section 2.4) of the gamification mechanism not finding unanimous approval. As soon as the distributive effects of the gamification scheme are amplified sufficiently, through α^* , unanimous approval is always an equilibrium outcome if the gamification mechanism is voted upon.

Coase's *invariance claim* is, to some extent, even established for status-quo *liability rules*, as defined in Section 2.4: On the one hand, *arbitrary* variations in status-quo liability rules would shift agents' interim reservation utilities *beyond any* interim-expected utilities from gamification, the latter of which are bounded by the maximum α^* feasible for given individual endowments. On the other hand, within those boundaries, any *moderate* variation in status-quo liability rules leaves the gamification mechanism with maximum α^* unaffected.

From a purely theoretical perspective, Theorem 1 refutes the assumption that extending the scope of asymmetric information must, generally, further confine the set of 'feasible' mechanisms. When extending the scope of asymmetric information from agents' *allocative* to their *distributive* preferences, a possibility result becomes tangible. Notice also that the constraint of moderately altruistic or spiteful agents, $|\delta_{ij}| < \frac{1}{2n-3}$, does not constrain the possibility result qualitatively: The sufficient (and necessary) condition regarding the distribution of social types is that those have positive variance. Even for arbitrarily large network size n , the possibility result is established if it is *not common knowledge* that agents are selfish.

5.2 Variants of Gamification in Autonomous Networks

In autonomous social networks, too, gamification schemes can be implemented with qualitatively different, while ex-post payoff-equivalent, game forms. These provide agents with team- or relative-performance incentives, or hybrids of those. This variety of feasible game forms, too, justifies the term ‘gamification’ for the incentive-theoretical concept under investigation.

The game forms of Proposition 7 are readily adapted to the gamification scheme of Proposition 8 when replacing δ_j^T with δ_j^* from equation (17) while letting $\mu = \max_{j \neq M} |\delta_j^*|$.

It is also possible to extend Proposition 8 to a scenario in which several agents collectively take on the role of the mediator, with equal shares in losses and benefits; gamification is thus feasible if as few as two agents are strategically active. This possibility implies that Theorem 1 can be established if only the interpersonal preferences within a single dyad are subject to asymmetric information.

While the above variants all leave at least one agent, the ‘mediator,’ out of play, a final variant shall illustrate how gamification in networks of *at least four* agents can involve *every* network member strategically. Theoretically, the idea is to provide two teams of agents with independent gamification schemes and let each of them take on the role of a ‘mediator’ for the respective other team. Gamification can then be implemented with a team competition that involves each and every network member.

Let $n \geq 4$, and consider a network partition into two teams: $\mathcal{I} = \mathcal{T}^1 \cup \mathcal{T}^2$, with $|\mathcal{T}^1|, |\mathcal{T}^2| \geq 2$. Denote by $\mathcal{T}_i \in \{\mathcal{T}^1, \mathcal{T}^2\}$ the team of agent i and by \mathcal{T}_{-i} the respective other team. Define transfers $s^* = (s_i^*)_{i \in \mathcal{I}} : \Delta \rightarrow \mathbb{R}^n$ by

$$(18) \quad s_i^*(\delta) = \sigma_i(\delta) - \frac{1}{|\mathcal{T}_i|} \sum_{j \in \mathcal{T}_{-i}} \sigma_j(\delta),$$

$$(19) \quad \sigma_i(\delta) = g_i(\delta_i^*) - \delta_i^* g'_i(\delta_i^*) + \sum_{j \in \mathcal{T}_i \setminus \{i\}} g'_j(\delta_j^*),$$

$$(20) \quad g_i(\delta_i^*) = \text{Var}_{\delta_i}[\delta_i^*] + (\delta_i^* - \mathbb{E}_{\delta_i}[\delta_i^*])^2,$$

$$(21) \quad \delta_i^* = \frac{\sum_{j \in \mathcal{T}_i \setminus \{i\}} \delta_{ij} - \frac{|\mathcal{T}_i| - 1}{|\mathcal{T}_{-i}|} \sum_{j \in \mathcal{T}_{-i}} \delta_{ij}}{\delta_{ii} - \frac{1}{|\mathcal{T}_{-i}|} \sum_{j \in \mathcal{T}_{-i}} \delta_{ij}}.$$

Under s^* , as I show below, each agent i has the strictly dominant strategy to report $\hat{\delta}_i^* = \delta_i^*$. As δ_i^* increases in i 's altruism toward the members of his own team, and

decreases in his average altruism towards the opposing team, δ_i^* can be interpreted as i 's *relative altruism towards his own team*. Notice that $|\delta_i^*| < 2n/(1 - \max_{j,\ell \neq j} \delta_{j\ell}^{\max})$.

