

Daske, Thomas

**Preprint**

## Externality Assessments, Welfare Judgments, and Mechanism Design

*Suggested Citation:* Daske, Thomas (2017) : Externality Assessments, Welfare Judgments, and Mechanism Design, Technische Universität München, München

This Version is available at:

<https://hdl.handle.net/10419/172494>

**Standard-Nutzungsbedingungen:**

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

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

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

**Terms of use:**

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

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

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

# Externality Assessments, Welfare Judgments, and Mechanism Design\*

Thomas Daske <sup>†</sup>

December 27, 2017

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 the sum of 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, robust implementation, social welfare, bargaining

---

\*I am 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. For their critical remarks and helpful comments I am particularly indebted to Michael Kurschilgen, Christoph March, Marco Sahn, and Robert von Weizsäcker.

<sup>†</sup>Technical University of Munich, TUM School of Management, Arcisstr. 21, 80333 Munich, Germany. Email: [thomas.daske@tum.de](mailto:thomas.daske@tum.de).

# 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.<sup>1</sup>

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.<sup>2</sup>

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.<sup>3</sup> 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

---

<sup>1</sup>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.

<sup>2</sup>This scenario has been analyzed by [Harstad \(2007\)](#), under the assumption of commonly known externalities though.

<sup>3</sup>E.g., [Becker \(1981\)](#).

to despise each other (their externality types).<sup>4</sup> On the other hand, empirical studies have found that many, if not all, people care about their relative standing in society.<sup>5</sup> 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 (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.<sup>6</sup> 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

---

<sup>4</sup>With regard to cross-ownership as outlined above, one can also think of two rulers in the cameralist era of European history who are related by marriage and negotiate the division of land.

<sup>5</sup>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).

<sup>6</sup>Bergemann and Morris (2005) show that Bayesian implementable allocation rules can, in many cases, no longer be ex post implemented when requiring budget balance.

their types truthfully.<sup>7</sup> 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 social 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 the sum of private payoffs exclusive of externalities. These allocations are Pareto-efficient 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’, social welfare judgments of the mechanism designer, as it implies that she is not indifferent to even small changes in payoff types.<sup>8</sup>

---

<sup>7</sup>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.

<sup>8</sup>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.

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 the sum of private payoffs exclusive of externalities; I call the social 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 a social 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 social 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 social welfare judgments and budget balance. Theorem 2 shows that, even with asymmetric information about externality assessments, nearly any social 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 social 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 social welfare judgments is obtained. Section 5 interprets results for strategic bargaining under incomplete information. Section 6 expands the central result to social welfare measures that incorporate the redistributive effects of the transfer scheme itself. Section 7 concludes.

## 2 Related Literature

This study relates to three strands of literature: ‘robust’ implementation, implementation in the presence of externalities, and (perhaps most importantly) 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.<sup>9</sup> Jehiel et al. (2006) consider a model framework that entails the one presented here, with 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. The implications of their result for the questions addressed here are discussed in detail at the end of Section 4.

Several studies have explored ex post or Bayesian implementation under the assumption that externalities are common knowledge.<sup>10</sup> The present study considers Bayesian implementation while relaxing this assumption.<sup>11</sup> 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.<sup>12</sup> 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 social 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.<sup>13</sup> The present study, in a slightly different

---

<sup>9</sup>E.g., Mookherjee and Reichelstein (1992), Dasgupta and Maskin (2000), Bergemann and Morris (2005, 2011), Chung and Ely (2007), Gershkov et al. (2013).

<sup>10</sup>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).

<sup>11</sup>Many of the studies on implementation in the presence of externalities are devoted to auction theory. Notice that the here derived propositions have only limited relevance for auctions, since I am concerned with budget balance while auction theory is typically concerned with revenue maximization. Moreover, I deal with continuous allocation rules whereas, in auctions, allocation rules are typically discrete.

<sup>12</sup>They refer to an allocation as ‘efficient’ if it maximizes the sum of payoffs *inclusive* of externalities.

<sup>13</sup>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.



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 the sum of 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 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.<sup>14</sup> 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](#)).<sup>15</sup> 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.<sup>16</sup>

More generally, Theorems 1 to 3 suggest that theories of ‘efficient’ implementation depend critically on their underlying social 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:<sup>17</sup> 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 will affect themselves.<sup>18</sup> It seems plausible in all these cases that a social welfare measure should not condition on such ‘behavioral’ externality assessments, and that mechanisms designed to implement ‘efficient’ allocations should be externality-robust.

