Externality Assessments, Welfare Judgments, and Mechanism Design*

Thomas Daske †

January 30, 2018

How agents assess the (in)tangible externalities that others might impose on them can strongly influence strategic interaction. This study explores mechanism design for agents whose externality assessments and private payoffs, exclusive of externalities, are all subject to asymmetric information; utility is quasi-linear and transferable. An allocation rule will be called strongly Bayesian implementable if it is Bayesian implementable for arbitrary type distributions. Under reasonable assumptions, the following result is established: A Paretian allocation rule is strongly Bayesian implementable through budget-balanced transfers *if and only if* it maximizes aggregate private payoffs *exclusive* of externalities. The corresponding mechanism is necessarily *externality-robust* in that it leaves agents' externality assessments strategically inoperative.

The result emphasizes the critical incentive-theoretical role of the welfare judgment inherent to social choice. Strong Bayesian implementation of a welfare judgment inconsistent with *externality-ignoring utilitarianism* violates budget balance and thus entails incentive costs.

JEL classification: C70, C72, D62, D63, D82

Keywords: (behavioral) mechanism design, externalities, social preferences, robust implementation, social welfare

*This paper is based on chapter I of my doctoral thesis entitled "The Impact of Other-Regarding Preferences on Moral Hazard and Adverse Selection". For their critical remarks and helpful comments I am particularly indebted to Christian Feilcke, Michael Kurschilgen, Christoph March, Marco Sahm, and Robert von Weizsäcker. I am also grateful for helpful comments by participants of the Asian Meeting of the Econometric Society in Hong Kong, the European Meeting on Game Theory (SING13) in Paris, the Annual Congress of the Association for Public Economic Theory in Paris, the Annual Congress of the International Institute of Public Finance in Tokyo, the European Winter Meeting of the Econometric Society in Barcelona, and seminars at the University of Bamberg and the Technical University of Munich.

†TUM School of Management, Technical University of Munich, Arcisstr. 21, 80333 Munich, Germany. Email: thomas.daske@tum.de. Phone: +49-89-289-25713. Fax: +49-89-289-25702.

1 Introduction

The theory of mechanism design is devoted to the question of how to render collective action efficient if the agents involved hold private information—typically about their valuations of tangible assets. In many economic environments, however, this challenge is exacerbated by the fact that agents do also hold private information about their (rational or *ex post* irrational) assessments of the externalities that others might impose on them. These externalities can be tangible, for instance due to spillover effects between firms or local economies, or intangible—if agents derive (dis)utility directly from how tangible assets are *distributed* among them.[1]

This study explores ex post Pareto-efficient (and, thus, ex post budget-balanced), mechanism design for two agents whose externality assessments and private payoffs are all subject to asymmetric information. Each agent's utility is taken as a weighted sum of her own payoff and her opponent's payoff, while the real-valued weight on the latter determines an agent's externality assessment, her *externality type*. An agent's payoff is additively separable in a numeraire good (money) and a payoff component (subject to the economic environment under investigation) which is taken affine in her real-valued *payoff type*. An agent's externality type and payoff type are exogenously given, not perfectly correlated, and private information; types are independent across agents.—The central question is to what extent collective action can, or must, condition on agents' externality assessments in order to be ex post Pareto-efficient and incentivize agents to reveal their preferences truthfully.

With externalities taken tangible, the model captures bargaining between competing nations about scarce resources, with each nation having its private expectations about the benefit from that resource but also having its private expectations about the threat of the resource when being in the other nation's hands. Another example are neighboring municipalities negotiating harmonized public expenditure if there are spillovers from locally provided public goods.[2]

[1]Agents might also derive (dis)utility from—or change their preferences according to—the *process* through which final allocations are realized; see, e.g., Bowles and Hwang (2008). This line of reasoning is beyond the scope of the present study. Here, I take intangible externalities as outcome-dependent, being determined by agents' judgments about the final distribution of wealth.

[2]This scenario has been analyzed by Harstad (2007), under the assumption of commonly known externalities though.

With externalities taken intangible, the model captures other-regarding preferences in the form of altruism, spite, or status. Altruism and spite are often deployed in the range of family economics. The model captures bargaining problems like inheritance disputes and divorce battles, given that family members are privately informed about their valuations of the goods at stake (their payoff types) and about the extent to which they have come to despise each other (their externality types). On the other hand, empirical studies have found that many, if not all, people care about their relative standing in society.[3] The model applies, for instance, to bargaining situations the outcomes of which will affect the income opportunities of bargainers, provided that the respective income expectations (payoff types) as well as relative standing considerations (externality types) are private information.

In order to implement ex post Pareto-efficient allocations, a mechanism provides agents with incentives such that they truthfully reveal their preferences in equilibrium.—What is the appropriate equilibrium concept if there is asymmetric information about externality as well as payoff types?—This question is central not only to the design but also to the applicability of mechanisms, since different equilibrium concepts differ in their *common knowledge* assumptions about agents' information, preferences, and rationality. The aim to successively weaken common knowledge assumptions in game theory is sometimes referred to as the 'Wilson doctrine':

> "Game theory has a great advantage in explicitly analyzing the consequences of trading rules that presumably are really common knowledge; it is deficient to the extent it assumes other features to be common knowledge, such as one player's probability assessment about another's preferences or information.
> I foresee the progress of game theory as depending on successive reductions in the base of common knowledge required to conduct useful analyses of practical problems. Only by repeated weakening of common knowledge assumptions will the theory approximate reality." (Wilson, 1987)

The equilibrium concept with the weakest information requirement is that of dominant strategy implementation in the manner of Vickrey (1961), Clarke (1971), and Groves

[3]For empirical evidence on status considerations see, e.g., Clark, Frijters, and Shields (2008), Heffetz and Frank (2008), Tran and Zeckhauser (2012), and the survey by Weiss and Fershtman (1998). For a theoretical foundation of status preferences see, e.g., Bisin and Verdier (1998).

(1973). Unfortunately, with externalities, whether private information or common knowledge, dominant strategy implementation is typically not feasible. A weaker notion is that of ex post implementation, which requires that truthful revelation is each agent's best strategy in response to each and every realization of her opponents' (truthfully revealed) types. Under ex post implementation, knowledge of type distributions is not required. However, even if externality types are common knowledge, the imposition of budget balance restricts its applicability immensely.[4] The equilibrium concept I deploy is that of Bayesian implementation, which requires that truthful revelation maximizes each agent's von Neumann-Morgenstern (interim) expected utility provided all other agents reveal their types truthfully.[5] As Bayesian implementation collides with the 'Wilson doctrine', I will put emphasis on how the assumption of common knowledge about the distribution of externality types can (and even must) be avoided.

In the environment under investigation, a mechanism specifies an *allocation rule*, specifying collective action based on the agents' preferences, and a *transfer scheme*, incentivizing agents to reveal those preferences. The challenge involved with private information about externality assessments is the following: Suppose the allocation rule conditions on externality assessments. Then the transfer scheme must elicit payoff types as well as externality types. However, through their externality assessments, agents internalize the distributive effects of the transfer scheme itself. Hence, the mechanism itself might deliver incentives to misrepresent preferences. Bayesian incentive compatibility demands counterbalance of these adverse incentives. Requiring budget balance further restricts the domain of adequate transfer schemes.

I show that the welfare judgment inherent to an allocation rule is decisive for *whether* and *how* that allocation rule can be Bayesian implemented with a budget-balanced mechanism. Specifically, I obtain the following results.

By Proposition 2, the renowned 'expected externality mechanism' (AGV-mechanism), due to Arrow (1979) and d'Aspremont and Gérard-Varet (1979), Bayesian implements in a budget-balanced way the allocation rule that, for each realization of types, maximizes aggregate private payoffs exclusive of externalities. These allocations are Pareto-efficient

[4] Bergemann and Morris (2005) show that Bayesian implementable allocation rules can, in many cases, no longer be ex post implemented when requiring budget balance.

[5] To be sure, the term *type* refers to the *pair* of an agent's externality and payoff type. Notice that a property which is possessed by the class of Bayesian implementable allocation rules is necessarily possessed by allocation rules that are ex post implementable.

if each agent's marginal utility from her own payoff exceeds her marginal (dis)utility from her opponent's payoff. The AGV-mechanism is *externality-robust* in the sense that it requires neither agents nor the mechanism designer to have any knowledge of the statistical distribution of externality types.

I then ask for conditions that an ex post Pareto-efficient allocation rule *must* satisfy in order to be Bayesian implementable with a budget-balanced mechanism. For this purpose, I introduce the notions of *sensitive allocation rules* and *strong Bayesian implementability.*

An allocation rule will be called *sensitive* if, in the respective economic environment, it is the unique maximizer of a social-welfare measure which satisfies the Pareto property. Furthermore, a sensitive allocation rule is required to be *non-constant* in payoff types and to be *symmetric* in the sense that the effect of an increase in one agent's externality or payoff type on the other agent's private payoff is qualitatively similar for both agents. *Non-constancy* reflects strong, or 'sensitive', welfare judgments of the mechanism designer, as it implies that she is not indifferent to even small changes in payoff types.[6]

An allocation rule will be called *strongly* Bayesian implementable if, for *any* set of (non-degenerate) type distributions, there exists a mechanism that Bayesian implements it. That is, strongly Bayesian implementable allocation rules may not condition on the specifics of type distributions. This requirement accounts for the 'Wilson doctrine' in so far as it avoids making common knowledge assumptions from the outset. By Proposition 2, the allocation rule associated with externality-ignoring utilitarianism is sensitive and strongly Bayesian implementable.

I show that the converse of Proposition 2 is also true if one asks for strong Bayesian implementation of sensitive allocation rules, which yields the following equivalence (Theorem 1): A sensitive allocation rule can be strongly Bayesian implemented with a budget-balanced mechanism *if and only if* it maximizes aggregate private payoffs exclusive of externalities; I call the welfare judgment inherent to these allocations *externality-ignoring utilitarianism.* The respective mechanism takes the form of the AGV-mechanism.

Loosely speaking, a sensitive allocation rule can be strongly Bayesian implemented in a budget-balanced way if and only if it results from a form of utilitarianism that approves individual achievements but ignores 'help' or 'harm' from others. Implementation of

[6] Examples of sensitive social-welfare measures are given by utilitarian welfare, either *inclusive* or *exclusive* of externalities. When restricting the economic environment to linear utilities and non-negative externalities, several classical social-welfare measures qualify as sensitive; they are listed in Proposition 1.

a welfare judgment inconsistent with externality-ignoring utilitarianism violates budget balance and thus requires either an external source of money or that 'money is burned'. The associated costs can be interpreted as the incentive costs of the welfare judgment. Furthermore, costless implementation of a sensitive allocation rule requires an externality-robust mechanism; all mechanisms having this property are of AGV-type. That is, the requirement of externality robustness does not only serve the purpose of satisfying the 'Wilson doctrine' but is even necessary from a welfarist point of view.

Finally, I outline the antagonistic roles of welfare judgments and budget balance. Theorem 2 shows that, even with asymmetric information about externality assessments, nearly any welfare judgment can be Bayesian implemented if one waives the requirement of budget balance. On the other hand, with privately observed payoff types but common knowledge of externality types, nearly any allocation rule can be Bayesian implementable in a budget-balanced way (Theorem 3). Hence, it is not externality assessments per se that render welfare judgments critical but rather the asymmetry of information about them combined with the efficiency request of budget balance.

The paper proceeds as follows. Section 2 reviews the related literature. Section 3 outlines the basic model. Section 4 identifies conditions that are necessary and sufficient for ex post Pareto-efficient Bayesian implementation; the central result on the allocative implications of welfare judgments is obtained. Section 5 expands the central result to social-welfare measures that incorporate the redistributive effects of the transfer scheme itself. Section 6 interprets results for strategic bargaining under incomplete information. Section 7 concludes.

2 Related Literature

This study bridges three strands of literature: those on robust implementation, implementation in the presence of externalities, and the measurement of social welfare.

In order to come by the criticism pointed at unrealistic common knowledge assumptions (Wilson, 1987), many studies have characterized conditions under which Bayesian implementable allocation rules are ex post or even dominant strategy implementable.[7] Jehiel et al. (2006) consider a model framework that entails the one presented here, with

[7]E.g., Mookherjee and Reichelstein (1992), Dasgupta and Maskin (2000), Bergemann and Morris (2005, 2011), Chung and Ely (2007), Gershkov et al. (2013).

the exception that agents do not internalize the distributive effects of transfers. They show that only those allocation rules can be ex post implemented that appoint the very same allocation for any realization of types.

Several studies have explored ex post or Bayesian implementation under the assumption that externalities are common knowledge.[8] The studies closest to the present one are those of Jehiel and Moldovanu (2001) and Bierbrauer and Netzer (2016).

Jehiel and Moldovanu (2001) investigate the feasibility of 'efficient' Bayesian implementation in the presence of (allocative or informative) externalities.[9] In their model, each agent i is privately informed about her private payoff, exclusive of externalities, and about the externality she imposes on another agent j. Agent j's externality type, in the language of the present study, is assumed common knowledge. The present study expands the work of Jehiel and Moldovanu (2001) to the extent that it takes the externality of i on j as a composite of two pieces of private information, one held by i, the other one held by j. However, in order to expose the critical role of welfare judgments, attention is restricted to more specific economic environments.

Bierbrauer and Netzer (2016) explore the design of mechanisms for agents who exhibit intention-based social preferences in the manner of Rabin (1993). In a novel attempt, they allow for private information about social types and identify *sufficient* conditions for externality-robust Bayesian implementation.[10] The present study, in a slightly different setting, supplements their work by asking for *necessary and sufficient* conditions for budget-balanced Bayesian implementation.

This study bridges normative and positive theory based on incentive theoretical grounds. With regard to 'efficient' implementation, the mechanism design literature typically takes a utilitarian view. In the presence of externalities, the allocation rule is typically taken to maximize aggregate private payoffs *inclusive* of externalities (e.g., Jehiel and Moldovanu, 2001). Theorem 1 provides a positive rationale for the utilitarian view in mechanism design theory, however complemented with the somewhat surprising qualification that, if externality assessments are private information, externalities must

[8]E.g., Jehiel, Moldovanu, and Stacchetti (1996, 1999), Jehiel and Moldovanu (2001), Goeree et al. (2005), Kucuksenel (2012), Lu (2012), and Tang and Sandholm (2012).

[9]They refer to an allocation as 'efficient' if it maximizes aggregate payoffs *inclusive* of externalities.

[10]Bierbrauer et al. (2017) provide empirical evidence for the relevance of 'social-preference robust' implementation in the range of bilateral trade as well as income taxation. Bartling and Netzer (2016) follow a similar line for the design of auctions if bidders are privately informed about their spiteful preferences.

be *ignored* in order to achieve both incentive compatibility and budget balance. Other foundations of utilitarianism have been provided on axiomatic, or say normative, grounds (e.g., Harsanyi, 1955, d'Aspremont and Gevers, 1977, and Maskin, 1978) and in the range of decision-making under ignorance (e.g., Maskin, 1979).

Theorem 1 is bad news for the proponents of non-utilitarian measures of social welfare.[11] Examples for alternative concepts are the maximin principle of Rawls (1971), the CES welfare measures proposed by Arrow (1973), and welfare measures that explicitly condition on indices of inequality (e.g., on the inequality index of Atkinson, 1970).[12] When interpreting agents' externality assessments as their individual, privately known preferences for redistribution, Theorem 1 implies that incentive-compatible redistributive policies (beyond externality-ignoring utilitarianism) come at a price, embodied in the violation of budget balance.[13]

More generally, Theorems 1 to 3 suggest that theories of 'efficient' implementation depend critically on their underlying welfare judgments, and their results might not pertain when introducing asymmetric information about agents' externality assessments. This particularly involves theories of optimal taxation based on "social utility weights". From another angle, the result contributes to the growing field of behavioral mechanism design:[14] With regard to their externality assessments, agents might not be able to fully process the information available (e.g., McFadden, 2009). Other agents might believe that there are externalities even though there are objectively none. Likewise, agents might be overly optimistic, or pessimistic, about how the well-being of others would affect themselves (e.g., Hirschman and Rothschild, 1973). It seems plausible in all these cases that social planners should not condition their policies on such 'behavioral' externality assessments, and that mechanisms designed to implement 'efficient' allocations should be externality-robust.

[11]For critical reflections of utilitarianism see, e.g., Posner (1979) and Sen (1973, 1979).

[12]For a discussion of the CES welfare measures see also Sen (1974).

[13]Saez and Stantcheva (2016), for instance, characterize optimal taxation under non-utilitarian social-welfare measures—in the (somewhat unrealistic) absence of externalities.

[14]E.g., Glazer and Rubinstein (1998), Cabrales and Serrano (2011), de Clippel (2014), Bierbrauer and Netzer (2016), and Bartling and Netzer (2016).