Proposition 9 *The transfer scheme s^* defined by (18) to (21) is a gamification scheme in the manner of Definition 1.*

Proof. *Strategy proofness:* Under s^* , each $i \in \mathcal{I}$ reports a social type $\hat{\delta}_i$ so as to maximize ex-post utility $u_i(\hat{\delta}) = \sum_{j \in \mathcal{I}} \delta_{ij} s_j^*(\hat{\delta})$, which is strategically equivalent to reporting some signal $\hat{\delta}_i^* \in \mathbb{R}$. In s_j^* , for each $j \in \mathcal{I}$, the components strategically relevant to i are given by $g_i(\hat{\delta}_i^*) - \hat{\delta}_i^* g'_i(\hat{\delta}_i^*)$ if $j = i$, by $g'_i(\hat{\delta}_i^*)$ if $j \in \mathcal{T}_i \setminus \{i\}$, and by $-\frac{1}{|\mathcal{T}_{-i}|} [g_i(\hat{\delta}_i^*) - \hat{\delta}_i^* g'_i(\hat{\delta}_i^*)] - \frac{|\mathcal{T}_i| - 1}{|\mathcal{T}_{-i}|} g'_i(\hat{\delta}_i^*)$ if $j \in \mathcal{T}_{-i}$. For a term C independent of $\hat{\delta}_i^*$, agent i 's ex-post utility is thus given by $u_i(\hat{\delta}) = C + (\delta_{ii} - \frac{1}{|\mathcal{T}_{-i}|} \sum_{j \in \mathcal{T}_{-i}} \delta_{ij}) [g_i(\hat{\delta}_i^*) - \hat{\delta}_i^* g'_i(\hat{\delta}_i^*)] + (\sum_{j \in \mathcal{T}_i \setminus \{i\}} \delta_{ij} - \frac{|\mathcal{T}_i| - 1}{|\mathcal{T}_{-i}|} \sum_{j \in \mathcal{T}_{-i}} \delta_{ij}) g'_i(\hat{\delta}_i^*)$, where $\delta_{ii} = 1 > \frac{1}{|\mathcal{T}_{-i}|} \sum_{j \in \mathcal{T}_{-i}} \delta_{ij}$. Hence, when substituting for δ_i^* from (21), agent i maximizes $g_i(\hat{\delta}_i^*) + (\delta_i^* - \hat{\delta}_i^*) g'_i(\hat{\delta}_i^*)$ over the choice of $\hat{\delta}_i^*$. As $g''_i > 0$, agent i has the strictly dominant strategy to report $\hat{\delta}_i^* = \delta_i^*$.

Ex-post budget balance: Take the perspective of any specific agent i and observe that $\sum_{j \in \mathcal{I}} s_j^*(\delta) = \sum_{j \in \mathcal{T}_i} s_j^*(\delta) + \sum_{j \in \mathcal{T}_{-i}} s_j^*(\delta) = \sum_{j \in \mathcal{T}_i} \sigma_j(\delta) - \sum_{j \in \mathcal{T}_{-i}} \sigma_j(\delta) + \sum_{j \in \mathcal{T}_{-i}} \sigma_j(\delta) - \sum_{j \in \mathcal{T}_i} \sigma_j(\delta) = 0$.

Interim-expected Pareto improvement: By Lemma 2, $\mathbb{E}_{\delta_j} [g_j(\delta_j^*) - \delta_j^* g'_j(\delta_j^*)] = 0 = \mathbb{E}_{\delta_j} [g'_j(\delta_j^*)]$ for all $j \in \mathcal{I}$. Hence, each $i \in \mathcal{I}$ derives positive interim-expected utility from unanimous participation: $\sum_{j \in \mathcal{I}} \delta_{ij} \mathbb{E}_{\delta_{-i}} [s_j^*(\delta)] = (1 - \frac{1}{|\mathcal{T}_{-i}|} \sum_{j \in \mathcal{T}_{-i}} \delta_{ij}) g_i(\delta_i^*) > 0$. ■

By the *reversed* logic of the *revelation principle* (see Section 4.4), s^* can be implemented with a team competition under which the relative (in-)efficiency of one team's performance is rewarded (sanctioned) at the expense (to the benefit) of the other team:

Proposition 10 *Let $\mu = 2n/(1 - \max_{j,\ell \neq j} \delta_{j\ell}^{\max})$. The gamification scheme of Proposition 9 can be implemented with a competition between teams \mathcal{T}^1 and \mathcal{T}^2 under which investments $x_i \geq 0$ result in payoffs $\tilde{s}_i((x_j)_{j \in \mathcal{I}}) = \tilde{\sigma}_i((x_j)_{j \in \mathcal{T}_i}) - \frac{1}{|\mathcal{T}_i|} \sum_{j \in \mathcal{T}_{-i}} \tilde{\sigma}_j((x_j)_{j \in \mathcal{T}_{-i}})$, where $\tilde{\sigma}_i((x_j)_{j \in \mathcal{T}_i}) = \tilde{c}_i - x_i + 2\mu\sqrt{x_i} + 2 \sum_{j \in \mathcal{T}_i \setminus \{i\}} \sqrt{x_j}$, with an appropriate constant \tilde{c}_i . In this game, dominant-strategy investments $\tilde{x}_i = (\mu + \delta_i^*)^2$ increase in δ_i^* .*