---

<sup>14</sup>For critical reflections of utilitarianism see, e.g., [Posner \(1979\)](#) and [Sen \(1973, 1979\)](#).

<sup>15</sup>For a discussion of the CES-welfare measures see also [Sen \(1974\)](#).

<sup>16</sup>[Saez and Stantcheva \(2016\)](#), for instance, characterize optimal taxation under non-utilitarian social welfare measures—in the (somewhat unrealistic) absence of externalities.

<sup>17</sup>E.g., [Glazer and Rubinstein \(1998\)](#), [Cabrales and Serrano \(2011\)](#), [de Clippel \(2014\)](#), [Bierbrauer and Netzer \(2016\)](#), and [Bartling and Netzer \(2016\)](#).

<sup>18</sup>In this respect, this study draws a mechanism design perspective on the ‘tunnel effect’ of [Hirschman and Rothschild \(1973\)](#).



### 3 The Model

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 payoff

$$(1) \quad \pi_i(k, t_i | \theta_i) = \theta_i v_i(k) + h_i(k) + t_i,$$

where the functions  $v_i : K \rightarrow [0, \infty)$  and  $h_i : K \rightarrow \mathbb{R}$  are twice continuously differentiable and satisfy  $\partial^2 \pi_i(k, t_i | \theta_i) / \partial k^2 < 0$  for all  $i, k$ , and  $\theta_i$ ; furthermore, either  $dv_i/dk > 0$  for all  $k$  and  $i$ , or  $dv_i/dk < 0$  for all  $k$  and  $i$ . The functions  $v_i, h_i$  are common knowledge. Agent  $i$ 's *payoff type*  $\theta_i$  is drawn from an interval  $\Theta_i = (\theta_i^{\min}, \theta_i^{\max})$ , with  $0 \leq \theta_i^{\min} < \theta_i^{\max}$ . Payoff types are private information and are distributed according to a continuous density function  $f_i > 0$ . From the allocation of payoffs, agent  $i$  gains utility

$$(2) \quad u_i(k, t_i, t_{-i}, \theta_{-i} | \theta_i, \delta_i) = \pi_i(k, t_i | \theta_i) + \delta_i \cdot \pi_{-i}(k, t_{-i} | \theta_{-i}),$$

where  $i$ 's *externality type*  $\delta_i$  is drawn from an interval  $\Delta_i = (\delta_i^{\min}, \delta_i^{\max}) \subset [-1, 1]$ , with  $\delta_i^{\min} < \delta_i^{\max}$ . Externality types are private information and are distributed according to a continuous density function  $g_i(\cdot | \theta_i) > 0$ . That is, an agent's externality type may correlate with her payoff type, not perfectly though. Notice also that 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.

Denote by  $H_i$  the joint c.d.f. of agent  $i$ 's *type*,  $(\theta_i, \delta_i)$ . While types are private information, type distributions  $H_i$  are common knowledge. Types are independent across agents; that is,  $H_1$  and  $H_2$  are stochastically independent.

Denote by  $\Theta$  and  $\Delta$ , 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  $X : \Theta \times \Delta \rightarrow \mathbb{R}$ , denote by  $\mathbb{E}_{\theta_i, \delta_i} [X(\theta, \delta)]$  the expected value of  $X$  for given values of  $\theta_{-i}$  and  $\delta_{-i}$ .<sup>19</sup>

$$\mathbb{E}_{\theta_i, \delta_i} [X(\theta, \delta)] = \int_{\theta_i^{\min}}^{\theta_i^{\max}} \left( \int_{\delta_i^{\min}}^{\delta_i^{\max}} X(\theta, \delta) g_i(\delta_i | \theta_i) d\delta_i \right) f_i(\theta_i) d\theta_i.$$

A *direct revelation mechanism* involves the agents in a strategic game. In this game, agents are asked to report their types truthfully.<sup>20</sup> Based on their reports, a social alternative will be implemented and transfers will be made. Specifically, the mechanism is defined by an *allocation rule*  $k : \Theta \times \Delta \rightarrow K$  and a *transfer scheme*  $T = (t_1, t_2) : \Theta \times \Delta \rightarrow \mathbb{R}^2$ . In what follows, attention will be restricted to transfer schemes  $T$  that are

<sup>19</sup>Likewise, denote by  $\mathbb{E}_{\theta_i} [Y(\theta)]$  the expected value of  $Y : \Theta \rightarrow \mathbb{R}$  for a given value of  $\theta_{-i}$ .