3 The Model

3.1 The Basic Setup

There is an interval $K = [k^{\min}, k^{\max}]$ of social alternatives, with $k^{\min} < k^{\max}$, and there are two agents, indexed by $i \in \{1, 2\}$. The agent other than i is denoted by $-i$. From alternative $k \in K$ and a monetary transfer $t_i \in \mathbb{R}$, agent i gains a *private payoff*

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

where the functions $v_i : K \to [0, \infty)$ and $h_i : K \to \mathbb{R}$ are twice continuously differentiable and satisfy $\partial^2 \pi_i(k, t_i \,|\, \theta_i)/\partial k^2 < 0$ for all i, k, and θ_i; furthermore, either $dv_i/dk > 0$ for all k and i, or $dv_i/dk < 0$ for all k and i. Agent i's *payoff type* θ_i is drawn from an interval $\Theta_i = (\theta_i^{\min}, \theta_i^{\max}) \subset (0, \infty)$. From the allocation of payoffs, i gains *utility*

$$u_i(k, t_i, t_{-i}, \theta_{-i} \,|\, \theta_i, \delta_i) \quad = \quad \pi_i(k, t_i \,|\, \theta_i) + \delta_i \cdot \pi_{-i}(k, t_{-i} \,|\, \theta_{-i}),$$

where i's *externality type* δ_i is drawn from an interval $\Delta_i = (\delta_i^{\min}, \delta_i^{\max}) \subset [-1, 1]$. Externality types take absolute values smaller than one, such that each agent's marginal utility from her own payoff exceeds her marginal (dis)utility from her opponent's payoff. The pair (θ_i, δ_i) will be referred to as i's *type*. For convenience, define also $\pi_i(k \,|\, \theta_i) = \theta_i v_i(k) + h_i(k)$ and $u_i(k, \theta_{-i} \,|\, \theta_i, \delta_i) = \pi_i(k \,|\, \theta_i) + \delta_i \pi_{-i}(k \,|\, \theta_{-i})$.

3.2 Information and Incentives

The functions $\{v_i, h_i\}$ are common knowledge. Payoff types, θ_i, and externality types, δ_i, are private information and are distributed according to continuous density functions $f_i : \Theta_i \to (0, \infty)$ and $g_i(\,\cdot \,|\, \theta_i) : \Delta_i \to (0, \infty)$; that is, an agent's externality type may correlate with that agent's payoff type, not perfectly though. Denote by H_i the joint c.d.f. of i's type. While types are private information, type distributions, $\{H_i\}$, are common knowledge. H_1 and H_2 are stochastically independent.

Denote by Θ and Δ, respectively, the Cartesian products $\Theta_1 \times \Theta_2$ and $\Delta_1 \times \Delta_2$, and let $\theta = (\theta_1, \theta_2)$ and $\delta = (\delta_1, \delta_2)$. For a random variable $X : \Theta \times \Delta \to \mathbb{R}$, denote by $\mathbb{E}_{\theta_i,\delta_i}\big[X(\theta, \delta)\big]$ the expected value of X for a given type $(\theta_{-i}, \delta_{-i})$:[15]

A *direct revelation mechanism* involves the agents in a strategic game of incomplete information. In this game, agents are asked to report their types truthfully.[16] Based on their reports, a social alternative will be implemented and transfers will be made. Specifically, the mechanism is defined by an *allocation rule* $k : \Theta \times \Delta \to K$ and a *transfer scheme* $T = (t_1, t_2) : \Theta \times \Delta \to \mathbb{R}^2$. In what follows, I restrict attention to transfer schemes that are continuous on the externality-type space Δ.[17] An allocation rule k is said to be *Bayesian implementable* if there exists a transfer scheme $T = (t_1, t_2)$ such that truthful revelation maximizes each agent's interim expected utility provided the respective other agent reveals her type truthfully:

$$
\begin{aligned}
(\theta_1, \delta_1) \;&\in\; \arg\max_{\hat\theta_1,\hat\delta_1} \mathbb{E}_{\theta_2,\delta_2}\Big[u_1\big(k(\hat\theta_1,\hat\delta_1,\theta_2,\delta_2), t_1(\hat\theta_1,\hat\delta_1,\theta_2,\delta_2), t_2(\hat\theta_1,\hat\delta_1,\theta_2,\delta_2), \theta_2 \,\big|\, \theta_1,\delta_1\big)\Big],\\
(\theta_2, \delta_2) \;&\in\; \arg\max_{\hat\theta_2,\hat\delta_2} \mathbb{E}_{\theta_1,\delta_1}\Big[u_2\big(k(\theta_1,\delta_1,\hat\theta_2,\hat\delta_2), t_2(\theta_1,\delta_1,\hat\theta_2,\hat\delta_2), t_1(\theta_1,\delta_1,\hat\theta_2,\hat\delta_2), \theta_1 \,\big|\, \theta_2,\delta_2\big)\Big].
\end{aligned}
$$

The mechanism is said to be *ex post budget-balanced* if transfers satisfy $t_1 + t_2 = 0$ for each realization of types.

3.3 Further Definitions

The following two definitions restrict the set of allocation rules that are to be considered in the next Sections. For that purpose, denote by $\operatorname{sgn} : \mathbb{R} \to \{-1, 0, 1\}$ the *sign function.*[18]

Definition 1 (Sensitivity)
Let $W : \mathbb{R}^4 \to \mathbb{R}$ be twice partially continuously differentiable, and define $V : K \to \mathbb{R}$ by $V(k) = W\big(\pi_1(k \,|\, \theta_1)\,,\, \delta_1\pi_2(k \,|\, \theta_2)\,,\, \pi_2(k \,|\, \theta_2)\,,\, \delta_2\pi_1(k \,|\, \theta_1)\big)$. Then W is said to be a sensitive social-welfare measure if it has the following properties.

(i) Payoff sensitivity: $\partial W(\pi_1, \delta_1\pi_2, \pi_2, \delta_2\pi_1)/\partial\pi_i > 0$ for each $i \in \{1, 2\}$.

[15] Likewise, denote by $\mathbb{E}_{\theta_i}\big[Y(\theta)\big]$ the expected value of $Y : \Theta \to \mathbb{R}$ for a given payoff type θ_{-i}.

[16] By the *revelation principle*, which applies to the present setup (Myerson, 1979), there is no loss of generality in identifying message sets, from which agents draw their reports, with agents' type sets.

[17] This assumption more restrictive than necessary. However, some assumption about the 'smoothness' of T with respect to externality types will be required in the proof of Lemma 1 below.

[18] For $x \in \mathbb{R}$, the sign of x is defined as $\operatorname{sgn}(x) = 1$ if $x > 0$, $\operatorname{sgn}(x) = -1$ if $x < 0$, and $\operatorname{sgn}(0) = 0$.

(ii) Pareto property: If there exist $k_1, k_2 \in K$ and $i \in \{1,2\}$ such that $u_i(k_1, \theta_{-i} \,|\, \theta_i, \delta_i) > u_i(k_2, \theta_{-i} \,|\, \theta_i, \delta_i)$ while $u_{-i}(k_1, \theta_i \,|\, \theta_{-i}, \delta_{-i}) \geq u_{-i}(k_2, \theta_i \,|\, \theta_{-i}, \delta_{-i})$, then $V(k_1) > V(k_2)$.

(iii) Implication of payoff-type sensitive allocations and symmetric effects: There exists a partially continuously differentiable allocation rule $k^* : \Theta \times \Delta \to K$ satisfying $k^*(\theta, \delta) = \arg\max_{k \in K} V(k)$,

$$1 = \operatorname{sgn}\left(\frac{\partial v_1(k^*)}{\partial \theta_2}\right) \cdot \operatorname{sgn}\left(\frac{\partial v_2(k^*)}{\partial \theta_1}\right), \quad \textit{and} \tag{1}$$

$$0 = \operatorname{sgn}\left(\frac{\partial \pi_1(k^* \,|\, \theta_1)}{\partial \delta_2}\right) - \operatorname{sgn}\left(\frac{\partial \pi_2(k^* \,|\, \theta_2)}{\partial \delta_1}\right). \tag{2}$$

The allocation rule k^* is said to be sensitive.

Sensitive social-welfare measures account separately for private payoffs, $\pi_i(k \,|\, \theta_i)$, and externalities, $\delta_i \pi_{-i}(k \,|\, \theta_{-i})$.[19] This serves the purpose of clearly isolating the extent to which Pareto-efficient allocation rules may condition on agents' externality assessments if they are to be Bayesian implemented through budget-balanced transfers.

By condition (i), a marginal increase in an agent's private payoff contributes to social welfare. Conditions (ii) and (iii), jointly, ensure that the allocation rule unambiguously specifies *some* allocation on the ex post Pareto frontier.[20] Given that individual utility is affine in transfers, full ex post Pareto efficiency is realized if transfers are budget-balanced.

Conditions (1) and (2), in essence, require that the effect of an increase in one agent's payoff type or externality type on the other agent's payoff is qualitatively the same across agents. As the functions v_i are assumed to be either strictly increasing or strictly decreasing, condition (1) requires in particular that sensitive allocation rules are responsive to changes in payoff types. By contrast, they are allowed to not respond to changes in externality types.[21]

Several well-known social-welfare measures qualify as sensitive.

[19] This specification is without loss of generality in that it takes payoffs, π_i, and externality types, δ_i, as independent variables. For example, $V(k) = (1+\delta_1^2)\pi_2 + (1+\delta_2^2)\pi_1$ can be written as $V(k) = (\pi_1) + (\pi_2) + \frac{(\delta_2\pi_1)}{(\pi_1)}(\delta_2\pi_1) + \frac{(\delta_1\pi_2)}{(\pi_2)}(\delta_1\pi_2)$.

[20] Furthermore, condition (iii) requires sensitive social-welfare measures as well as the economic environment to allow for interior solutions to $\max_{k \in K} V(k)$. Hence, k^* satisfies the first- and second-order conditions $dV(k^*(\theta,\delta))/dk = 0$ and $d^2V(k^*(\theta,\delta))/dk^2 < 0$ for all $(\theta,\delta) \in \Theta \times \Delta$.

[21] Notice that condition (2) precludes the dictatorial social-welfare measure, $W = u_i(k, \theta_{-i} \,|\, \theta_i, \delta_i)$ for some agent i, from being sensitive, since then $\partial \pi_i(k^* \,|\, \theta_i)/\partial \delta_{-i} = 0$, whereas $\partial \pi_{-i}(k^* \,|\, \theta_{-i})/\partial \delta_i \neq 0$.

Proposition 1 *Each of the following social-welfare measures W is sensitive if the economic environment is such that W induces a unique partially continuously differentiable allocation rule $k^*(\theta, \delta) = \arg\max_{k \in K} V(k)$ satisfying $\partial k^* / \partial \theta_i \neq 0$ for all $(\theta, \delta) \in \Theta \times \Delta$ and all $i \in \{1, 2\}$.*

(i) Externality-ignoring utilitarianism: $W = \pi_1(k \,|\, \theta_1) + \pi_2(k \,|\, \theta_2)$.

(ii) Bentham utilitarianism: $W = u_1(k, \theta_2 \,|\, \theta_1, \delta_1) + u_2(k, \theta_1 \,|\, \theta_2, \delta_2)$.

If the economic environment is restricted to $h_i = 0$ and $\Delta_i \subset (0, 1)$ for all $i \in \{1, 2\}$, then the following social-welfare measures are sensitive.[22]

(iii) "Social utility weights", inclusive of externalities:
$W = \alpha_1 u_1(k, \theta_2 \,|\, \theta_1, \delta_1) + \alpha_2 u_2(k, \theta_1 \,|\, \theta_2, \delta_2)$, *with* $\alpha_1, \alpha_2 > 0$.

(iv) The Nash product, inclusive of externalities:
$W = u_1(k, \theta_2 \,|\, \theta_1, \delta_1) \cdot u_2(k, \theta_1 \,|\, \theta_2, \delta_2)$.

(v) CES welfare (Arrow, 1973), inclusive of externalities:
$W = \left[[u_1(k, \theta_2 \,|\, \theta_1, \delta_1)]^{-\rho} + [u_2(k, \theta_1 \,|\, \theta_2, \delta_2)]^{-\rho} \right]^{-\frac{1}{\rho}}$, *with* $\rho \in (-1, \infty) \setminus \{0\}$.

Proof. Externality-ignoring utilitarianism will be addressed separately in Proposition 2. Proofs are straightforward for (ii) and (iii) and are omitted therefore. See the Appendix for (iv) and (v). ■

By means of the next definition, I restrict attention to those social-welfare measures that do not vary with (higher moments of) type distributions and still imply ex post allocations that are Bayesian implementable, irrespective of type distributions.

Definition 2 (Strong Bayesian Implementability)
An allocation rule $k^* : \Theta \times \Delta \to K$ is said to be strongly Bayesian implementable if it is Bayesian implementable for any set of (non-degenerate) type distributions, $\{H_1, H_2\}$.

Strong Bayesian implementability does not require the mechanism as a whole to be independent from type distributions. It rather makes a qualitative distinction between 'means' (the transfer scheme) and 'ends' (the allocation rule). The welfare judgment

[22] One can show that the externality-ignoring versions of these welfare measures satisfy conditions (i) and (iii) of Definition 1 but might yield allocations that are not Pareto-efficient.

inherent to this concept is that ex post allocations ought not depend on what agents' types *could have been* but only on what agents' types *are*.[23]

4 The Incentive Costs of Welfare Judgments

This Section proves the following theorem (employing Propositions 2 to 4) and discusses it from various angles (through Theorems 2 and 3).

Theorem 1 *A sensitive allocation rule $k^* : \Theta \times \Delta \to K$ is strongly Bayesian implementable through budget-balanced transfers if and only if it maximizes aggregate private payoffs exclusive of externalities: $k^*(\theta, \delta) = \arg\max_{k \in K} \pi_1(k \,|\, \theta_1) + \pi_2(k \,|\, \theta_2)$ for all (θ, δ); in particular, k^* is independent from externality types: $k^* = k^*(\theta)$.*

The budget-balanced transfer schemes $T^* = (t_1^*, t_2^*)$ that (ordinarily) Bayesian implement k^* are of AGV-type: For reported types $(\hat{\theta}, \hat{\delta}) \in \Theta \times \Delta$, transfers are given by

$$
\begin{aligned}
(3) \qquad t_1^*(\hat{\theta}, \hat{\delta}) \;&=\; \mathbb{E}_{\theta_2}\big[\pi_2(k^*(\hat{\theta}_1, \theta_2) \,|\, \theta_2)\big] - \mathbb{E}_{\theta_1}\big[\pi_1(k^*(\theta_1, \hat{\theta}_2) \,|\, \theta_1)\big] + s(\hat{\theta}, \hat{\delta}), \\
(4) \qquad t_2^*(\hat{\theta}, \hat{\delta}) \;&=\; \mathbb{E}_{\theta_1}\big[\pi_1(k^*(\theta_1, \hat{\theta}_2) \,|\, \theta_1)\big] - \mathbb{E}_{\theta_2}\big[\pi_2(k^*(\hat{\theta}_1, \theta_2) \,|\, \theta_2)\big] - s(\hat{\theta}, \hat{\delta}),
\end{aligned}
$$

where $s : \Theta \times \Delta \to \mathbb{R}$ must be chosen such that $\mathbb{E}_{\theta_{-i}, \delta_{-i}}[s(\theta, \delta)]$ is constant on $\Theta_i \times \Delta_i$ for all $i \in \{1, 2\}$.[24]

By Theorem 1, Bayesian implementation of a welfare judgment inconsistent with externality-ignoring utilitarianism violates budget balance and thus entails incentive costs.

In the following, I refer to the mechanisms (k^*, T^*) specified by Theorem 1 as *AGV-type mechanisms* (after Arrow, 1979, and d'Aspremont and Gérard-Varet, 1979). *Ex interim*, AGV-type mechanisms leave externality assessments strategically inoperative. If the distribution of externality types is not common knowledge, one can let $s = 0$.

The sufficient conditions of Theorem 1 are to be addressed first.

Proposition 2 *Suppose the allocation rule $k^* : \Theta \to K$ is partially continuously differentiable and satisfies $k^*(\theta) = \arg\max_{k \in K} \pi_1(k \,|\, \theta_1) + \pi_2(k \,|\, \theta_2)$ and $\partial k^* / \partial \theta_i \neq 0$ for all*

[23] An established social-welfare measure that does condition on type distributions is the *generalized Nash product* of Harsanyi and Selten (1972).

[24] Such functions s can be smooth and non-constant; for example, $s(\theta, \delta) = (\theta_1 - \mathbb{E}_{\theta_1}[\theta_1])(\theta_2 - \mathbb{E}_{\theta_2}[\theta_2]) + (\delta_1 - \mathbb{E}_{\delta_1}[\delta_1])(\delta_2 - \mathbb{E}_{\delta_2}[\delta_2])$.

$\theta \in \Theta$. *Then k^* is sensitive and can be strongly Bayesian implemented through AGV-type transfers.*[25]

Proof. Regarding the sensitivity of k^*, I focus on the Pareto property. Verification of the remaining properties of Definition 1 is straightforward.