Proof. Observe that

$$\begin{aligned}
\sigma_i(\delta) &= \mathbb{E}_{\delta_i}[(\delta_i^*)^2] - (\delta_i^*)^2 + 2 \sum_{j \in \mathcal{T}_i \setminus \{i\}} (\delta_j^* - \mathbb{E}_{\delta_j}[\delta_j^*]) \\
&= \tilde{c}_i - (\mu + \delta_i^*)^2 + 2\mu(\mu + \delta_i^*) + 2 \sum_{j \in \mathcal{T}_i \setminus \{i\}} (\mu + \delta_j^*) \\
&= \tilde{\sigma}_i((\tilde{x}_j)_{j \in \mathcal{T}_i})
\end{aligned}$$

when letting $\tilde{c}_i = 2\mu(1 - |\mathcal{T}_i|) - \mu^2 + \mathbb{E}_{\delta_i}[(\delta_i^*)^2] - 2 \sum_{j \in \mathcal{T}_i \setminus \{i\}} \mathbb{E}_{\delta_j}[\delta_j^*]$ and $\tilde{x}_j = (\mu + \delta_j^*)^2$. Notice that $d\tilde{x}_j/d\delta_j^* > 0$, since $\delta_j^* > -\mu = -2n/(1 - \max_{j, \ell \neq j} \delta_{j\ell}^{\max})$. ■

6 Concluding Remarks

It is considered commonplace in contemporary economics that the Coase theorem cannot hold under incomplete information and strategic interaction. The present study puts this perception into perspective. I have explored mechanism design for groups of agents whose allocative and distributive preferences are *all* subject to asymmetric information. I have shown that the Coase theorem can be established, in its typical mechanism-design interpretation, for groups of at least three agents.

The theorem can be established not despite but precisely because agents' distributive preferences are their private information. When extending the scope of asymmetric information from agents' allocative preferences to their distributive preferences, then the latter can be operationalized to satisfy agents' participation constraints. From a theoretical perspective, then, the finding refutes the assumption that extending the scope of asymmetric information must, generally, further confine the set of 'feasible' mechanisms.

I have argued that this result is more than a mathematical curiosity. In order to illustrate its economic relevance, I have interpreted the solution concept as *gamification*. In a nutshell: Attract agents' participation in an otherwise not individually rational mechanism by complementing this mechanism with a 'joyful' game, where 'joy from playing' emerges endogenously from the asymmetry of information about agents' interpersonal concerns.

As [Medema \(2020\)](#) argues convincingly, the Coase theorem can serve as a useful benchmark in economic theory. The same is true for the possibility result presented here. As I have argued, the very principle of gamification can be recognized in the manifold

approaches that real-world groups of people take in trying to resolve their conflicts of interest,—be it in the form of complementing fundraisers with auctions, raffles, or contests, or be it in the form of complementing the ‘peace-keeping process’ with Olympic games. In so far, this study provides an economical rationale to Farrell’s (1987) conviction that “people can be creative and can bypass unsatisfactory institutions.” Real-world gamification, of course, might result in merely second-best allocations (e.g., due to endowment constraints), or might come at transaction costs that are simply too high; but from a benchmark perspective, people have good reason to try.

From another angle, I have derived answers to the following questions which are also fundamental to the vast non-economical literature on gamification: What’s the economic role of people playing joyful, interactive games? And what’s the origin of such joy? To me, it is a charming side-kick of this incentive-theoretical study to take game theory back to where it started: with von Neumann’s (1928) seminal paper “Zur Theorie der Gesellschaftsspiele,” or in words to that effect, ‘toward a theory of parlor games.’