<sup>20</sup>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.

continuous on the externality-type space  $\Delta$ . An allocation rule  $k$  is said to be *Bayesian implementable*, if there exists a transfer scheme  $(t_1, t_2)$  such that both

$$\begin{aligned} (\theta_1, \delta_1) &\in \arg \max_{\hat{\theta}_1, \hat{\delta}_1} \mathbb{E}_{\theta_2, \delta_2} \left[ u_1(k(\hat{\theta}_1, \hat{\delta}_1, \theta_2, \delta_2), t_1(\hat{\theta}_1, \hat{\delta}_1, \theta_2, \delta_2), t_2(\hat{\theta}_1, \hat{\delta}_1, \theta_2, \delta_2), \theta_2 \mid \theta_1, \delta_1) \right], \\ (\theta_2, \delta_2) &\in \arg \max_{\hat{\theta}_2, \hat{\delta}_2} \mathbb{E}_{\theta_1, \delta_1} \left[ u_2(k(\theta_1, \delta_1, \hat{\theta}_2, \hat{\delta}_2), t_1(\theta_1, \delta_1, \hat{\theta}_2, \hat{\delta}_2), t_2(\theta_1, \delta_1, \hat{\theta}_2, \hat{\delta}_2), \theta_1 \mid \theta_2, \delta_2) \right]. \end{aligned}$$

That is, truthful revelation maximizes each agent's interim expected utility provided the respective other agent reveals her type truthfully.

The mechanism is said to be *ex post budget-balanced* if the transfer scheme satisfies  $t_1 + t_2 = 0$  for any realization of types, such that agents neither have to have access to an external source of money, nor that 'money is burned'.

The following two definitions restrict the domain of allocation rules to be considered in the next sections. For that purpose, define

$$\begin{aligned} \pi_i(k \mid \theta_i) &= \theta_i v_i(k) + h_i(k), \quad \text{and} \\ u_i(k, \theta_{-i} \mid \theta_i, \delta_i) &= \pi_i(k \mid \theta_i) + \delta_i \pi_{-i}(k \mid \theta_{-i}), \end{aligned}$$

and denote by  $\text{sgn} : \mathbb{R} \rightarrow \{-1, 0, 1\}$  the *sign function*.<sup>21</sup>

### Definition 1 (Sensitivity)

Let  $W : \mathbb{R}^4 \rightarrow \mathbb{R}$  be twice partially continuously differentiable, and let  $V : K \rightarrow \mathbb{R}$ ,  $V(k) = W(\pi_1(k \mid \theta_1), \delta_1 \pi_2(k \mid \theta_2), \pi_2(k \mid \theta_2), \delta_2 \pi_1(k \mid \theta_1))$ .<sup>22</sup>  $W$  is said to be a *sensitive social welfare measure* if it has the following properties.

- (i)  $\partial W(\pi_1, \delta_1 \pi_2, \pi_2, \delta_2 \pi_1) / \partial \pi_i > 0$  for each  $i \in \{1, 2\}$ .
- (ii) *Pareto property*: If there exist  $k_1, k_2 \in K$  and  $i \in \{1, 2\}$  such that  $u_i(k_1, \theta_{-i} \mid \theta_i, \delta_i) > u_i(k_2, \theta_{-i} \mid \theta_i, \delta_i)$  and  $u_{-i}(k_1, \theta_i \mid \theta_{-i}, \delta_{-i}) \geq u_{-i}(k_2, \theta_i \mid \theta_{-i}, \delta_{-i})$ , then  $V(k_1) > V(k_2)$ .
- (iii) There exists a unique partially continuously differentiable allocation rule  $k^* : \Theta \times \Delta \rightarrow K$  such that  $k^*(\theta, \delta) = \arg \max_{k \in K} V(k)$ ,

$$(3) \quad 1 = \text{sgn} \left( \frac{\partial v_1(k^*)}{\partial \theta_2} \right) \cdot \text{sgn} \left( \frac{\partial v_2(k^*)}{\partial \theta_1} \right), \quad \text{and}$$

$$(4) \quad 0 = \text{sgn} \left( \frac{\partial \pi_1(k^* \mid \theta_1)}{\partial \delta_2} \right) - \text{sgn} \left( \frac{\partial \pi_2(k^* \mid \theta_2)}{\partial \delta_1} \right).$$