Suppose there exists an allocation $k'(\theta, \delta)$ that, for some types (θ, δ), Pareto-improves upon $k^*(\theta)$. Since $\pi_i(k \,|\, \theta_i)$ is concave, $\pi_1(k'(\theta, \delta) \,|\, \theta_1) + \pi_2(k'(\theta, \delta) \,|\, \theta_2) < \pi_1(k^*(\theta) \,|\, \theta_1) + \pi_2(k^*(\theta) \,|\, \theta_2)$. Suppose agent 1 suffers the (weakly) greater loss in private payoffs. Then the differences $d_i = \pi_i(k^*(\theta) \,|\, \theta_i) - \pi_i(k'(\theta, \delta) \,|\, \theta_i)$ satisfy $d_1 > 0$ and $d_1 \geq d_2 > -d_1$. Since $\delta_1 \in (-1, 1)$,

$$u_1(k'(\theta, \delta), \theta_2 \,|\, \theta_1, \delta_1) - u_1(k^*(\theta), \theta_2 \,|\, \theta_1, \delta_1) \;=\; -(d_1 + \delta_1 d_2) \;<\; 0.$$

Hence, agent 1 is worse of under $k'(\theta, \delta)$ than under $k^*(\theta)$; a contradiction.[26]

Under AGV-type mechanisms, which are budget-balanced, and under the assumption that agent 2 reveals her type (θ_2, δ_2) truthfully, agent 1 reports $(\hat{\theta}_1, \hat{\delta}_1)$ so as to maximize her interim expected utility. Without loss of generality, normalize $s(\hat{\theta}, \hat{\delta}) = 0$. By equations (3) and (4),

$$\begin{aligned} \mathbb{E}_{\theta_2,\delta_2}\big[u_1(\,\cdot\,)\big] \;&=\; \mathbb{E}_{\theta_2}\big[\pi_1(k^*(\hat{\theta}_1, \theta_2) \,|\, \theta_1) + \pi_2(k^*(\hat{\theta}_1, \theta_2) \,|\, \theta_2)\big] \\ &\quad -(1-\delta_1)\mathbb{E}_{\theta_1,\theta_2}\big[\pi_1(k^*(\theta_1, \theta_2) \,|\, \theta_1)\big], \end{aligned}$$

where the second term in the last line is independent from $\hat{\theta}_1$. If truthfully reporting θ_1 is strictly inferior to some report $\hat{\theta}_1 \neq \theta_1$, then there must exist some θ_2 such that

$$\pi_1(k^*(\hat{\theta}_1, \theta_2) \,|\, \theta_1) + \pi_2(k^*(\hat{\theta}_1, \theta_2) \,|\, \theta_2) \;>\; \pi_1(k^*(\theta_1, \theta_2) \,|\, \theta_1) + \pi_2(k^*(\theta_1, \theta_2) \,|\, \theta_2),$$

which contradicts the definition of k^*. Hence, agent 1 has no incentive to misreport her payoff type. Obviously, she has no incentive to misreport her externality type. By symmetry, agent 2 can do no better than reporting (θ_2, δ_2). As the argument holds for any

[25] That AGV-type mechanisms are Bayesian incentive-compatible for other-regarding, spiteful agents has been shown earlier by Bartling and Netzer (2016) and Bierbrauer and Netzer (2016).

[26] This line of reasoning shows also that for "excessive" externalities, $|\delta_i| > 1$, externality-ignoring utilitarianism might not yield Pareto-efficient allocations.

set of type distributions, AGV-type transfers *strongly* Bayesian implement k^*. ■

The next two Propositions give proof of the necessary conditions of Theorem 1. They successively constrain the domain of sensitive allocation rules that are strongly Bayesian implementable through budget-balanced transfers. A Lemma eases the exposition.

Lemma 1 *Suppose the partially continuously differentiable allocation rule $k^* : \Theta \times \Delta \to K$ is strongly Bayesian implementable through budget-balanced transfers. Then k^* satisfies*

$$(5) \qquad (1-\delta_i)\frac{\partial v_i(k^*(\theta,\delta))}{\partial \delta_i} \quad = \quad \left[\frac{d\pi_i(k^*(\theta,\delta)\,|\,\theta_i)}{dk} + \frac{d\pi_{-i}(k^*(\theta,\delta)\,|\,\theta_{-i})}{dk}\right]\frac{\partial k^*(\theta,\delta)}{\partial \theta_i}$$

for all $(\theta,\delta) \in \Theta \times \Delta$ and all $i \in \{1,2\}$. If k^* is independent from externality types, $k^* = k^*(\theta)$, then k^* is (ordinarily) Bayesian implementable through budget-balanced transfers only if the transfer to each agent i satisfies

$$\mathbb{E}_{\theta_{-i},\delta_{-i}}\big[t_i(\theta,\delta)\big] \quad = \quad \alpha_i + \mathbb{E}_{\theta_{-i}}\big[\pi_{-i}(k^*(\theta)\,|\,\theta_{-i})\big]$$

for all $(\theta_i,\delta_i) \in \Theta_i \times \Delta_i$ and some constant $\alpha_i \in \mathbb{R}$.

Proof. See the Appendix. ■

In light of the second part of the Lemma, Proposition 3 indicates also that the desired mechanisms leave agents' externality assessments strategically inoperative.

Proposition 3 *A sensitive allocation rule $k^* : \Theta \times \Delta \to K$ is strongly Bayesian implementable through budget-balanced transfers only if it is independent from externality types.*

Proof. Let $k^* : \Theta\times\Delta \to K$ be the sensitive allocation rule that corresponds to a sensitive social-welfare measure $W : \mathbb{R}^4 \to \mathbb{R}$, and suppose k^* is strongly Bayesian implementable through budget-balanced transfers. It has to be shown that $\partial k^*/\partial\delta_i = 0$.

For $x \in \mathbb{R}^4$ and $j = \{1,...,4\}$, write $W_j(x) = \partial W(x)/\partial x_j$, and define

$$(6) \qquad W_j \quad = \quad W_j\big(\pi_1(k^*\,|\,\theta_1), \delta_1\pi_2(k^*\,|\,\theta_2), \pi_2(k^*\,|\,\theta_2), \delta_2\pi_1(k^*\,|\,\theta_1)\big).$$

By Definition 1, k^* satisfies the first-order condition

$$(7) \qquad 0 = \frac{dV(k^*)}{dk} = [W_1 + \delta_2 W_4]\frac{d\pi_1(k^* \mid \theta_1)}{dk} + [W_3 + \delta_1 W_2]\frac{d\pi_2(k^* \mid \theta_2)}{dk},$$

where $\partial W/\partial \pi_1 = W_1 + \delta_2 W_4 > 0$ and $\partial W/\partial \pi_2 = W_3 + \delta_1 W_2 > 0$. By Lemma 1, k^* satisfies also

$$(8) \qquad (1-\delta_1)\frac{\partial v_1(k^*)}{\partial \delta_1} = \left[\frac{d\pi_1(k^* \mid \theta_1)}{dk} + \frac{d\pi_2(k^* \mid \theta_2)}{dk}\right]\frac{\partial k^*}{\partial \theta_1},$$

$$(9) \qquad (1-\delta_2)\frac{\partial v_2(k^*)}{\partial \delta_2} = \left[\frac{d\pi_1(k^* \mid \theta_1)}{dk} + \frac{d\pi_2(k^* \mid \theta_2)}{dk}\right]\frac{\partial k^*}{\partial \theta_2}.$$

Substituting (7) into (8) and (9) yields

$$(10) \qquad (1-\delta_1)\frac{\partial v_1(k^*)}{\partial \delta_1} = \left[1 - \frac{W_1 + \delta_2 W_4}{W_3 + \delta_1 W_2}\right]\frac{d\pi_1(k^* \mid \theta_1)}{dk}\frac{\partial k^*}{\partial \theta_1},$$

$$(11) \qquad (1-\delta_2)\frac{\partial v_2(k^*)}{\partial \delta_2} = \left[1 - \frac{W_3 + \delta_1 W_2}{W_1 + \delta_2 W_4}\right]\frac{d\pi_2(k^* \mid \theta_2)}{dk}\frac{\partial k^*}{\partial \theta_2}.$$

On the other hand, as $\partial k^*/\partial \theta_i \neq 0$ by Definition 1(iii), identities (8) and (9) imply that

$$(12) \qquad (1-\delta_1)\frac{\partial v_1(k^*)}{\partial \delta_1}\frac{\partial k^*}{\partial \theta_2} = (1-\delta_2)\frac{\partial v_2(k^*)}{\partial \delta_2}\frac{\partial k^*}{\partial \theta_1}.$$

As $\delta_i < 1$ and $dv_i/dk \neq 0$ by assumption, identity (12) implies that $\partial k^*/\partial \delta_1 = 0$ if and only if $\partial k^*/\partial \delta_2 = 0$.

Suppose $\partial k^*(\theta, \delta)/\partial \delta_i \neq 0$ for some (θ, δ) and all i. Then each of the factors on the right-hand sides of (10) and (11) is non-zero. In this case, (10) and (11) yield

$$(13) \qquad \frac{(W_3 + \delta_1 W_2)(1-\delta_1)\frac{\partial v_1(k^*)}{\partial \delta_1}}{\frac{d\pi_1(k^* \mid \theta_1)}{dk}\frac{\partial k^*}{\partial \theta_1}} = [(W_3 + \delta_1 W_2) - (W_1 + \delta_2 W_4)]$$

$$= -\left[(W_1 + \delta_2 W_4) - (W_3 + \delta_1 W_2)\right]$$

$$= -\frac{(W_1 + \delta_2 W_4)(1-\delta_2)\frac{\partial v_2(k^*)}{\partial \delta_2}}{\frac{d\pi_2(k^* \mid \theta_2)}{dk}\frac{\partial k^*}{\partial \theta_2}}.$$

Rearranging (13), while writing $\frac{\partial v_i(k^*)}{\partial \delta_i} = \frac{dv_i(k^*)}{dk}\frac{\partial k^*}{\partial \delta_i}$, yields the identity

$$(14) \qquad \begin{aligned} & (W_3 + \delta_1 W_2)(1-\delta_1)\frac{dv_1(k^*)}{dk}\frac{\partial k^*}{\partial \delta_1}\frac{d\pi_2(k^* \mid \theta_2)}{dk}\frac{\partial k^*}{\partial \theta_2} \\ = \; & -(W_1 + \delta_2 W_4)(1-\delta_2)\frac{dv_2(k^*)}{dk}\frac{\partial k^*}{\partial \delta_2}\frac{d\pi_1(k^* \mid \theta_1)}{dk}\frac{\partial k^*}{\partial \theta_1}. \end{aligned}$$

Since $(W_1 + \delta_2 W_4), (W_3 + \delta_1 W_2), (1-\delta_i) > 0$, applying the sign function to (14) yields

$$(15) \qquad \operatorname{sgn}\left(\frac{d\pi_2(k^* \mid \theta_2)}{dk}\frac{\partial k^*}{\partial \delta_1}\frac{dv_1(k^*)}{dk}\frac{\partial k^*}{\partial \theta_2}\right) = -\operatorname{sgn}\left(\frac{d\pi_1(k^* \mid \theta_1)}{dk}\frac{\partial k^*}{\partial \delta_2}\frac{dv_2(k^*)}{dk}\frac{\partial k^*}{\partial \theta_1}\right).$$

By Definition 1(iii), $\operatorname{sgn}(\partial v_1(k^*)/\partial \theta_2) \cdot \operatorname{sgn}(\partial v_2(k^*)/\partial \theta_1) = 1$, such that (15) implies

$$(16) \qquad \operatorname{sgn}\left(\frac{\partial \pi_2(k^* \mid \theta_2)}{\partial \delta_1}\right) = -\operatorname{sgn}\left(\frac{\partial \pi_1(k^* \mid \theta_1)}{\partial \delta_2}\right).$$

Equation (16) contradicts condition (2) of Definition 1, unless $\partial \pi_i(k^* \mid \theta_i)/\partial \delta_{-i} = 0$ for all i. Suppose $\partial \pi_1(k^* \mid \theta_1)/\partial \delta_2 = 0$; then multiplying (10) with $\partial k^*(\theta, \delta)/\partial \delta_2$ implies that

$$(17) \qquad (1-\delta_1)\frac{\partial v_1(k^*)}{\partial \delta_1}\frac{\partial k^*(\theta,\delta)}{\partial \delta_2} = 0.$$

As $\delta_i < 1$ and $dv_i/dk \neq 0$, (17) yields $\frac{\partial k^*(\theta,\delta)}{\partial \delta_1}\frac{\partial k^*(\theta,\delta)}{\partial \delta_2} = 0$, such that $\partial k^*/\partial \delta_i = 0$, due to (12) and the reasoning thereafter. Altogether, $\partial k^*/\partial \delta_i = 0$. ■

The next Proposition indicates that, from a social point of view, the desired mechanisms treat agents' private payoffs as perfect substitutes.

Proposition 4 *A sensitive allocation rule $k^* : \Theta \to K$, which is independent from externality types, is strongly Bayesian implementable through budget-balanced transfers only if $k^*(\theta) = \arg\max_{k \in K} \pi_1(k \mid \theta_1) + \pi_2(k \mid \theta_2)$ for all $\theta \in \Theta$. Any budget-balanced transfer scheme that (ordinarily) Bayesian implements k^* is necessarily of AGV-type.*

Proof. If k^* is independent from externality types, identity (5) of Lemma 1 reads

$$0 = \left[\frac{d\pi_i(k^*(\theta) \mid \theta_i)}{dk} + \frac{d\pi_{-i}(k^*(\theta) \mid \theta_{-i})}{dk}\right]\frac{\partial k^*(\theta)}{\partial \theta_i}.$$

By Definition 1(iii), $\partial k^*/\partial \theta_i \neq 0$ for all θ_i. As $\pi_i(k \mid \theta_i)$ is concave in the choice of k, $k^*(\theta) = \arg\max_{k \in K} \pi_1(k \mid \theta_1) + \pi_2(k \mid \theta_2)$.

Suppose there exists a budget-balanced transfer scheme $T^* = (t_1^*, t_2^*) : \Theta \times \Delta \to \mathbb{R}^2$ that Bayesian implements k^*. Notice that one can always write

$$\begin{aligned} t_1^*(\hat\theta, \hat\delta) &= \mathbb{E}_{\theta_2}\big[\pi_2(k^*(\hat\theta_1, \theta_2) \,|\, \theta_2)\big] - \mathbb{E}_{\theta_1}\big[\pi_1(k^*(\theta_1, \hat\theta_2) \,|\, \theta_1)\big] + s_1(\hat\theta, \hat\delta), \\ t_2^*(\hat\theta, \hat\delta) &= \mathbb{E}_{\theta_1}\big[\pi_1(k^*(\theta_1, \hat\theta_2) \,|\, \theta_1)\big] - \mathbb{E}_{\theta_2}\big[\pi_2(k^*(\hat\theta_1, \theta_2) \,|\, \theta_2)\big] + s_2(\hat\theta, \hat\delta), \end{aligned}$$

for appropriate functions $s_1, s_2 : \Theta \times \Delta \to \mathbb{R}$ satisfying $s_1 + s_2 = 0$ on $\Theta \times \Delta$. But then, for each $i \in \{1, 2\}$ and all $(\theta_i, \delta_i) \in \Theta_i \times \Delta_i$,

$$\mathbb{E}_{\theta_{-i},\delta_{-i}}\big[t_i^*(\theta, \delta)\big] = \mathbb{E}_{\theta_{-i}}\big[\pi_{-i}(k^*(\theta) \,|\, \theta_{-i})\big] - \mathbb{E}_{\theta_i,\theta_{-i}}\big[\pi_i(k^*(\theta) \,|\, \theta_i)\big] + \mathbb{E}_{\theta_{-i},\delta_{-i}}\big[s_i(\theta, \delta)\big].$$

On the other hand, for $k^* : \Theta \to K$, Lemma 1 states that

$$\mathbb{E}_{\theta_{-i},\delta_{-i}}\big[t_i^*(\theta, \delta)\big] = \alpha_i + \mathbb{E}_{\theta_{-i}}\big[\pi_{-i}(k^*(\theta) \,|\, \theta_{-i})\big]$$

for all $(\theta_i, \delta_i) \in \Theta_i \times \Delta_i$ and some constant $\alpha_i \in \mathbb{R}$. Jointly, these identities imply that $\mathbb{E}_{\theta_{-i},\delta_{-i}}\big[s_i(\theta, \delta)\big] = \alpha_i + \mathbb{E}_{\theta_i,\theta_{-i}}\big[\pi_i(k^*(\theta) \,|\, \theta_i)\big]$ for all (θ_i, δ_i), which is constant on $\Theta_i \times \Delta_i$. Hence, (k^*, T^*) is of *AGV-type*. ■

Propositions 2 to 4 give proof of Theorem 1.

The next result shows that the set of Bayesian implementable welfare judgments expands substantially if one waives the requirement of budget balance.

Theorem 2 *If one waives budget balance, then any twice continuously differentiable allocation rule $k^* : \Theta \times \Delta \to K$ satisfying $\min_{\theta_i,\delta_i} \frac{\partial}{\partial \theta_i}\mathbb{E}_{\theta_{-i},\delta_{-i}}\big[v_i(k^*(\theta, \delta))\big] > 0$ for all $(\theta_i, \delta_i) \in \Theta_i \times \Delta_i$ and all $i \in \{1, 2\}$ is strongly Bayesian implementable.*[27]

[27]The sufficient condition is fairly weak; as indicated by condition (36) in the proof of Lemma 1, any Bayesian implementable allocation rule necessarily satisfies $\frac{\partial}{\partial \theta_i}\mathbb{E}_{\theta_{-i},\delta_{-i}}\big[v_i(k^*(\theta, \delta))\big] \geq 0$.