A Appendix

Proof of Lemma 1

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

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

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

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

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

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

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

References

- Andreoni, James and John Miller. 2002. “Giving according to GARP: An experimental test of the consistency of preferences for altruism.” *Econometrica* 70 (2):737–753.
- Arrow, Kenneth 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.
- Bergemann, Dirk and Stephen Morris. 2005. “Robust mechanism design.” *Econometrica* 73 (6):1771–1813.
- Bergstrom, Theodore C. 1999. “Systems of benevolent utility functions.” *Journal of Public Economic Theory* 1 (1):71–100.
- . 2017. “When was Coase right?” *Mimeo*. Available under: <https://escholarship.org/uc/item/6136k9kh>.
- Bertrand, Elodie. 2019. “Much ado about nothing? The controversy over the validity of the Coase theorem.” *European Journal of the History of Economic Thought* 26 (3):502–536.
- 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.
- Bourlès, Renaud, Yann Bramoullé, and Eduardo Perez-Richet. 2017. “Altruism in networks.” *Econometrica* 85 (2):675–689.
- Bruhin, Adrian, Ernst Fehr, and Daniel Schunk. 2019. “The many faces of human sociality: Uncovering the distribution and stability of social preferences.” *Journal of the European Economic Association* 17 (4):1025–1069.
- Charness, Gary and Matthew Rabin. 2002. “Understanding social preferences with simple tests.” *Quarterly Journal of Economics* 117 (3):817–869.
- Coase, Ronald H. 1960. “The problem of social cost.” *Journal of Law and Economics* 3 (October):1–44.
- Daske, Thomas. 2016. “Pooling hawks and doves: Interim-efficient labor contracts for other-regarding agents.” *Mimeo*. Available under: <http://hdl.handle.net/10419/145951>.
- . 2017. “Externality assessments, welfare judgments, and mechanism design.” *Mimeo*. Available under: <http://hdl.handle.net/10419/172494>.
- d’Aspremont, Claude and Louis-André Gérard-Varet. 1979. “Incentives and incomplete information.” *Journal of Public Economics* 11 (1):25–45.
- Dixit, Avinash and Mancur Olson. 2000. “Does voluntary participation undermine the Coase Theorem?” *Journal of Public Economics* 76 (3):309–335.
- Ellingsen, Tore and Elena Paltseva. 2016. “Confining the Coase theorem: Contracting, ownership, and free-riding.” *Review of Economic Studies* 83 (2):547–586.
- Farrell, Joseph. 1987. “Information and the Coase theorem.” *Journal of Economic Perspectives* 1 (2):113–129.
- Fehr, Ernst, Karla Hoff, and Mayuresh Kshetramade. 2008. “Spite and development.” *American Economic Review: Papers & Proceedings* 98 (2):494–99.
- Hamari, Juho, Kai Huotari, and Juha Tolvanen. 2015. “Gamification and economics.” In *The gameful world: Approaches, issues, applications*, edited by Steffen P. Walz and Sebastian Deterding, chap. 5. MIT Press Cambridge, MA, 139–162.
- Jackson, Matthew O and Simon Wilkie. 2005. “Endogenous games and mechanisms: Side payments among players.” *Review of Economic Studies* 72 (2):543–566.
- 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.
- Koivisto, Jonna and Juho Hamari. 2019. “The rise of motivational information systems: A review of gamification research.” *International Journal of Information Management* 45:191–210.

- Mailath, George J and Andrew Postlewaite. 1990. "Asymmetric information bargaining problems with many agents." *Review of Economic Studies* 57 (3):351–367.
- McKelvey, Richard D and Talbot Page. 2002. "Status quo bias in bargaining: An extension of the Myerson–Satterthwaite theorem with an application to the Coase theorem." *Journal of Economic Theory* 107 (2):336–355.
- Medema, Steven. 2020. "The Coase theorem at sixty." *Journal of Economic Literature*. Forthcoming.
- Myerson, Roger B. 1979. "Incentive compatibility and the bargaining problem." *Econometrica* 47 (1):61–73.
- Myerson, Roger B and Mark A Satterthwaite. 1983. "Efficient mechanisms for bilateral trading." *Journal of Economic Theory* 29 (2):265–281.
- Pelozo, 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.
- Prediger, Sebastian, Björn Vollan, and Benedikt Herrmann. 2014. "Resource scarcity and antisocial behavior." *Journal of Public Economics* 119:1–9.
- 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.
- Saijo, Tatsuyoshi and Hideki Nakamura. 1995. "The "spite" dilemma in voluntary contribution mechanism experiments." *Journal of Conflict Resolution* 39 (3):535–560.
- Seaborn, Katie and Deborah I Fels. 2015. "Gamification in theory and action: A survey." *International Journal of Human-Computer Studies* 74:14–31.
- Segal, Ilya and Michael D Whinston. 2016. "Property rights and the efficiency of bargaining." *Journal of the European Economic Association* 14 (6):1287–1328.
- von Neumann, John. 1928. "Zur Theorie der Gesellschaftsspiele." *Mathematische Annalen* 100 (1):295–320.
- 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.
- Williams, Steven R and Roy Radner. 1988. "Informational externalities and the scope of efficient dominant strategy mechanisms." Discussion paper 761, Northwestern University. Available under: <https://ideas.repec.org/p/nwu/cmsems/761.html>.
- 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.