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

<sup>21</sup>For  $x \in \mathbb{R}$ , the sign of  $x$  is defined as  $\text{sgn}(x) = 1$  for  $x > 0$ ,  $\text{sgn}(x) = -1$  for  $x < 0$ , and  $\text{sgn}(0) = 0$ .

<sup>22</sup>This specification of a welfare measure with regard to the choice of  $k$  is without loss of generality as it allows for taking private payoffs,  $\pi_i$ , and externality types,  $\delta_i$ , as independent variables. For instance,  $V(k) = (1 + \delta_1^2)\pi_2 + (1 + \delta_2^2)\pi_1$  can be written as  $V(k) = (\pi_1) + (\pi_2) + \frac{(\delta_2 \pi_1)}{(\pi_1)}(\delta_2 \pi_1) + \frac{(\delta_1 \pi_2)}{(\pi_2)}(\delta_1 \pi_2)$ .

Whether a function qualifies as a sensitive social welfare measure is context-dependent, since the above conditions involve the functions  $v_i$  and  $h_i$ . A sensitive social welfare measure  $V(k)$  accounts separately for private payoffs,  $\pi_i(k | \theta_i)$ , and externalities,  $\delta_i \pi_{-i}(k | \theta_{-i})$ . This serves the purpose of clearly isolating the extent to which ‘efficient’ allocation rules may condition on externality assessments if they are to be Bayesian implemented in a budget-balanced way.

By condition (i), a marginal increase in one 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. According to equations (1) and (2), full ex post Pareto efficiency is realized if, in addition, transfers are budget-balanced.

Identities (3) and (4) are symmetry assumptions. Identity (3) requires that the effect of an increase in agent  $i$ ’s payoff type on agent  $-i$ ’s payoff, exclusive of  $h_{-i}(k^*)$ , is similar for all agents. As the functions  $v_i$  are assumed to be either strictly increasing or strictly decreasing, and since  $\partial v_i(k^*)/\partial \theta_{-i} = (dv_i(k^*)/dk)(\partial k^*/\partial \theta_{-i})$ , equation (3) requires in particular that a sensitive allocation rule is either strictly increasing or strictly decreasing in each agent’s payoff type. In this respect, it responds *sensitively* to changes in agents’ payoff characteristics.<sup>23</sup> Finally, identity (4) requires that the effect of an increase in one agent’s externality type on the other agent’s payoff is similar for all agents.

Several “classic” social welfare measures qualify as sensitive.<sup>24</sup>

**Proposition 1** *With notation as in Definition 1, each of the following social welfare measures  $W : \mathbb{R}^4 \rightarrow \mathbb{R}$  is sensitive if the economic environment is such that  $W$  induces a unique partially continuously differentiable function  $k^*(\theta, \delta) = \arg \max_{k \in K} V(k)$  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) *Externality-sensitive 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  $h_i \equiv 0$  and  $\Delta_i \subset [0, 1]$  for all  $i \in \{1, 2\}$ , then the following social welfare measures are sensitive.*

(iii) *“Social utility weights”, inclusive of externalities:*

$$W = \alpha_1 u_1(k, \theta_2 | \theta_1, \delta_1) + \alpha_2 u_2(k, \theta_1 | \theta_2, \delta_2), \text{ with } \alpha_1, \alpha_2 > 0.$$

<sup>23</sup>Notice also that condition (iii) of Definition 1 requires the economic environment as well as a sensitive social welfare measure to allow for *interior* solutions to  $\max_{k \in K} V(k)$ . Hence,  $k^*$  must satisfy the first-order condition  $dV(k^*(\theta, \delta))/dk = 0$  and the second-order condition  $d^2V(k^*(\theta, \delta))/dk^2 < 0$  for each  $(\theta, \delta) \in \Theta \times \Delta$ .