Proof. Be k^* as described, with $\beta_i > 0$ for $\beta_i = \min_{\theta_i,\delta_i} \frac{\partial}{\partial\theta_i}\mathbb{E}_{\theta_{-i},\delta_{-i}}\big[v_i(k^*(\theta,\delta))\big]$. For functions $p_i : \Delta_i \to \mathbb{R}$, define transfers $T^* = (t_1^*, t_2^*)$ by

$$\begin{aligned} t_i^*(\hat{\theta},\hat{\delta}) \;=\;& p_i(\hat{\delta}_i) - \hat{\delta}_i \frac{\partial p_i(\hat{\delta}_i)}{\partial\hat{\delta}_i} + \int_{\theta_i^{\min}}^{\hat{\theta}_i} \mathbb{E}_{\theta_{-i},\delta_{-i}}\big[v_i(k^*(s,\hat{\theta}_{-i},\hat{\delta}))\big]\,ds \\ &+ \frac{\partial p_{-i}(\hat{\delta}_{-i})}{\partial\hat{\delta}_{-i}} + \frac{\partial}{\partial\hat{\delta}_{-i}} \int_{\theta_{-i}^{\min}}^{\hat{\theta}_{-i}} \mathbb{E}_{\theta_i,\delta_i}\big[v_{-i}(k^*(\hat{\theta}_i,s,\hat{\delta}))\big]\,ds \\ &- \hat{\delta}_i \frac{\partial}{\partial\hat{\delta}_i} \int_{\theta_i^{\min}}^{\hat{\theta}_i} \mathbb{E}_{\theta_{-i},\delta_{-i}}\big[v_i(k^*(s,\hat{\theta}_{-i},\hat{\delta}))\big]\,ds \\ &- \mathbb{E}_{\theta_{-i},\delta_{-i}}\big[\pi_i(k^*(\hat{\theta},\hat{\delta}) \,|\, \hat{\theta}_i)\big] - \mathbb{E}_{\theta_i,\delta_i}\big[\pi_i(k^*(\hat{\theta},\hat{\delta}) \,|\, \hat{\theta}_i)\big]. \end{aligned}$$

Then T^* strongly Bayesian implements k^* if the functions p_i are chosen such that the following condition holds for all (θ_i,δ_i) and all i:

$$(18) \qquad \frac{\Big[\frac{\partial}{\partial\delta_i}\mathbb{E}_{\theta_{-i},\delta_{-i}}\big[v_i(k^*(\theta,\delta))\big]\Big]^2}{\frac{\partial}{\partial\theta_i}\mathbb{E}_{\theta_{-i},\delta_{-i}}\big[v_i(k^*(\theta,\delta))\big]} \;<\; \frac{\partial^2}{\partial\delta_i^2}\left[p_i(\delta_i) + \int_{\theta_i^{\min}}^{\theta_i} \mathbb{E}_{\theta_{-i},\delta_{-i}}\big[v_i(k^*(s,\theta_{-i},\delta))\big]\,ds\right].$$

For example, one can choose $p_i(\delta_i) = \frac{1}{2}c_i\delta_i^2$, with

$$(19) \qquad c_i \;=\; \gamma_i - \min_{\theta_i,\delta_i} \frac{\partial^2}{\partial\delta_i^2} \int_{\theta_i^{\min}}^{\theta_i} \mathbb{E}_{\theta_{-i},\delta_{-i}}\big[v_i(k^*(s,\theta_{-i},\delta))\big]\,ds$$

for some constant γ_i satisfying $\beta_i \cdot \gamma_i > \max_{\theta_i,\delta_i}\Big[\frac{\partial}{\partial\delta_i}\mathbb{E}_{\theta_{-i},\delta_{-i}}\big[v_i(k^*(\theta,\delta))\big]\Big]^2$.[28] For an extensive proof of this claim as well as a derivation of T^*, see the Appendix. ■

The final result of this Section emphasizes the critical role of information about agents' externality assessments.

Theorem 3 *Suppose externality types are common knowledge. Then any differentiable allocation rule* $k^* : \Theta \times \Delta \to K$ *satisfying* $\frac{\partial}{\partial\theta_i}\mathbb{E}_{\theta_{-i}}\big[v_i(k^*(\theta,\delta))\big] \geq 0$ *for all* $(\theta_i,\delta) \in \Theta_i \times \Delta$ *and all* $i \in \{1,2\}$ *is strongly Bayesian implementable through budget-balanced transfers.*[29]

[28] The latter maximum value exists, since v_i and k^* are continuously differentiable and K is compact.
[29] As implied by condition (36) in the proof of Lemma 1, the sufficient condition of Theorem 3 is even necessary to render k^* (ordinarily) Bayesian incentive compatible.

Proof. Be k^* as described. For agents $i \in \{1,2\}$ of commonly known externality types $\delta = (\delta_1, \delta_2)$, define the functions $S_i : \Theta \times \Delta \to \mathbb{R}$ by

$$S_i(\hat{\theta}, \delta) \;=\; \int_{\theta_i^{\min}}^{\hat{\theta}_i} v_i(k^*(s, \hat{\theta}_{-i}, \delta))\, ds - \pi_i(k^*(\hat{\theta}, \delta) \,|\, \hat{\theta}_i) - \delta_i \pi_{-i}(k^*(\hat{\theta}, \delta) \,|\, \hat{\theta}_{-i}).$$

Then the budget-balanced transfer scheme $T^* = (t_1^*, t_2^*)$ defined by

$$\begin{aligned} t_1^*(\hat{\theta}, \delta) \;&=\; \frac{1}{1-\delta_1}\Big[S_1(\hat{\theta}, \delta) - \mathbb{E}_{\theta_1}\big[S_1(\theta_1, \hat{\theta}_2, \delta)\big]\Big] \\ &\quad + \frac{1}{1-\delta_2}\Big[-S_2(\hat{\theta}, \delta) + \mathbb{E}_{\theta_2}\big[S_2(\hat{\theta}_1, \theta_2, \delta)\big]\Big], \\ t_2^*(\hat{\theta}, \delta) \;&=\; \frac{1}{1-\delta_1}\Big[-S_1(\hat{\theta}, \delta) + \mathbb{E}_{\theta_1}\big[S_1(\theta_1, \hat{\theta}_2, \delta)\big]\Big] \\ &\quad + \frac{1}{1-\delta_2}\Big[S_2(\hat{\theta}, \delta) - \mathbb{E}_{\theta_2}\big[S_2(\hat{\theta}_1, \theta_2, \delta)\big]\Big] \end{aligned}$$

strongly Bayesian implements k^*. For an extensive proof of this claim as well as a derivation of T^*, see the Appendix. ■

By Theorem 3, it is not externality assessments per se that constrains the implementability of welfare judgments, but rather the asymmetry of information about them.

5 Holistic Social Welfare Measures

Up to this point, I have focused on the welfare judgment inherent to the allocation rule. How does the result of Theorem 1 expand to welfare judgments that are *holistic* in the sense that they incorporate the distributive effects of the transfer scheme?

To answer this question, consider a differentiable social-welfare measure $W : \mathbb{R}^4 \to \mathbb{R}$, and define $V : K \times \mathbb{R}^2 \to \mathbb{R}$ by

$$V(k, t_1, t_2) \;=\; W\big(\pi_1(k, t_1 \,|\, \theta_1)\,,\, \delta_1 \pi_2(k, t_2 \,|\, \theta_2)\,,\, \pi_2(k, t_2 \,|\, \theta_2)\,,\, \delta_2 \pi_1(k, t_1 \,|\, \theta_1)\big).$$

Suppose W is an *ex post* social-welfare measure in that it is invariant to changes in type distributions. Assume also that W is *payoff-sensitive* in that it satisfies

$$(20) \qquad \frac{\partial W}{\partial \pi_1} \;=\; W_1 + \delta_2 W_4 \;>\; 0 \qquad \text{and} \qquad \frac{\partial W}{\partial \pi_2} \;=\; W_3 + \delta_1 W_2 \;>\; 0.$$

The *social-choice rule* (k^*, t_1^*, t_2^*), with allocation rule $k^* : \Theta \times \Delta \to K$ and transfer scheme $(t_1^*, t_2^*) : \Theta \times \Delta \to \mathbb{R}^2$, is budget-balanced and maximizes V if and only if $t_2^* = -t_1^*$ and $(k^*, t_1^*) = \arg\max_{(k,t_1)\in K\times\mathbb{R}} V(k, t_1, -t_1)$. When assuming that W and the economic environment allow for an interior solution, (k^*, t_1^*) satisfies the first-order conditions

$$(21) \quad 0 = \frac{\partial V(k^*, t_1^*, -t_1^*)}{\partial k} = [W_1 + \delta_2 W_4]\frac{d\pi_1(k^* \,|\, \theta_1)}{dk} + [W_3 + \delta_1 W_2]\frac{d\pi_2(k^* \,|\, \theta_2)}{dk},$$

$$(22) \quad 0 = \frac{\partial V(k^*, t_1^*, -t_1^*)}{\partial t_1} = [W_1 + \delta_2 W_4] - [W_3 + \delta_1 W_2].$$

Jointly, conditions (20), (21), and (22) imply that the socially efficient allocation rule is consistent with externality-ignoring utilitarianism: $k^*(\theta, \delta) = \arg\max_{k\in K} \pi_1(k \,|\, \theta_1) + \pi_2(k \,|\, \theta_2)$. In other words, under holistic social-welfare measures, social choice differs merely in the extent of redistributive taxation. The problem thus reduces to the question: Which welfare judgments yield redistributive tax tariffs that are Bayesian incentive-compatible?

By Theorem 1, the socially efficient allocation rule k^* can be Bayesian implemented through budget-balanced transfers if and only if transfers are of AGV-type. As AGV-type transfers vary with changes in (payoff-)type distributions, the *ex post* social-welfare measure W must be invariant to changes in transfers. That is, agents' private payoffs must be perfect substitutes from a social planner's point of view. This proves the following theorem.

Theorem 4 *A budget-balanced social-choice rule that is interior solution to the maximization of a differentiable, payoff-sensitive ex post social-welfare measure W is Bayesian incentive-compatible if and only if W is consistent with externality-ignoring utilitarianism. The respective mechanism is of AGV-type.*

Under holistic social-welfare measures, too, incentives must be externality-robust. The assumptions of Theorem 4 apply in particular to the measures listed in Proposition 1.

A final remark can be made on Rawlsian justice (Rawls, 1971). While the non-differentiable Rawlsian maximin welfare function does not meet with the requirements of the above analyses, Theorem 1 still proves useful to obtain the following result.

Proposition 5 *A budget-balanced social-choice rule satisfying Rawls' maximin principle, inclusive or exclusive of externalities, is not Bayesian incentive-compatible.*

Proof. Consider the maximin principle inclusive of externalities and define the budget-balanced social-choice rule $(k^*, t^*, -t^*)$ through

$$(k^*, t^*) \quad = \quad \arg \max_{(k,t) \in K \times \mathbb{R}} \min\{\pi_1(k, t \,|\, \theta_1) + \delta_1 \pi_2(k, -t \,|\, \theta_2) \,;\, \pi_2(k, -t \,|\, \theta_2) + \delta_2 \pi_1(k, t \,|\, \theta_1)\}.$$

As individual utility is affine in transfers, t^* must equalize utilities:

$$\pi_1(k^* \,|\, \theta_1) + \delta_1 \pi_2(k^* \,|\, \theta_2) + (1 - \delta_1) t^* \quad = \quad \pi_2(k^* \,|\, \theta_2) + \delta_2 \pi_1(k^* \,|\, \theta_1) - (1 - \delta_2) t^*.$$

Therefore, $t^* = \frac{1-\delta_1}{2-\delta_1-\delta_2} \pi_2(k^* \,|\, \theta_2) - \frac{1-\delta_2}{2-\delta_1-\delta_2} \pi_1(k^* \,|\, \theta_1)$, and utilities are given by

$$u_i \quad = \quad \frac{1 - \delta_1 \delta_2}{2 - \delta_1 - \delta_2} \left[\pi_1(k^* \,|\, \theta_1) + \pi_2(k^* \,|\, \theta_2) \right].$$

Hence, $k^* = \arg \max_{k \in K} \, \pi_1(k \,|\, \theta_1) + \pi_2(k \,|\, \theta_2)$, since $\delta_i \in (-1, 1)$. By Theorem 1, transfers must be of AGV-type in order to Bayesian implement k^*. As t^* is not of AGV-type, the social-choice rule $(k^*, t^*, -t^*)$ is not Bayesian incentive-compatible.

When letting $\delta_i = 0$ in the above line of reasoning, the proof is obtained for the maximin principle exclusive of externalities. ■

6 Bargaining with Side-Payments

This Section applies the above results to the following question: How, by what means and what ends, do two agents come to an agreement upon the division of a given 'pie' which is currently owned by neither of them? With 'means' I refer to the bargaining procedure, with 'ends' to those allocations that are feasible under that procedure. In particular, how are means and ends restricted if agents are privately informed about how they value shares of pie as well as how they assess the externalities, tangible or intangible, that they might impose on each other?

The bargaining literature can be broadly separated into two strands. The 'means'-strand, initiated by Rubinstein (1982), starts out from specific bargaining procedures and takes ends as equilibrium outcomes of the respective non-cooperative games.[30] The 'ends'-strand, initiated earlier by Nash (1950), is often referred to as "axiomatic bargaining"

[30] See Ausubel, Cramton, and Deneckere (2002) for a survey on non-cooperative bargaining under incomplete information.

and asks how sharing rules, or *bargaining solutions*, are restricted, if not determined, by collections of reasonable, pre-specified properties.[31] Naturally, these properties are preference-contingent, which makes preference revelation a critical issue.

In the following, I discuss the bargaining problem from a mechanism design perspective, for "this allows us to identify properties shared by all Bayesian equilibria of any bargaining game" (Ausubel, Cramton, and Deneckere, 2002). I ask which bargaining solutions are strongly Bayesian implementable if utility is transferable by means of budget-balanced side-payments between agents.[32]

Consider two agents, 1 and 2, who bargain over the division of a pie of size 1. Specify the model framework of Section 3 by letting $v_1(k) = v(k)$ and $v_2(k) = v(1-k)$, where $v : [0,1] \to [0,1]$ is twice continuously differentiable and satisfies $v(0) = 0$, $v(1) = 1$, $v' > 0$, and $v'' < 0$. Let $h_i = 0$ for all i. From their shares k and $1-k$, respectively, and transfers t_1 and t_2, agents 1 and 2 draw ex post utilities

$$\begin{aligned} u_1(k, t_1, t_2) &= \big[\theta_1 v(k) + t_1\big] + \delta_1 \cdot \big[\theta_2 v(1-k) + t_2\big], \\ u_2(k, t_1, t_2) &= \big[\theta_2 v(1-k) + t_2\big] + \delta_2 \cdot \big[\theta_1 v(k) + t_1\big]. \end{aligned}$$

By Theorem 1, the only *sensitive* bargaining solution $k : \Theta \times \Delta \to [0,1]$ that can be *strongly* Bayesian implemented through budget-balanced transfers is the one associated with externality-ignoring utilitarianism: $k^*(\theta) = \arg\max_{k \in [0,1]} \theta_1 v(k) + \theta_2 v(1-k)$. Transfers are necessarily of *AGV-type*: If agents 1 and 2 claim to be of types $(\hat{\theta}_1, \hat{\delta}_1)$ and $(\hat{\theta}_2, \hat{\delta}_2)$, transfers are given by

$$\begin{aligned} t_1(\hat{\theta}, \hat{\delta}) &= \mathbb{E}_{\theta_2}\big[\theta_2 v(1 - k^*(\hat{\theta}_1, \theta_2))\big] - \mathbb{E}_{\theta_1}\big[\theta_1 v(k^*(\theta_1, \hat{\theta}_2))\big] + s(\hat{\theta}, \hat{\delta}), \\ t_2(\hat{\theta}, \hat{\delta}) &= \mathbb{E}_{\theta_1}\big[\theta_1 v(k^*(\theta_1, \hat{\theta}_2))\big] - \mathbb{E}_{\theta_2}\big[\theta_2 v(1 - k^*(\hat{\theta}_1, \theta_2))\big] - s(\hat{\theta}, \hat{\delta}), \end{aligned}$$

where s must be chosen such that $\mathbb{E}_{\theta_{-i}, \delta_{-i}}[s(\theta, \delta)]$ is constant on $\Theta_i \times \Delta_i$ for each i, such that externality assessments are strategically inoperative. That is, bargaining must focus on private payoffs, irrespective of externalities. When letting $s = 0$, as agents' externality assessments might not be common knowledge, agents make mutual concessions amounting to the expected externalities they impose on each other under the bargaining solution k^*.

[31] See Thomson (1994) for a survey.

[32] The results are also informative for "pure" bargaining (i.e., if utility is not transferable), since side-payments could be zero if the bargaining solution was incentive-compatible on its own.

The necessity of externality robustness seems particularly plausible in the range of conflict resolution. An arbitrator, seeking to resolve dispute between hostile parties, should rather claim "Let's focus on the issue!" than care about who dislikes whom how much (and is thus more or less spiteful).

The results of the preceding Sections preclude the most prominent solutions to axiomatic bargaining from being strongly Bayesian implemented without incentive costs. At best, they could be Bayesian implemented through budget-balanced transfers only for very specific type distributions.

Proposition 6 *The bargaining solutions of Nash (1950), Kalai (1977), and Kalai and Smorodinsky (1975), all of these either externality-sensitive or externality-ignoring, cannot be strongly Bayesian implemented through budget-balanced transfers.*[33]

Proof. The bargaining solutions under consideration each assume that agents would end up with an inferior allocation if they did not come to a mutual agreement upon pie-division. For the present purpose, it suffices to assume that this "threat point" yields both agents a zero-payoff: $\pi_i = 0$.