<sup>24</sup>Notice that condition (4) of Definition 1 precludes the dictatorial social welfare measure  $V(k) = u_i(k, \theta_{-i} | \theta_i, \delta_i)$  from being sensitive, since then  $\partial \pi_i(k^* | \theta_i)/\partial \delta_{-i} = 0$ , whereas  $\partial \pi_{-i}(k^* | \theta_{-i})/\partial \delta_i \neq 0$ . Notice further that externality-ignoring discriminatory utilitarianism of the form  $W = \alpha_1 \pi_1(k | \theta_1) + \alpha_2 \pi_2(k | \theta_2)$ , with  $\alpha_1, \alpha_2 > 0$  and  $\alpha_1 \neq \alpha_2$ , satisfies all the conditions of Definition 1 but might not have the Pareto property.

(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, inclusive of externalities:*

$$W = [(u_1(k, \theta_2 | \theta_1, \delta_1))^{-\rho} + (u_2(k, \theta_1 | \theta_2, \delta_2))^{-\rho}]^{-\frac{1}{\rho}}, \text{ with } \rho \in (-1, \infty) \setminus \{0\}.$$

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

By means of the next definition, attention will be further restricted to those Bayesian implementable allocation rules that do not condition on (moments of) type distributions.

**Definition 2 (Strong Bayesian implementability)**

*An allocation rule  $k^* : \Theta \times \Delta \rightarrow K$  is said to be strongly Bayesian implementable if it is Bayesian implementable for any set of (non-degenerate) type distributions,  $\{F_1, G_1, F_2, G_2\}$ .*

Strong Bayesian implementability is critical to the results obtained below.<sup>25</sup> It 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 social welfare judgment 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 ex post*.<sup>26</sup>

### 3.1 Altruism, Spite, and Status Considerations

Evidently, the model captures the linear conceptions of altruism and spite when interpreting externality types as the intensity of altruism or spite. It also captures linear conceptions of preferences for status:<sup>27</sup> Suppose the allocation of payoffs,  $\pi_1$  and  $\pi_2$ , yields agent  $i$  a utility level of  $u_i = \pi_i + \sigma_i(\pi_i - \pi_{-i})$ , with  $\sigma_i > 0$  determining  $i$ ’s preference for status. Maximizing  $u_i$  is then equivalent to maximizing  $\hat{u}_i = u_i/(1 + \sigma_i) = \pi_i + \delta_i\pi_{-i}$ , with externality type  $\delta_i = -\sigma_i/(1 + \sigma_i) \in (-1, 0)$ .

## 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).

<sup>25</sup>Strong Bayesian implementability should not be confused with notions of ‘robust’ implementation in the manner of Bergemann and Morris (2009, 2013).

<sup>26</sup>An example of a social welfare measure that *does* condition on type distributions is the *generalized Nash product* of Harsanyi and Selten (1972).

<sup>27</sup>See, e.g., Boskin and Sheshinski (1978) and Bisin and Verdier (1998). By the same token, the model captures linear versions of interdependent utilities in the manner of Hirschman and Rothschild (1973).

**Theorem 1** *A sensitive allocation rule  $k^* : \Theta \times \Delta \rightarrow K$  can be strongly Bayesian implemented with an ex post budget-balanced mechanism if and only if it maximizes the sum of 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  $(\theta, \delta)$ ; in particular,  $k^*$  is independent from externality types:  $k^* = k^*|_{\Theta}$ .*

*Any mechanism that (ordinarily) Bayesian implements  $k^*(\theta) = \arg \max_{k \in K} \pi_1(k | \theta_1) + \pi_2(k | \theta_2)$  is of AGV-type: For reported types  $(\hat{\theta}, \hat{\delta}) \in \Theta \times \Delta$ , transfers are given by*

$$(5) \quad t_1(\hat{\theta}, \hat{\delta}) = \mathbb{E}_{\theta_2}[\pi_2(k^*(\hat{\theta}_1, \theta_2) | \theta_2)] - \mathbb{E}_{\theta_1}[\pi_1(k^*(\theta_1, \hat{\theta}_2) | \theta_1)] + s(\hat{\theta}, \hat{\delta}),$$

$$(6) \quad t_2(\hat{\theta}, \hat{\delta}) = \mathbb{E}_{\theta_1}[\pi_1(k^*(\theta_1, \hat{\theta}_2) | \theta_1)] - \mathbb{E}_{\theta_2}[\pi_2(k^*(\hat{\theta}_1, \theta_2) | \theta_2)] - s(\hat{\theta}, \hat{\delta}),$$

where  $s : \Theta \times \Delta \rightarrow \mathbb{R}$  must be chosen such that  $\mathbb{E}_{\theta_{-i}, \delta_{-i}}[s(\theta, \delta)]$  is constant on  $\Theta_i \times \Delta_i$  for each  $i \in \{1, 2\}$ .<sup>28</sup>