The externality-sensitive *Nash solution* is given by

$$(23) \qquad k^*(\theta, \delta) \;=\; \arg\max_{k \in [0,1]} \big[\theta_1 v(k) + \delta_1 \theta_2 v(1-k)\big] \cdot \big[\theta_2 v(1-k) + \delta_2 \theta_1 v(k)\big].$$

By Proposition 1(iv) and Theorem 1, the Nash solution is not strongly Bayesian implementable through budget-balanced transfers.

The externality-sensitive *Kalai solution* requires, in the manner of Rawls (1971), to maximize the minimum of agents' ex post utilities. This is equivalent to $k^* = k^*(\theta, \delta)$ equalizing utilities, such that[34]

$$(24) \qquad 0 \;=\; \theta_2(1-\delta_1)v(1-k^*) - \theta_1(1-\delta_2)v(k^*) \;=\; F(k^*, \theta, \delta).$$

[33]By Proposition 1(iv) and Theorem 3, the opposite implication would hold for the Nash solution if externality types were common knowledge. The same is true for the Kalai-Smorodinsky solution if $\Delta_i \subset (0, \frac{\theta_i^{\min}}{\theta_{-i}^{\max}})$ for all i; this is indicated by Theorem 3 and identity (63) in the Appendix. It is also shown in the Appendix that the Kalai solution satisfies $\partial k^*/\partial \theta_1 < 0$ and thereby violates the necessary condition for Bayesian implementability, $\frac{\partial}{\partial \theta_i} \mathbb{E}_{\theta_{-i}} \big[v_i(k^*(\theta, \delta))\big] \geq 0$; that is, the Kalai solution is not Bayesian implementable even if externality types are common knowledge.

[34]Due to the assumptions on v and $\Theta \times \Delta$, a solution to (24) always exists.

The externality-sensitive *Kalai-Smorodinsky solution* requires k^* to equalize the ratio of agents' ex post utilities and the ratio of agents' maximum potential gains: $\frac{u_1(k^*)}{u_2(k^*)} = \frac{u_1(1)}{u_2(0)}$, where $u_i(k) = \theta_i v_i(k) + \delta_i \theta_{-i} v_{-i}(k)$. This is equivalent to k^* satisfying[35]

$$(25) \qquad 0 \;=\; \theta_2(\theta_1 - \delta_1\theta_2)v(1-k^*) - \theta_1(\theta_2 - \delta_2\theta_1)v(k^*) \;=\; G(k^*, \theta, \delta).$$

Condition (5) of Lemma 1 implies in particular that a partially differentiable bargaining solution k^* that *does not* maximize aggregate private payoffs is strongly Bayesian implementable through budget-balanced transfers *only if* the following holds for all (θ, δ):[36]

$$(26) \qquad \operatorname{sgn}\left(\frac{\partial k^*}{\partial \theta_1}\frac{\partial k^*}{\partial \theta_2}\right) \;=\; -\operatorname{sgn}\left(\frac{\partial k^*}{\partial \delta_1}\frac{\partial k^*}{\partial \delta_2}\right).$$

It will be shown in the Appendix that the bargaining solutions (24) and (25) each satisfy $\operatorname{sgn}(\frac{\partial k^*}{\partial \theta_1}\frac{\partial k^*}{\partial \theta_2}) = -1 = \operatorname{sgn}(\frac{\partial k^*}{\partial \delta_1}\frac{\partial k^*}{\partial \delta_2})$ and thereby violate condition (26).

The respective externality-ignoring versions of the above bargaining solutions are obtained when letting $\delta_i = 0$ in (23) to (25). These solutions obviously violate condition (26), since then $\frac{\partial k^*}{\partial \delta_i} = 0$, whereas $\frac{\partial k^*}{\partial \theta_i} \neq 0$. ■

7 Conclusion

I have presented an incentive theory of normative principles. For this purpose, I have explored mechanism design for agents whose assessments of (in)tangible externalities and private payoffs are all subject to asymmetric information.

Under reasonable assumptions, Pareto-efficient allocations are Bayesian implementable through budget-balanced transfers *if and only if* the normative principle underlying the choice of allocations is that of *externality-ignoring utilitarianism*, which requires to maximize aggregate private payoffs exclusive of externalities.

Intangible externalities may be associated with altruism, spite, or the quest for status. In order to attain allocative efficiency, social planners must ignore such other-regarding

[35] Condition (25) is well-defined on $\Theta \times \Delta$ if and only if $\delta_i^{\max} \leq \frac{\theta_i^{\min}}{\theta_{-i}^{\max}}$ for all i: A solution k^* exists if and only if either $(\theta_1 - \delta_1\theta_2), (\theta_2 - \delta_2\theta_1) \geq 0$, or $(\theta_1 - \delta_1\theta_2), (\theta_2 - \delta_2\theta_1) < 0$; however, the latter condition would imply that $(1-\delta_1)\theta_2 + (1-\delta_2)\theta_1 < 0$, which contradicts the assumptions on $\Theta \times \Delta$.

[36] To be sure, multiplying condition (5) in the version of $i = 1$ with (5) in the version of $i = 2$ and applying the sign function to the resulting identity yields the condition $\operatorname{sgn}\left(\frac{\partial k^*}{\partial \theta_1}\frac{\partial k^*}{\partial \theta_2}\right) = \operatorname{sgn}\left(\frac{\partial v_1(k^*)}{\partial \delta_1}\frac{\partial v_2(k^*)}{\partial \delta_2}\right)$ which, in the present context, is equivalent to (26).

preferences. In the range of conflict resolution, this insight provides a rationale for the common-sense approach many people would adopt when arbitrating between conflicting parties: to *not* condition the arbitration process or final resolution on the extent to which the opponents despise each other but to rather "focus on the issue" and base arbitration solely on how it would affect the opponents' *material* well-being. One may think of how judges approach the resolution of divorce battles, how a mother tends to resolve animosity between her children, or how third-party diplomats try to conciliate rival tribes or nations.

When interpreting intangible externalities as people's distributive preferences, the result suggests that public economic policies dedicated to maximize a social-welfare measure inconsistent with externality-ignoring utilitarianism do either provide people with adverse incentives (e.g., to reduce their labor supply beyond the efficient level) or are not budget-balanced, leading either to a waste of money or an increase in public debt.

The Pareto-efficient, budget-balanced mechanism corresponding to externality-ignoring utilitarianism *necessarily* takes the form of the renowned AGV-mechanism. This mechanism is *externality-robust* in that it leaves agents' externality assessments strategically inoperative—and even allows to ignore them completely. Externality robustness thus turns out to be more than just a desirable property in order to avoid unrealistic common knowledge assumptions (about type distributions), as urged by Wilson (1987). Externality robustness is actually necessary from an incentive compatibility and allocative point of view.

Appendix

Proof of Proposition 1(iv) and (v)

Suppose in the following that $h_i = 0$ and $\Delta_i \subset (0,1)$ for all i. Obviously, the social-welfare measures (iv) and (v) satisfy payoff sensitivity and the Pareto property. In order to verify identities (1) and (2), ease notation by letting $\pi_i = \pi_i(k \,|\, \theta_i)$ and $v_i = v_i(k)$.

Proof of Proposition 1(iv)

Let $V(k) = (\pi_1 + \delta_1\pi_2)(\pi_2 + \delta_2\pi_1)$. By assumption, k^* satisfies the first-order condition

$$(27) \quad 0 = \frac{dV(k^*)}{dk} = \left(\frac{d\pi_1}{dk} + \delta_1\frac{d\pi_2}{dk}\right)(\pi_2 + \delta_2\pi_1) + \left(\frac{d\pi_2}{dk} + \delta_2\frac{d\pi_1}{dk}\right)(\pi_1 + \delta_1\pi_2)\,.$$

Define $x_1 = \pi_1 + \delta_1 \pi_2$ and $x_2 = \pi_2 + \delta_2 \pi_1$, and notice that $x_i > 0$. Then (27) can be rewritten so as to obtain

$$0 = (x_1 + \delta_1 x_2) \frac{d\pi_2}{dk} + (x_2 + \delta_2 x_1) \frac{d\pi_1}{dk}, \tag{28}$$

where $x_1 + \delta_1 x_2 > 0$ and $x_2 + \delta_2 x_1 > 0$. Implicit differentiation of (27) with respect to θ_1 yields $\partial k^*/\partial\theta_1 = -X_1/[d^2V(k^*)/dk^2]$, where

$$X_1 = x_2 \frac{dv_1}{dk} + \delta_2 v_1 \left(\frac{d\pi_1}{dk} + \delta_1 \frac{d\pi_2}{dk} \right) + \delta_2 x_1 \frac{dv_1}{dk} + v_1 \left(\frac{d\pi_2}{dk} + \delta_2 \frac{d\pi_1}{dk} \right).$$

Since $d^2V(k^*)/dk^2 < 0$ by the second-order condition, $\operatorname{sgn}(\partial k^*/\partial\theta_1) = \operatorname{sgn}(X_1)$. Since $h_i = 0$, one can make use of the identities $v_1 \frac{d\pi_1}{dk} = \pi_1 \frac{dv_1}{dk}$ and (28) to rewrite X_1 as

$$\begin{aligned} X_1 &= (x_2 + \delta_2 x_1) \frac{dv_1}{dk} + v_1 \left[2\delta_2 \frac{d\pi_1}{dk} + (1 + \delta_1\delta_2) \frac{d\pi_2}{dk} \right] \\ &= (x_2 + \delta_2 x_1) \frac{dv_1}{dk} + v_1 \frac{d\pi_1}{dk} \left[2\delta_2 - (1 + \delta_1\delta_2) \frac{(x_2 + \delta_2 x_1)}{(x_1 + \delta_1 x_2)} \right] \\ &= \frac{dv_1}{dk} \frac{Y_1}{(x_1 + \delta_1 x_2)}, \end{aligned}$$

where $Y_1 = [(x_1 + \delta_1 x_2)(x_2 + \delta_2 x_1) + \pi_1(1 - \delta_1\delta_2)(\delta_2 x_1 - x_2)]$. As $\delta_i \in (0,1)$ and $\pi_i, x_i > 0$, letting $\delta_i = 0$ yields the lower bound $Y_1 > x_1 x_2 + \pi_1(-x_2) = (x_1 - \pi_1)x_2 = \delta_1 \pi_2 x_2 > 0$. Hence, $\operatorname{sgn}(\partial k^*/\partial\theta_1) = \operatorname{sgn}(X_1) = \operatorname{sgn}(dv_1/dk)$, while, by assumption, $\operatorname{sgn}(\partial k^*/\partial\theta_i) \neq 0$ and $\operatorname{sgn}(dv_i/dk) \neq 0$. Hence, $1 = \operatorname{sgn}^2(\partial k^*/\partial\theta_1) = \operatorname{sgn}(dv_1/dk)\operatorname{sgn}(\partial k^*/\partial\theta_i) = \operatorname{sgn}(\partial v_1/\partial\theta_1)$. By symmetry, $1 = \operatorname{sgn}(\partial v_2/\partial\theta_2)$. Hence, $1 = \operatorname{sgn}(\partial v_1/\partial\theta_1 \cdot \partial v_2/\partial\theta_2) = \operatorname{sgn}(\partial v_1/\partial\theta_2 \cdot \partial v_2/\partial\theta_1) = \operatorname{sgn}(\partial v_1/\partial\theta_2)\operatorname{sgn}(\partial v_2/\partial\theta_1)$, as required.

On the other hand, implicit differentiation of (27) with respect to δ_1 yields $\partial k^*/\partial\delta_1 = -Z_1/[d^2V(k^*)/dk^2]$, where

$$Z_1 = x_2 \frac{d\pi_2}{dk} + \pi_2 \left(\frac{d\pi_2}{dk} + \delta_2 \frac{d\pi_1}{dk} \right).$$

Since $d^2V(k^*)/dk^2 < 0$ by the second-order condition, $\operatorname{sgn}(\partial k^*/\partial\delta_1) = \operatorname{sgn}(Z_1)$. By making use of (28), Z_1 can be written as

$$Z_1 = \frac{d\pi_2}{dk} \left[x_2 + \pi_2 - \delta_2 \pi_2 \frac{(x_1 + \delta_1 x_2)}{(x_2 + \delta_2 x_1)} \right] = x_2 \frac{d\pi_2}{dk} \left[1 + \pi_2 \frac{(1 - \delta_1\delta_2)}{(x_2 + \delta_2 x_1)} \right].$$

Hence, $\operatorname{sgn}(\partial k^*/\partial\delta_1) = \operatorname{sgn}(Z_1) = \operatorname{sgn}(\pi_2/dk)$, such that

$$\operatorname{sgn}\left(\frac{\partial\pi_2}{\partial\delta_1}\right) = \operatorname{sgn}\left(\frac{d\pi_2}{dk}\right)\operatorname{sgn}\left(\frac{dk^*}{\partial\delta_1}\right) = \operatorname{sgn}^2\left(\frac{d\pi_2}{dk}\right) \in \{0,1\}.$$

By symmetry, $\operatorname{sgn}(\partial\pi_1/\partial\delta_2) = \operatorname{sgn}^2(d\pi_1/dk) \in \{0,1\}$. As $x_1+\delta_1 x_2 > 0$ and $x_2+\delta_2 x_1 > 0$, condition (27) implies that $d\pi_1/dk = 0$ if and only if $d\pi_2/dk = 0$. Hence, as required, $\operatorname{sgn}(\partial\pi_1/\partial\delta_2) = \operatorname{sgn}(\partial\pi_2/\partial\delta_1)$. Altogether, W is sensitive.

Proof of Proposition 1(v)

Let $V(k) = [(\pi_1+\delta_1\pi_2)^{-\rho} + (\pi_2+\delta_2\pi_1)^{-\rho}]^{-\frac{1}{\rho}}$, with $\rho \in (-1,\infty)\setminus\{0\}$. By assumption, k^* satisfies the first-order condition

$$(29) \qquad 0 = \frac{dV(k^*)}{dk} = \left[V(k^*)\right]^{1+\rho}\left[(\pi_1+\delta_1\pi_2)^{-\rho-1}\left(\frac{d\pi_1}{dk}+\delta_1\frac{d\pi_2}{dk}\right) + (\pi_2+\delta_2\pi_1)^{-\rho-1}\left(\frac{d\pi_2}{dk}+\delta_2\frac{d\pi_1}{dk}\right)\right].$$

Define $x_1 = \pi_1+\delta_1\pi_2$ and $x_2 = \pi_2+\delta_2\pi_1$, and notice that $x_i > 0$. By (29),

$$(30) \qquad 0 = \left(x_1^{-\rho-1}+\delta_2 x_2^{-\rho-1}\right)\frac{d\pi_1}{dk} + \left(x_2^{-\rho-1}+\delta_1 x_1^{-\rho-1}\right)\frac{d\pi_2}{dk},$$

where $x_1^{-\rho-1}+\delta_2 x_2^{-\rho-1} > 0$ and $x_2^{-\rho-1}+\delta_1 x_1^{-\rho-1} > 0$. Implicit differentiation of (29) with respect to θ_1 yields $\partial k^*/\partial\theta_1 = -X_1[V(k^*)]^{1+\rho}/[d^2V(k^*)/dk^2]$, where

$$\begin{aligned} X_1 &= \left(x_1^{-\rho-1}+\delta_2 x_2^{-\rho-1}\right)\frac{dv_1}{dk} - (1+\rho)x_1^{-\rho-2}v_1\left(\frac{d\pi_1}{dk}+\delta_1\frac{d\pi_2}{dk}\right) \\ &\quad -(1+\rho)x_2^{-\rho-2}\delta_2 v_1\left(\frac{d\pi_2}{dk}+\delta_2\frac{d\pi_1}{dk}\right). \end{aligned}$$

Since $d^2V(k^*)/dk^2 < 0$ by the second-order condition, $\operatorname{sgn}(\partial k^*/\partial\theta_1) = \operatorname{sgn}(X_1)$. Since $h_i = 0$, one can make use of the identities $v_1\frac{d\pi_1}{dk} = \pi_1\frac{dv_1}{dk}$ and (30) to rewrite X_1 as

$$\begin{aligned} X_1 &= \frac{dv_1}{dk}\frac{Y_1}{x_2^{-\rho-1}+\delta_1 x_1^{-\rho-1}}, \quad \text{where} \\ Y_1 &= \left(x_1^{-\rho-1}+\delta_2 x_2^{-\rho-1}\right)\left(x_2^{-\rho-1}+\delta_1 x_1^{-\rho-1}\right) \\ &\quad + (1+\rho)(1-\delta_1\delta_2)x_1^{-\rho-2}x_2^{-\rho-2}(\delta_2 x_1 - x_2)\pi_1. \end{aligned}$$

Hence, $\operatorname{sgn}(\partial k^*/\partial\theta_1) = \operatorname{sgn}(dv_1/dk)\operatorname{sgn}(Y_1)$. By exchanging the roles of 1 and 2, one obtains $\operatorname{sgn}(\partial k^*/\partial\theta_2) = \operatorname{sgn}(dv_2/dk)\operatorname{sgn}(Y_2)$, where Y_2 is defined as

$$\begin{aligned} Y_2 &= \left(x_1^{-\rho-1} + \delta_2 x_2^{-\rho-1}\right)\left(x_2^{-\rho-1} + \delta_1 x_1^{-\rho-1}\right) \\ &\qquad + (1+\rho)(1-\delta_1\delta_2)x_1^{-\rho-2}x_2^{-\rho-2}(\delta_1 x_2 - x_1)\pi_2. \end{aligned}$$

Since $(\delta_2 x_1 - x_2)\pi_1 = -(1-\delta_1\delta_2)\pi_1\pi_2 = (\delta_1 x_2 - x_1)\pi_2$, one observes that $Y_1 = Y_2$. Hence, as required,

$$\begin{aligned} 1 &= \operatorname{sgn}^2\left(\frac{\partial k^*}{\partial\theta_1}\right)\operatorname{sgn}^2\left(\frac{\partial k^*}{\partial\theta_2}\right) \\ &= \operatorname{sgn}\left(\frac{\partial k^*}{\partial\theta_1}\right)\operatorname{sgn}\left(\frac{dv_1}{dk}\right)\operatorname{sgn}(Y_1)\operatorname{sgn}\left(\frac{\partial k^*}{\partial\theta_2}\right)\operatorname{sgn}\left(\frac{dv_2}{dk}\right)\operatorname{sgn}(Y_2) \\ &= \operatorname{sgn}\left(\frac{\partial v_1}{\partial\theta_2}\right)\operatorname{sgn}\left(\frac{\partial v_2}{\partial\theta_1}\right), \end{aligned} \tag{31}$$

where the first equality in (31) holds due to the assumption that $\partial k^*/\partial\theta i \neq 0$ for all i.

On the other hand, implicit differentiation of (29) with respect to δ_1 yields $\partial k^*/\partial\delta_1 = -Z_1[V(k^*)]^{1+\rho}/[d^2V(k^*)/dk^2]$, where

$$Z_1 = x_1^{-\rho-2}\left[x_1\frac{d\pi_2}{dk} - \pi_2(1+\rho)\left(\frac{d\pi_1}{dk} + \delta_1\frac{d\pi_2}{dk}\right)\right].$$

Since $d^2V(k^*)/dk^2 < 0$ by the second-order condition, $\operatorname{sgn}(\partial k^*/\partial\delta_1) = \operatorname{sgn}(Z_1)$. By making use of (30), Z_1 can be written as

$$\begin{aligned} Z_1 &= x_1^{-\rho-2}\frac{d\pi_2}{dk}\left[x_1 - \pi_2(1+\rho)\left(\delta_1 - \frac{x_2^{-\rho-1} + \delta_1 x_1^{-\rho-1}}{x_1^{-\rho-1} + \delta_2 x_2^{-\rho-1}}\right)\right] \\ &= x_1^{-\rho-2}\frac{d\pi_2}{dk}\left[x_1 + \pi_2(1+\rho)(1-\delta_1\delta_2)\frac{x_2^{-\rho-1}}{x_1^{-\rho-1} + \delta_2 x_2^{-\rho-1}}\right]. \end{aligned}$$

Hence, $\operatorname{sgn}(\partial k^*/\partial\delta_1) = \operatorname{sgn}(Z_1) = \operatorname{sgn}(\pi_2/dk)$, such that

$$\operatorname{sgn}\left(\frac{\partial\pi_2}{\partial\delta_1}\right) = \operatorname{sgn}\left(\frac{d\pi_2}{dk}\right)\operatorname{sgn}\left(\frac{dk^*}{\partial\delta_1}\right) = \operatorname{sgn}^2\left(\frac{d\pi_2}{dk}\right) \in \{0,1\}.$$

By symmetry, $\operatorname{sgn}(\partial\pi_1/\partial\delta_2) = \operatorname{sgn}^2(d\pi_1/dk) \in \{0,1\}$. Since $x_1^{-\rho-1} + \delta_2 x_2^{-\rho-1} > 0$ and $x_2^{-\rho-1} + \delta_1 x_1^{-\rho-1} > 0$, identity (30) implies that $d\pi_1/dk = 0$ if and only if $d\pi_2/dk = 0$. Hence, as required, $\operatorname{sgn}(\partial\pi_1/\partial\delta_2) = \operatorname{sgn}(\partial\pi_2/\partial\delta_1)$. Altogether, W is sensitive.

Proof of Lemma 1

Suppose the partially continuously differentiable allocation rule $k^* : \Theta \times \Delta \to \mathbb{R}$ can be strongly Bayesian implemented with the ex post budget-balanced transfer scheme $T = (t_1, t_2) : \Theta \times \Delta \to \mathbb{R}^2$. Define

$$\begin{aligned}
\bar{v}_i(\hat{\theta}_i, \hat{\delta}_i) &= \mathbb{E}_{\theta_{-i},\delta_{-i}}\big[v_i(k^*(\hat{\theta}_i, \hat{\delta}_i, \theta_{-i}, \delta_{-i}))\big], \\
\bar{h}_i(\hat{\theta}_i, \hat{\delta}_i) &= \mathbb{E}_{\theta_{-i},\delta_{-i}}\big[h_i(k^*(\hat{\theta}_i, \hat{\delta}_i, \theta_{-i}, \delta_{-i}))\big], \\
\bar{\pi}_i(\hat{\theta}_i, \hat{\delta}_i) &= \mathbb{E}_{\theta_{-i},\delta_{-i}}\big[\pi_i(k^*(\hat{\theta}_i, \hat{\delta}_i, \theta_{-i}, \delta_{-i}) \,|\, \hat{\theta}_i)\big], \\
\bar{\pi}_{-i}(\hat{\theta}_i, \hat{\delta}_i) &= \mathbb{E}_{\theta_{-i},\delta_{-i}}\big[\pi_{-i}(k^*(\hat{\theta}_i, \hat{\delta}_i, \theta_{-i}, \delta_{-i}) \,|\, \theta_{-i})\big], \\
\bar{t}_i(\hat{\theta}_i, \hat{\delta}_i) &= \mathbb{E}_{\theta_{-i},\delta_{-i}}\big[t_i(\hat{\theta}_i, \hat{\delta}_i, \theta_{-i}, \delta_{-i})\big], \\
\bar{t}_{-i}(\hat{\theta}_i, \hat{\delta}_i) &= \mathbb{E}_{\theta_{-i},\delta_{-i}}\big[t_{-i}(\hat{\theta}_i, \hat{\delta}_i, \theta_{-i}, \delta_{-i})\big].
\end{aligned}$$

Denote by $U_i(\hat{\theta}_i, \hat{\delta}_i \,|\, \theta_i, \delta_i)$ agent i's interim expected utility from reporting $(\hat{\theta}_i, \hat{\delta}_i)$ if her true type is (θ_i, δ_i) and if agent $-i$ reports her type truthfully:

$$U_i(\hat{\theta}_i, \hat{\delta}_i \,|\, \theta_i, \delta_i) = \theta_i \bar{v}_i(\hat{\theta}_i, \hat{\delta}_i) + \bar{h}_i(\hat{\theta}_i, \hat{\delta}_i) + \bar{t}_i(\hat{\theta}_i, \hat{\delta}_i) + \delta_i \bar{\pi}_{-i}(\hat{\theta}_i, \hat{\delta}_i) + \delta_i \bar{t}_{-i}(\hat{\theta}_i, \hat{\delta}_i).$$

Ease notation by also defining $U_i(\theta_i, \delta_i) = U_i(\theta_i, \delta_i \,|\, \theta_i, \delta_i)$. Then the following must hold for all $\theta_i, \hat{\theta}_i \in \Theta_i$ and all $\delta_i, \hat{\delta}_i \in \Delta_i$:

$$\begin{aligned}
U_i(\theta_i, \delta_i) &\geq U_i(\hat{\theta}_i, \delta_i \,|\, \theta_i, \delta_i) \\
&= U_i(\hat{\theta}_i, \delta_i) + (\theta_i - \hat{\theta}_i)\bar{v}_i(\hat{\theta}_i, \delta_i), && (32)\\
U_i(\hat{\theta}_i, \delta_i) &\geq U_i(\theta_i, \delta_i \,|\, \hat{\theta}_i, \delta_i) \\
&= U_i(\theta_i, \delta_i) + (\hat{\theta}_i - \theta_i)\bar{v}_i(\theta_i, \delta_i), && (33)\\
U_i(\theta_i, \delta_i) &\geq U_i(\theta_i, \hat{\delta}_i \,|\, \theta_i, \delta_i) \\
&= U_i(\theta_i, \hat{\delta}_i) + (\delta_i - \hat{\delta}_i)\Big[\bar{\pi}_{-i}(\theta_i, \hat{\delta}_i) + \bar{t}_{-i}(\theta_i, \hat{\delta}_i)\Big], && (34)\\
U_i(\theta_i, \hat{\delta}_i) &\geq U_i(\theta_i, \delta_i \,|\, \theta_i, \hat{\delta}_i) \\
&= U_i(\theta_i, \delta_i) + (\hat{\delta}_i - \delta_i)\Big[\bar{\pi}_{-i}(\theta_i, \delta_i) + \bar{t}_{-i}(\theta_i, \delta_i)\Big]. && (35)
\end{aligned}$$

Suppose $\hat{\theta}_i > \theta_i$. Then (32) and (33) imply that

$$\bar{v}_i(\hat{\theta}_i, \delta_i) \;\geq\; \frac{U_i(\hat{\theta}_i, \delta_i) - U_i(\theta_i, \delta_i)}{\hat{\theta}_i - \theta_i} \;\geq\; \bar{v}_i(\theta_i, \delta_i). \tag{36}$$

As $\bar{v}_i$ is continuous on Θ_i, letting $\hat{\theta}_i$ approach θ_i implies that $\partial U_i(\theta_i, \delta_i)/\partial\theta_i = \bar{v}_i(\theta_i, \delta_i)$. Integrating the latter with respect to θ_i yields the identity

$$U_i(\theta_i, \delta_i) \;=\; p_i(\delta_i) + \int_{\theta_i^{\min}}^{\theta_i} \bar{v}_i(s, \delta_i)\, ds, \tag{37}$$

with some function $p_i : \Delta_i \to \mathbb{R}$. Similarly, suppose $\hat{\delta}_i > \delta_i$. Then (34) and (35) imply that

$$\bar{\pi}_{-i}(\theta_i, \hat{\delta}_i) + \bar{t}_{-i}(\theta_i, \hat{\delta}_i) \;\geq\; \frac{U_i(\theta_i, \hat{\delta}_i) - U_i(\theta_i, \delta_i)}{\hat{\delta}_i - \delta_i} \;\geq\; \bar{\pi}_{-i}(\theta_i, \delta_i) + \bar{t}_{-i}(\theta_i, \delta_i).$$

As $\bar{\pi}_{-i}$ and $\bar{t}_{-i}$ are continuous on Δ_i by assumption, letting $\hat{\delta}_i$ approach δ_i implies that

$$\frac{\partial U_i(\theta_i, \delta_i)}{\partial \delta_i} \;=\; \bar{\pi}_{-i}(\theta_i, \delta_i) + \bar{t}_{-i}(\theta_i, \delta_i). \tag{38}$$

Integrating (38) with respect to δ_i yields the identity

$$U_i(\theta_i, \delta_i) \;=\; q_i(\theta_i) + \int_{\delta_i^{\min}}^{\delta_i} \bar{\pi}_{-i}(\theta_i, r)\, dr + \int_{\delta_i^{\min}}^{\delta_i} \bar{t}_{-i}(\theta_i, r)\, dr, \tag{39}$$

with some function $q_i : \Theta_i \to \mathbb{R}$. As $\bar{\pi}_{-i}$ and $\bar{t}_{-i}$ are continuous on Δ_i, identity (39) implies that $U_i(\theta_i, \delta_i)$ is differentiable in δ_i. As $\bar{v}_i$ is differentiable in δ_i, identity (37) implies that also p_i is differentiable in δ_i. Jointly, (37) and (39) yield

$$\int_{\delta_i^{\min}}^{\delta_i} \bar{t}_{-i}(\theta_i, r)\, dr \;=\; p_i(\delta_i) - q_i(\theta_i) + \int_{\theta_i^{\min}}^{\theta_i} \bar{v}_i(s, \delta_i)\, ds - \int_{\delta_i^{\min}}^{\delta_i} \bar{\pi}_{-i}(\theta_i, r)\, dr. \tag{40}$$

Differentiating (40) with respect to δ_i yields

$$\bar{t}_{-i}(\theta_i, \delta_i) \;=\; \frac{dp_i(\delta_i)}{d\delta_i} - \bar{\pi}_{-i}(\theta_i, \delta_i) + \frac{\partial}{\partial \delta_i} \int_{\theta_i^{\min}}^{\theta_i} \bar{v}_i(s, \delta_i)\, ds. \tag{41}$$

Ex post budget balance requires in particular that $\bar{t}_i(\theta_i, \delta_i) = -\bar{t}_{-i}(\theta_i, \delta_i)$ on $\Theta_i \times \Delta_i$, such that truthful revelation of (θ_i, δ_i) is Bayesian incentive-compatible for agent i only if θ_i satisfies the first-order condition

(42)

$$\begin{aligned}
0 &= \frac{\partial}{\partial \hat{\theta}_i}\Big[\theta_i \bar{v}_i(\hat{\theta}_i, \delta_i) + \bar{h}_i(\hat{\theta}_i, \delta_i) + \delta_i \bar{\pi}_{-i}(\hat{\theta}_i, \delta_i) - (1-\delta_i)\bar{t}_{-i}(\hat{\theta}_i, \delta_i)\Big]\Bigg|_{\hat{\theta}_i=\theta_i} \\
&= \theta_i \frac{\bar{v}_i(\theta_i, \delta_i)}{\partial \theta_i} + \frac{\bar{h}_i(\theta_i, \delta_i)}{\partial \theta_i} + \delta_i \frac{\bar{\pi}_{-i}(\theta_i, \delta_i)}{\partial \theta_i} - (1-\delta_i)\left[\frac{\bar{v}_i(\theta_i, \delta_i)}{\partial \delta_i} - \frac{\bar{\pi}_{-i}(\theta_i, \delta_i)}{\partial \theta_i}\right] \\
&= \mathbb{E}_{\theta_{-i}, \delta_{-i}}\left[\frac{d\pi_i(k^*(\theta, \delta) \,|\, \theta_i)}{dk}\frac{\partial k^*}{\partial \theta_i} + \frac{d\pi_{-i}(k^*(\theta, \delta) \,|\, \theta_{-i}))}{dk}\frac{\partial k^*}{\partial \theta_i} - (1-\delta_i)\frac{v_i(k^*(\theta, \delta))}{\partial \delta_i}\right],
\end{aligned}$$

where the second equality is implied by identity (41). In order to be Bayesian implementable through budget-balanced transfers, k^* must satisfy identity (42) irrespective of the specific form that the transfer scheme might take. As k^* is also assumed to be *strongly* Bayesian implementable, identity (42) holds for arbitrary (non-degenerate) type distributions H_{-i}. However, due to the assumptions on the economic environment, the argument of $\mathbb{E}_{\theta_{-i}, \delta_{-i}}[\,\cdot\,]$ in (42) is continuous in $(\theta_{-i}, \delta_{-i})$. Hence, k^* must satisfy

$$0 = \frac{d\pi_i(k^*(\theta, \delta) \,|\, \theta_i)}{dk}\frac{\partial k^*}{\partial \theta_i} + \frac{d\pi_{-i}(k^*(\theta, \delta) \,|\, \theta_{-i}))}{dk}\frac{\partial k^*}{\partial \theta_i} - (1-\delta_i)\frac{v_i(k^*(\theta, \delta))}{\partial \delta_i}$$

for all $(\theta, \delta) \in \Theta \times \Delta$, which proves the first part of the Lemma.

For the second part, reconsider identities (37) and (41). Under truthful revelation, they jointly imply that

$$\begin{aligned}
(43) \qquad p_i(\delta_i) + \int_{\theta_i^{\min}}^{\theta_i} \bar{v}_i(s, \delta_i)\, ds &= U_i(\theta_i, \delta_i) \\
&= \theta_i \bar{v}_i(\theta_i, \delta_i) + \bar{h}_i(\theta_i, \delta_i) + \bar{t}_i(\theta_i, \delta_i) \\
&\qquad + \delta_i \frac{dp_i(\delta_i)}{d\delta_i} + \delta_i \frac{\partial}{\partial \delta_i}\int_{\theta_i^{\min}}^{\theta_i} \bar{v}_i(s, \delta_i)\, ds.
\end{aligned}$$

Now suppose k^* is independent from externality types: $\partial k^*/\partial \delta_i = 0$ for all i. According to identities (43) and (41), respectively, $\bar{t}_i(\theta_i, \delta_i)$ and $\bar{t}_{-i}(\theta_i, \delta_i)$ then satisfy

$$\begin{aligned}
(44) \qquad \bar{t}_i(\theta_i, \delta_i) &= p_i(\delta_i) - \delta_i \frac{dp_i(\delta_i)}{d\delta_i} - \theta_i \bar{v}_i(\theta_i, \delta_i) - \bar{h}_i(\theta_i, \delta_i) + \int_{\theta_i^{\min}}^{\theta_i} \bar{v}_i(s, \delta_i)\, ds, \\
(45) \qquad \bar{t}_{-i}(\theta_i, \delta_i) &= \frac{dp_i(\delta_i)}{d\delta_i} - \bar{\pi}_{-i}(\theta_i, \delta_i),
\end{aligned}$$

where, now, only the terms containing p_i effectively depend on δ_i. Due to budget balance, identities (44) and (45) imply that p_i solves the differential equation

$$a_i = p_i(\delta_i) + (1-\delta_i)\frac{dp_i(\delta_i)}{d\delta_i}, \tag{46}$$

where a_i is some constant. Differentiating (46) with respect to δ_i yields $\partial^2 p_i(\delta_i)/\partial\delta_i^2 = 0$, such that $dp_i(\delta_i)/d\delta_i = -\alpha_i$ for some constant α_i. Hence, identity (45) reads $\bar{t}_{-i}(\theta_i,\delta_i) = -\alpha_i - \bar{\pi}_{-i}(\theta_i,\delta_i)$, implying that $\bar{t}_i(\theta_i,\delta_i) = \alpha_i + \bar{\pi}_{-i}(\theta_i,\delta_i) = \alpha_i + \mathbb{E}_{\theta_{-i}}\big[\pi_{-i}(k^*(\hat{\theta})\,|\,\theta_{-i})\big]$, due to budget balance and $\partial k^*/\partial\delta_i = 0$.

Proof of Theorem 2 Extended

With notation adopted from the proof of Lemma 1, T^* satisfies

$$\begin{aligned}
\bar{t}_i(\hat{\theta}_i,\hat{\delta}_i) &= a_i + p_i(\hat{\delta}_i) - \hat{\delta}_i\frac{\partial p_i(\hat{\delta}_i)}{\partial\hat{\delta}_i} - \bar{\pi}_i(\hat{\theta}_i,\hat{\delta}_i) \\
&\quad + \int_{\theta_i^{\min}}^{\hat{\theta}_i} \bar{v}_i(s,\hat{\delta}_i)\,ds - \hat{\delta}_i\frac{\partial}{\partial\hat{\delta}_i}\int_{\theta_i^{\min}}^{\hat{\theta}_i} \bar{v}_i(s,\hat{\delta}_i)\,ds, \\
\bar{t}_{-i}(\hat{\theta}_i,\hat{\delta}_i) &= b_i + \frac{\partial p_i(\hat{\delta}_i)}{\partial\hat{\delta}_i} - \bar{\pi}_{-i}(\hat{\theta}_i,\hat{\delta}_i) + \frac{\partial}{\partial\hat{\delta}_i}\int_{\theta_i^{\min}}^{\hat{\theta}_i} \bar{v}_i(s,\hat{\delta}_i)\,ds,
\end{aligned}$$

with appropriate constants $a_i, b_i \in \mathbb{R}$. Suppose agent $-i$ reports her type truthfully. From reporting some type $(\hat{\theta}_i,\hat{\delta}_i)$, agent i of true type (θ_i,δ_i) gains interim expected utility

$$\begin{aligned}
U_i(\hat{\theta}_i,\hat{\delta}_i\,|\,\theta_i,\delta_i) &= \theta_i\bar{v}_i(\hat{\theta}_i,\hat{\delta}_i) + \bar{h}_i(\hat{\theta}_i,\hat{\delta}_i) + \bar{t}_i(\hat{\theta}_i,\hat{\delta}_i) \\
&\quad + \delta_i\bar{\pi}_{-i}(\hat{\theta}_i,\hat{\delta}_i) + \delta_i\bar{t}_{-i}(\hat{\theta}_i,\hat{\delta}_i) \\
&= \theta_i\bar{v}_i(\hat{\theta}_i,\hat{\delta}_i) + a_i + p_i(\hat{\delta}_i) - \hat{\delta}_i\frac{\partial p_i(\hat{\delta}_i)}{\partial\hat{\delta}_i} - \hat{\theta}_i\bar{v}_i(\hat{\theta}_i,\hat{\delta}_i) \\
&\quad + \int_{\theta_i^{\min}}^{\hat{\theta}_i} \bar{v}_i(s,\hat{\delta}_i)\,ds - \hat{\delta}_i\frac{\partial}{\partial\hat{\delta}_i}\int_{\theta_i^{\min}}^{\hat{\theta}_i} \bar{v}_i(s,\hat{\delta}_i)\,ds \\
&\quad + \delta_i b_i + \delta_i\frac{\partial p_i(\hat{\delta}_i)}{\partial\hat{\delta}_i} + \delta_i\frac{\partial}{\partial\hat{\delta}_i}\int_{\theta_i^{\min}}^{\hat{\theta}_i} \bar{v}_i(s,\hat{\delta}_i)\,ds.
\end{aligned}$$

Partial derivatives thus satisfy

$$\frac{\partial}{\partial\hat{\theta}_i}U_i(\hat{\theta}_i,\hat{\delta}_i\,|\,\theta_i,\delta_i) \;=\; (\theta_i-\hat{\theta}_i)\frac{\partial}{\partial\hat{\theta}_i}\bar{v}_i(\hat{\theta}_i,\hat{\delta}_i)+(\delta_i-\hat{\delta}_i)\frac{\partial}{\partial\hat{\delta}_i}\bar{v}_i(\hat{\theta}_i,\hat{\delta}_i), \tag{47}$$

$$\begin{aligned}\frac{\partial}{\partial\hat{\delta}_i}U_i(\hat{\theta}_i,\hat{\delta}_i\,|\,\theta_i,\delta_i) \;&=\; (\theta_i-\hat{\theta}_i)\frac{\partial}{\partial\hat{\delta}_i}\bar{v}_i(\hat{\theta}_i,\hat{\delta}_i)\\ &\quad+(\delta_i-\hat{\delta}_i)\frac{\partial^2}{\partial\hat{\delta}_i^2}\left[p_i(\hat{\delta}_i)+\int_{\theta_i^{\min}}^{\hat{\theta}_i}\bar{v}_i(s,\hat{\delta}_i)\,ds\right].\end{aligned} \tag{48}$$

Ease notation by defining $A_i=\frac{\partial}{\partial\hat{\delta}_i}\bar{v}_i(\hat{\theta}_i,\hat{\delta}_i)$, $B_i=\frac{\partial}{\partial\hat{\theta}_i}\bar{v}_i(\hat{\theta}_i,\hat{\delta}_i)$, and

$$C_i \;=\; \frac{\partial^2}{\partial\hat{\delta}_i^2}\left[p_i(\hat{\delta}_i)+\int_{\theta_i^{\min}}^{\hat{\theta}_i}\bar{v}_i(s,\hat{\delta}_i)\,ds\right].$$

Then (47) and (48) read

$$\frac{\partial}{\partial\hat{\theta}_i}U_i(\hat{\theta}_i,\hat{\delta}_i\,|\,\theta_i,\delta_i) \;=\; (\theta_i-\hat{\theta}_i)B_i+(\delta_i-\hat{\delta}_i)A_i, \tag{49}$$

$$\frac{\partial}{\partial\hat{\delta}_i}U_i(\hat{\theta}_i,\hat{\delta}_i\,|\,\theta_i,\delta_i) \;=\; (\theta_i-\hat{\theta}_i)A_i+(\delta_i-\hat{\delta}_i)C_i. \tag{50}$$

By assumption, $B_i>0$. Choose $p_i(\delta_i)=\frac{1}{2}c_i\delta_i^2$, with c_i as defined by condition (19). Then $C_i>0$, and condition (18) is satisfied:

$$A_i^2 \;<\; B_iC_i.$$

Notice first that $(\hat{\theta}_i,\hat{\delta}_i)=(\theta_i,\delta_i)$ is the unique stationary point of $U_i(\hat{\theta}_i,\hat{\delta}_i\,|\,\theta_i,\delta_i)$, since $\frac{\partial}{\partial\hat{\theta}_i}U_i(\hat{\theta}_i,\hat{\delta}_i\,|\,\theta_i,\delta_i)=0=\frac{\partial}{\partial\hat{\delta}_i}U_i(\hat{\theta}_i,\hat{\delta}_i\,|\,\theta_i,\delta_i)$ implies that $(\theta_i-\hat{\theta}_i)=-(\delta_i-\hat{\delta}_i)\frac{A_i}{B_i}$ and, thus, $0=(\delta_i-\hat{\delta}_i)\frac{1}{B_i}(B_iC_i-A_i^2)$, where $B_i>0$ and $B_iC_i-A_i^2>0$. Evaluating the Hessian H_i of $U_i(\hat{\theta}_i,\hat{\delta}_i\,|\,\theta_i,\delta_i)$ at $(\hat{\theta}_i,\hat{\delta}_i)=(\theta_i,\delta_i)$ yields

$$H_i \;=\; \begin{pmatrix}-B_i & -A_i\\ -A_i & -C_i\end{pmatrix}. \tag{51}$$

The principal minors of (51), namely $-B_i<0$ and $\det(H_i)=B_iC_i-A_i^2>0$, are alternating in sign, with the first-order principal minor being negative. Hence, (θ_i,δ_i) is a local maximizer of $U_i(\hat{\theta}_i,\hat{\delta}_i\,|\,\theta_i,\delta_i)$. It remains to show that truth-telling is indeed the (unique) *global* expected utility maximizer for agent i. Given the above, it suffices to show that the extension of $U_i(\hat{\theta}_i, \hat{\delta}_i \,|\, \theta_i, \delta_i)$ to the closure of $\Theta_i \times \Delta_i$ has no local maximizer on the boundary of $\Theta_i \times \Delta_i$.

Suppose a local maximizer is located on $(\theta_i^{\min}, \theta_i^{\max}) \times \{\delta_i^{\min}\}$ or $(\theta_i^{\min}, \theta_i^{\max}) \times \{\delta_i^{\max}\}$. As $U_i(\hat{\theta}_i, \hat{\delta}_i \,|\, \theta_i, \delta_i)$ is twice partially continuously differentiable, this maximizer, $(\hat{\theta}_i, \hat{\delta}_i)$, must satisfy $0 = \frac{\partial}{\partial \hat{\theta}_i} U_i(\hat{\theta}_i, \hat{\delta}_i \,|\, \theta_i, \delta_i)$ and, thus, $(\theta_i - \hat{\theta}_i) = -(\delta_i - \hat{\delta}_i)\frac{A_i}{B_i}$. Substituting the latter into (50) yields $\frac{\partial}{\partial \hat{\delta}_i} U_i(\hat{\theta}_i, \hat{\delta}_i \,|\, \theta_i, \delta_i) = (\delta_i - \hat{\delta}_i)\frac{1}{B_i}(B_i C_i - A_i^2)$. As $\frac{1}{B_i}(B_i C_i - A_i^2) > 0$, the reporting of $\hat{\delta}_i \in \{\delta_i^{\min}, \delta_i^{\max}\}$ is not optimal, which contradicts the assumption. By a similar argument one can show that no local maximizer is located on $\{\theta_i^{\min}\} \times (\delta_i^{\min}, \delta_i^{\max})$ or $\{\theta_i^{\max}\} \times (\delta_i^{\min}, \delta_i^{\max})$. Hence, only the corners of the closure of $\Theta_i \times \Delta_i$ qualify as further local maximizers.

Suppose $(\theta_i^{\max}, \delta_i^{\max})$ is a local maximizer. Then $0 \leq \frac{\partial}{\partial \hat{\theta}_i} U_i(\theta_i^{\max}, \delta_i^{\max} \,|\, \theta_i, \delta_i)$ and $0 \leq \frac{\partial}{\partial \hat{\delta}_i} U_i(\theta_i^{\max}, \delta_i^{\max} \,|\, \theta_i, \delta_i)$ must hold. As $(\theta_i - \theta_i^{\max}), (\delta_i - \delta_i^{\max}) < 0$, while $B_i, C_i > 0$, this implies that $A_i < 0$. However, by (49) and (50), $(\delta_i - \delta_i^{\max}) \geq -(\theta_i - \theta_i^{\max})\frac{A_i}{C_i}$ and, thus,

$$0 \;\leq\; (\theta_i - \theta_i^{\max})B_i + (\delta_i - \delta_i^{\max})A_i \;\leq\; (\theta_i - \theta_i^{\max})\frac{1}{C_i}(B_i C_i - A_i^2) \;<\; 0.$$

Suppose $(\theta_i^{\max}, \delta_i^{\min})$ is a local maximizer. Then $0 \leq \frac{\partial}{\partial \hat{\theta}_i} U_i(\theta_i^{\max}, \delta_i^{\min} \,|\, \theta_i, \delta_i)$ and $0 \geq \frac{\partial}{\partial \hat{\delta}_i} U_i(\theta_i^{\max}, \delta_i^{\min} \,|\, \theta_i, \delta_i)$ must hold. As $(\theta_i - \theta_i^{\max}) < 0$, while $(\delta_i - \delta_i^{\min}), B_i, C_i > 0$, this implies that $A_i > 0$. However, by (49) and (50), $(\theta_i - \theta_i^{\max}) \geq -(\delta_i - \delta_i^{\min})\frac{A_i}{B_i}$ and, thus,

$$0 \;\geq\; (\theta_i - \theta_i^{\max})A_i + (\delta_i - \delta_i^{\min})C_i \;\geq\; (\delta_i - \delta_i^{\min})\frac{1}{B_i}(B_i C_i - A_i^2) \;>\; 0.$$

Suppose $(\theta_i^{\min}, \delta_i^{\min})$ is a local maximizer. Then $0 \geq \frac{\partial}{\partial \hat{\theta}_i} U_i(\theta_i^{\min}, \delta_i^{\min} \,|\, \theta_i, \delta_i)$ and $0 \geq \frac{\partial}{\partial \hat{\delta}_i} U_i(\theta_i^{\min}, \delta_i^{\min} \,|\, \theta_i, \delta_i)$ must hold. As $(\theta_i - \theta^{\min}), (\delta_i - \delta^{\min}), B_i, C_i > 0$, this implies that $A_i < 0$. However, by (49) and (50), $(\delta_i - \delta^{\min}) \leq -(\theta_i - \theta^{\min})\frac{A_i}{C_i}$ and, thus,

$$0 \;\geq\; (\theta_i - \theta_i^{\min})B_i + (\delta_i - \delta_i^{\min})A_i \;\geq\; (\theta_i - \theta_i^{\min})\frac{1}{C_i}(B_i C_i - A_i^2) \;>\; 0.$$

Finally, suppose $(\theta_i^{\min}, \delta_i^{\max})$ is a local maximizer. Then $0 \geq \frac{\partial}{\partial \hat{\theta}_i} U_i(\theta_i^{\min}, \delta_i^{\max} \,|\, \theta_i, \delta_i)$ and $0 \leq \frac{\partial}{\partial \hat{\delta}_i} U_i(\theta_i^{\min}, \delta_i^{\max} \,|\, \theta_i, \delta_i)$ must hold. As $(\delta_i - \delta_i^{\max}) < 0$ and $(\theta_i - \theta_i^{\min}), B_i, C_i > 0$, this implies that $A_i > 0$. However, by (49) and (50), $(\theta_i - \theta_i^{\min}) \leq -(\delta_i - \delta_i^{\max})\frac{A_i}{B_i}$ and, thus,

$$0 \;\leq\; (\theta_i - \theta_i^{\min})A_i + (\delta_i - \delta_i^{\max})C_i \;\leq\; (\delta_i - \delta_i^{\max})\frac{1}{B_i}(B_iC_i - A_i^2) \;<\; 0.$$

Altogether, (θ_i, δ_i) is the unique global maximizer of $U_i(\hat{\theta}_i, \hat{\delta}_i \,|\, \theta_i, \delta_i)$. As the above arguments hold for any set of type distributions, T^* *strongly* Bayesian implements k^*.

Derivation of the transfer scheme T^* in the proof of Theorem 2

Suppose the transfer scheme $T = (t_1, t_2)$ strongly Bayesian implements the partially continuously differentiable allocation rule k^*. With notation adopted from the proof of Lemma 1, condition (41) states that T must satisfy

$$\text{(52)} \qquad \bar{t}_{-i}(\theta_i, \delta_i) \;=\; \frac{dp_i(\delta_i)}{d\delta_i} - \bar{\pi}_{-i}(\theta_i, \delta_i) + \frac{\partial}{\partial \delta_i}\int_{\theta_i^{\min}}^{\theta_i} \bar{v}_i(s, \delta_i)\, ds,$$

where $p_i : \Delta_i \to \mathbb{R}$ is some differentiable function. By conditions (37) and (52),

$$\begin{aligned}
p_i(\delta_i) + \int_{\theta_i^{\min}}^{\theta_i} \bar{v}_i(s, \delta_i)\, ds \;&=\; U_i(\theta_i, \delta_i) \\
&=\; \bar{\pi}_i(\theta_i, \delta_i) + \bar{t}_i(\theta_i, \delta_i) + \delta_i \bar{\pi}_{-i}(\theta_i, \delta_i) + \delta_i \bar{t}_{-i}(\theta_i, \delta_i) \\
&=\; \bar{\pi}_i(\theta_i, \delta_i) + \bar{t}_i(\theta_i, \delta_i) + \delta_i \frac{dp_i(\delta_i)}{d\delta_i} + \delta_i \frac{\partial}{\partial \delta_i}\int_{\theta_i^{\min}}^{\theta_i} \bar{v}_i(s, \delta_i)\, ds.
\end{aligned}$$

Hence, T^* must also satisfy the identity

$$\begin{aligned}
\text{(53)} \qquad \bar{t}_i(\theta_i, \delta_i) \;=\;& p_i(\delta_i) - \delta_i \frac{dp_i(\delta_i)}{d\delta_i} - \bar{\pi}_i(\theta_i, \delta_i) \\
&+ \int_{\theta_i^{\min}}^{\theta_i} \bar{v}_i(s, \delta_i)\, ds - \delta_i \frac{\partial}{\partial \delta_i}\int_{\theta_i^{\min}}^{\theta_i} \bar{v}_i(s, \delta_i)\, ds.
\end{aligned}$$

From identities (52) and (53), T^* can be "guessed". The specific choice of p_i ensures that truth-telling is the unique best response to i's problem.

Proof of Theorem 3 Extended

Notice first that, for all $(\hat{\theta}_1, \theta_2) \in \Theta$ and all $\delta \in \Delta$, the functions S_i and T^* satisfy

$$\text{(54)} \qquad \mathbb{E}_{\theta_2}\big[t_1^*(\hat{\theta}_1, \theta_2, \delta) + \delta_1 t_2^*(\hat{\theta}_1, \theta_2, \delta)\big] \;=\; \mathbb{E}_{\theta_2}\big[S_1(\hat{\theta}_1, \theta_2, \delta)\big] - \mathbb{E}_{\theta_1,\theta_2}\left[S_1(\theta_1, \theta_2, \delta)\right],$$

$$\text{(55)} \qquad \mathbb{E}_{\theta_2}\big[S_1(\hat{\theta}_1, \theta_2, \delta)\big] \;=\; \int_{\theta_1^{\min}}^{\hat{\theta}_1} \mathbb{E}_{\theta_2}\big[v_1(k^*(s, \theta_2, \delta))\big]\, ds - \mathbb{E}_{\theta_2}\big[\pi_1(k^*(\hat{\theta}_1, \theta_2, \delta) \,|\, \hat{\theta}_1)\big] - \delta_1 \cdot \mathbb{E}_{\theta_2}\big[\pi_2(k^*(\hat{\theta}_1, \theta_2, \delta) \,|\, \theta_2\big].$$

Suppose agent 2 reports her payoff type truthfully. By (54) and (55), agent 1's interim expected utility from reporting $\hat{\theta}_1$ if her true type is θ_1 satisfies

$$\mathbb{E}_{\theta_2}\big[u_1(\,\cdot\,)\big] \;=\; \mathbb{E}_{\theta_2}\big[\pi_1(k^*(\hat{\theta}_1, \theta_2, \delta) \,|\, \theta_1)\big] - \mathbb{E}_{\theta_2}\big[\pi_1(k^*(\hat{\theta}_1, \theta_2, \delta) \,|\, \hat{\theta}_1)\big] - \mathbb{E}_{\theta_1,\theta_2}\left[S_1(\theta_1, \theta_2, \delta)\right] + \int_{\theta_1^{\min}}^{\hat{\theta}_1} \mathbb{E}_{\theta_2}\big[v_1(k^*(s, \theta_2, \delta))\big]\, ds.$$

Her marginal utility is given by $\frac{\partial}{\partial \hat{\theta}_1}\mathbb{E}_{\theta_2}\big[u_1(\,\cdot\,)\big] = (\theta_1 - \hat{\theta}_1) \cdot \frac{\partial}{\partial \hat{\theta}_1}\mathbb{E}_{\theta_2}\Big[v_1(k^*(\hat{\theta}_1, \theta_2, \delta))\Big]$, where the last factor is non-negative by assumption. Hence, truthful revelation is optimal for agent 1. By symmetry, $\hat{\theta}_2 = \theta_2$. As the above arguments hold for any set of type distributions, T^* *strongly* Bayesian implements k^*.

Derivation of the transfer scheme T^* in the proof of Theorem 3

Suppose externality types are common knowledge, and assume that the partially continuously differentiable allocation rule k^* is strongly Bayesian implemented by the ex post budget-balanced transfer scheme $T = (t_1, t_2)$. With notation adopted from the proof Lemma 1, agent i of true type (θ_i, δ_i) reports her payoff type $\hat{\theta}_i$ so as to maximize her interim expected utility,

$$U_i(\hat{\theta}_i \,|\, \theta_i, \delta) \;=\; \theta_i \bar{v}_i(\hat{\theta}_i, \delta) + \bar{h}_i(\hat{\theta}_i, \delta) + \bar{t}_i(\hat{\theta}_i, \delta) + \delta_i \bar{\pi}_{-i}(\hat{\theta}_i, \delta) + \delta_i \bar{t}_{-i}(\hat{\theta}_i, \delta).$$

Condition (41) in the proof Lemma 1 states that T must satisfy

$$\text{(56)} \qquad U_i(\theta_i \,|\, \theta_i, \delta) \;=\; p_i(\delta) + \int_{\theta_i^{\min}}^{\theta_i} \bar{v}_i(s, \delta)\, ds$$

for some function $p_i : \Delta \to \mathbb{R}$. For ease of notation, write $t_i = t_i(\theta, \delta)$ and $\pi_i = \pi_i(k^*(\theta, \delta) \,|\, \theta_i)$. Then, by (56), T satisfies the following identities:

$$\text{(57)} \qquad \mathbb{E}_{\theta_2}\big[t_1\big] + \delta_1 \mathbb{E}_{\theta_2}\big[t_2\big] = p_1(\delta) + \int_{\theta_1^{\min}}^{\theta_1} \bar{v}_1(s, \delta)\, ds - \mathbb{E}_{\theta_2}\big[\pi_1\big] - \delta_1 \mathbb{E}_{\theta_2}\big[\pi_2\big],$$

$$\text{(58)} \qquad \mathbb{E}_{\theta_1}\big[t_2\big] + \delta_2 \mathbb{E}_{\theta_1}\big[t_1\big] = p_2(\delta) + \int_{\theta_2^{\min}}^{\theta_2} \bar{v}_2(s, \delta)\, ds - \mathbb{E}_{\theta_1}\big[\pi_2\big] - \delta_2 \mathbb{E}_{\theta_1}\big[\pi_1\big].$$

Due to budget balance, (57) and (58) imply that interim expected transfers satisfy

$$\begin{aligned} (1-\delta_1)\mathbb{E}_{\theta_2}\big[t_1\big] &= p_1(\delta) + \int_{\theta_1^{\min}}^{\theta_1} \bar{v}_1(s, \delta)\, ds - \mathbb{E}_{\theta_2}\big[\pi_1\big] - \delta_1 \mathbb{E}_{\theta_2}\big[\pi_2\big], \\ -(1-\delta_2)\mathbb{E}_{\theta_1}\big[t_1\big] &= p_2(\delta) + \int_{\theta_2^{\min}}^{\theta_2} \bar{v}_2(s, \delta)\, ds - \mathbb{E}_{\theta_1}\big[\pi_2\big] - \delta_2 \mathbb{E}_{\theta_1}\big[\pi_1\big]. \end{aligned}$$

From these conditions, T^* can be "guessed".

Proof of Proposition 6 Continued

It remains to show that the bargaining solutions (24) and (25) each satisfy $\operatorname{sgn}(\frac{\partial k^*}{\partial \theta_1}\frac{\partial k^*}{\partial \theta_2}) = -1 = \operatorname{sgn}(\frac{\partial k^*}{\partial \delta_1}\frac{\partial k^*}{\partial \delta_2})$, where $\Delta_i \subset (-1, \frac{\theta_i^{\min}}{\theta_{-i}^{\max}})$ is assumed for (25).

Implicit differentiation of (24) yields

$$\begin{aligned} \frac{\partial F}{\partial k^*}\frac{\partial k^*}{\partial \delta_1} &= \theta_2 v(1-k^*), \\ \frac{\partial F}{\partial k^*}\frac{\partial k^*}{\partial \delta_2} &= -\theta_1 v(k^*), \\ \frac{\partial F}{\partial k^*}\frac{\partial k^*}{\partial \theta_1} &= (1-\delta_2) v(k^*), \\ \frac{\partial F}{\partial k^*}\frac{\partial k^*}{\partial \theta_2} &= -(1-\delta_1) v(1-k^*), \end{aligned}$$

where $\frac{\partial F}{\partial k^*} = -\theta_2(1-\delta_1)v'(1-k^*) - \theta_1(1-\delta_2)v'(k^*) < 0$. Hence, $\frac{\partial k^*}{\partial \delta_1} < 0 < \frac{\partial k^*}{\partial \delta_2}$ and $\frac{\partial k^*}{\partial \theta_1} < 0 < \frac{\partial k^*}{\partial \theta_2}$.

Implicit differentiation of (25) yields

$$\frac{\partial G}{\partial k^*}\frac{\partial k^*}{\partial \delta_1} = \theta_2^2 v(1-k^*), \tag{59}$$
$$\frac{\partial G}{\partial k^*}\frac{\partial k^*}{\partial \delta_2} = -\theta_1^2 v(k^*), \tag{60}$$
$$\frac{\partial G}{\partial k^*}\frac{\partial k^*}{\partial \theta_1} = (\theta_2 - 2\delta_2\theta_1)v(k^*) - \theta_2 v(1-k^*), \tag{61}$$
$$\frac{\partial G}{\partial k^*}\frac{\partial k^*}{\partial \theta_2} = \theta_1 v(k^*) - (\theta_1 - 2\delta_1\theta_2)v(1-k^*), \tag{62}$$

where $\frac{\partial G}{\partial k^*} = -\theta_2(\theta_1 - \delta_1\theta_2)v'(1-k^*) - \theta_1(\theta_2 - \delta_2\theta_1)v'(k^*) < 0$ for $\delta_i^{\max} < \frac{\theta_i^{\min}}{\theta_{-i}^{\max}}$. Hence, $\mathrm{sgn}(\frac{\partial k^*}{\partial \delta_1}\frac{\partial k^*}{\partial \delta_2}) = -1$. By substituting for (25) in (61) and (62), one observes that

$$\frac{\partial G}{\partial k^*}\frac{\partial k^*}{\partial \theta_1} = -(\delta_1\theta_2^2 - 2\delta_1\delta_2\theta_1\theta_2 + \delta_2\theta_1^2)\frac{v(k^*)}{\theta_1 - \delta_1\theta_2} = -\frac{\theta_2}{\theta_1}\frac{\partial G}{\partial k^*}\frac{\partial k^*}{\partial \theta_2}. \tag{63}$$

Hence, $\mathrm{sgn}(\frac{\partial k^*}{\partial \theta_1}\frac{\partial k^*}{\partial \theta_2}) = -1$. In particular, $\frac{\partial k^*}{\partial \theta_2} < 0 < \frac{\partial k^*}{\partial \theta_1}$ if $\Delta_i \subset (0, \frac{\theta_i^{\min}}{\theta_{-i}^{\max}})$, since then $\delta_1\theta_2^2 - 2\delta_1\delta_2\theta_1\theta_2 + \delta_2\theta_1^2 = (\delta_1\theta_2 - \delta_2\theta_1)^2 + \delta_1(1-\delta_1)\theta_2^2 + \delta_2(1-\delta_2)\theta_1^2 > 0$.

References

Arrow, Kenneth J. 1973. “Some ordinalist-utilitarian notes on Rawls’s theory of justice.” *Journal of Philosophy* 70 (9):245–263.

———. 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.

Atkinson, Anthony B. 1970. “On the measurement of inequality.” *Journal of Economic Theory* 2 (3):244–263.

Ausubel, Lawrence M, Peter Cramton, and Raymond J Deneckere. 2002. “Bargaining with incomplete information.” *Handbook of Game Theory with Economic Applications* 3:1897–1945.

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.

———. 2011. "Robust implementation in general mechanisms." *Games and Economic Behavior* 71 (2):261–281.

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.

Bisin, Alberto and Thierry Verdier. 1998. "On the cultural transmission of preferences for social status." *Journal of Public Economics* 70 (1):75–97.

Bowles, Samuel and Sung-Ha Hwang. 2008. "Social preferences and public economics: Mechanism design when social preferences depend on incentives." *Journal of Public Economics* 92 (8):1811–1820.

Cabrales, Antonio and Roberto Serrano. 2011. "Implementation in adaptive better-response dynamics: Towards a general theory of bounded rationality in mechanisms." *Games and Economic Behavior* 73 (2):360–374.

Chung, Kim-Sau and Jeffrey C Ely. 2007. "Foundations of dominant-strategy mechanisms." *Review of Economic Studies* 74 (2):447–476.

Clark, Andrew E, Paul Frijters, and Michael A Shields. 2008. "Relative income, happiness, and utility: An explanation for the Easterlin paradox and other puzzles." *Journal of Economic Literature* 46 (1):95–144.

Clarke, Edward H. 1971. "Multipart pricing of public goods." *Public Choice* 11 (1):17–33.

Dasgupta, Partha and Eric Maskin. 2000. "Efficient auctions." *Quarterly Journal of Economics* 115 (2):341–388.

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

d'Aspremont, Claude and Louis Gevers. 1977. "Equity and the informational basis of collective choice." *Review of Economic Studies* 44 (2):199–209.

de Clippel, Geoffroy. 2014. "Behavioral implementation." *American Economic Review* 104 (10):2975–3002.

Gershkov, Alex, Jacob K Goeree, Alexey Kushnir, Benny Moldovanu, and Xianwen Shi. 2013. "On the equivalence of Bayesian and dominant strategy implementation." *Econometrica* 81 (1):197–220.

Glazer, Jacob and Ariel Rubinstein. 1998. "Motives and implementation: On the design of mechanisms to elicit opinions." *Journal of Economic Theory* 79 (2):157–173.

Goeree, Jacob K, Emiel Maasland, Sander Onderstal, and John L Turner. 2005. "How (not) to raise money." *Journal of Political Economy* 113 (4):897–918.

Groves, Theodore. 1973. "Incentives in teams." *Econometrica* 41 (4):617–631.

Harsanyi, John C. 1955. "Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility." *Journal of Political Economy* 63 (4):309–321.

Harsanyi, John C and Reinhard Selten. 1972. "A generalized Nash solution for two-person bargaining games with incomplete information." *Management Science* 18 (5-part-2):80–106.

Harstad, Bård. 2007. "Harmonization and side payments in political cooperation." *American Economic Review* 97 (3):871–889.

Heffetz, Ori and Robert H Frank. 2008. "Preferences for status: Evidence and economic implications." In *Handbook of Social Economics*, vol. 1, edited by Jess Benhabib, Alberto Bisin, and Matthew Jackson. Netherlands: North Holland, 69–91.

Hirschman, Albert O and Michael Rothschild. 1973. "The changing tolerance for income inequality in the course of economic development: With a mathematical appendix." *Quarterly Journal of Economics* 87 (4):544–566.

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.

Jehiel, Philippe, Benny Moldovanu, and Ennio Stacchetti. 1996. “How (not) to sell nuclear weapons.” *American Economic Review* 86 (4):814–829.

———. 1999. “Multidimensional mechanism design for auctions with externalities.” *Journal of Economic Theory* 85 (2):258–293.

Kalai, Ehud. 1977. “Proportional solutions to bargaining situations: Interpersonal utility comparisons.” *Econometrica* 45 (7):1623–1630.

Kalai, Ehud and Meir Smorodinsky. 1975. “Other solutions to Nash’s bargaining problem.” *Econometrica* 43 (3):513–518.

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

Lu, Jingfeng. 2012. “Optimal auctions with asymmetric financial externalities.” *Games and Economic Behavior* 74 (2):561–575.

Maskin, Eric. 1978. “A theorem on utilitarianism.” *Review of Economic Studies* 45 (1):93–96.

———. 1979. “Decision-making under ignorance with implications for social choice.” In *Game Theory, Social Choice and Ethics*. Springer, 319–337.

McFadden, Daniel. 2009. “The human side of mechanism design: A tribute to Leo Hurwicz and Jean-Jacque Laffont.” *Review of Economic Design* 13 (1-2):77–100.

Mookherjee, Dilip and Stefan Reichelstein. 1992. “Dominant strategy implementation of Bayesian incentive compatible allocation rules.” *Journal of Economic Theory* 56 (2):378–399.

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

Nash, John F. 1950. “The bargaining problem.” *Econometrica* 18 (2):155–162.

Posner, Richard A. 1979. "Utilitarianism, economics, and legal theory." *Journal of Legal Studies* 8 (1):103–140.

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

Rawls, John B. 1971. *A theory of justice*. Harvard University Press.

Rubinstein, Ariel. 1982. "Perfect equilibrium in a bargaining model." *Econometrica* 50 (1):97–109.

Saez, Emmanuel and Stefanie Stantcheva. 2016. "Generalized social marginal welfare weights for optimal tax theory." *American Economic Review* 106 (1):24–45.

Sen, Amartya. 1973. *On economic inequality*. Oxford University Press.

———. 1974. "Informational bases of alternative welfare approaches: Aggregation and income distribution." *Journal of Public Economics* 3 (4):387–403.

———. 1979. "Utilitarianism and welfarism." *Journal of Philosophy* 76 (9):463–489.

Tang, Pingzhong and Tuomas Sandholm. 2012. "Optimal auctions for spiteful bidders." In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*. pp. 1457–1463.

Thomson, William. 1994. "Cooperative models of bargaining." *Handbook of Game Theory with Economic Applications* 2:1237–1284.

Tran, Anh and Richard Zeckhauser. 2012. "Rank as an inherent incentive: Evidence from a field experiment." *Journal of Public Economics* 96 (9):645–650.

Vickrey, William. 1961. "Counterspeculation, auctions, and competitive sealed tenders." *Journal of Finance* 16 (1):8–37.

Weiss, Yoram and Chaim Fershtman. 1998. "Social status and economic performance: A survey." *European Economic Review* 42 (3):801–820.

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.