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

In the following, I refer to the mechanisms specified by Theorem 1 as *AGV-type mechanisms* (after Arrow, 1979, and d'Aspremont and Gérard-Varet, 1979). Notice that, *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 as well as the sensitivity of externality-ignoring utilitarianism are to be addressed first.

**Proposition 2** *Suppose the allocation rule  $k^* : \Theta \rightarrow K$  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  $\theta \in \Theta$ . Then  $k^*$  is sensitive and can be strongly Bayesian implemented with the ex post budget-balanced AGV-type mechanisms.<sup>29</sup>*

**Proof.** In order to prove the sensitivity of  $k^*$ , it suffices to show that  $k^*$  does specify Pareto-efficient allocations. (Verification of the remaining properties of a sensitive allocation rule follows the lines of the proof of Proposition 1(ii).)

Suppose there exists an allocation  $k'(\theta, \delta)$  that, for some types  $(\theta, \delta)$ , 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  $\delta_1 \in \Delta_1 \subset (-1, 1)$ ,

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

<sup>28</sup>Such functions  $s$  can be smooth and non-constant; for instance,  $s(\theta, \delta) = (\theta_1 - \mathbb{E}_{\theta_1}[\theta_1])(\theta_2 - \mathbb{E}_{\theta_2}[\theta_2]) + (\delta_1 - \mathbb{E}_{\delta_1}[\delta_1])(\delta_2 - \mathbb{E}_{\delta_2}[\delta_2])$ .

<sup>29</sup>That AGV-type mechanisms are Bayesian incentive-compatible for other-regarding, spiteful agents has been shown earlier by Bartling and Netzer (2016).

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

Under AGV-type mechanisms in the manner of Theorem 1 (which, evidently, are ex post budget-balanced), and under the assumption that agent 2 reveals her type  $(\theta_2, \delta_2)$  truthfully, agent 1 chooses  $(\hat{\theta}_1, \hat{\delta}_1)$  so as to maximize her interim expected utility. Without loss of generality, normalize  $s(\hat{\theta}, \hat{\delta}) = 0$ . By equations (5) and (6),

$$\begin{aligned} & \mathbb{E}_{\theta_2, \delta_2} \left[ u_1(k^*(\hat{\theta}_1, \theta_2), t_1(\hat{\theta}_1, \theta_2, \hat{\delta}), t_2(\hat{\theta}_1, \theta_2, \hat{\delta}), \theta_2 \mid \theta_1, \delta_1) \right] \\ = & \mathbb{E}_{\theta_2} \left[ [\pi_1(k^*(\hat{\theta}_1, \theta_2) \mid \theta_1) + t_1(\hat{\theta}_1, \theta_2, \hat{\delta})] + \delta_1 \cdot [\pi_2(k^*(\hat{\theta}_1, \theta_2) \mid \theta_2) + t_2(\hat{\theta}_1, \theta_2, \hat{\delta})] \right] \\ = & \mathbb{E}_{\theta_2} [\pi_1(k^*(\hat{\theta}_1, \theta_2) \mid \theta_1) + \pi_2(k^*(\hat{\theta}_1, \theta_2) \mid \theta_2)] - (1 - \delta_1) \mathbb{E}_{\theta_1, \theta_2} [\pi_1(k^*(\theta_1, \theta_2) \mid \theta_1)], \end{aligned}$$

where the second term in the last line is independent from  $\hat{\theta}_1$ . Suppose truthfully reporting  $\theta_1$  is strictly inferior to some report  $\hat{\theta}_1 \neq \theta_1$ . Then there must exist some  $\theta_2$  such that

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

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 cannot do better than reporting  $(\theta_2, \delta_2)$ . As the argument holds for any set of type distributions, AGV-type mechanisms *strongly* Bayesian implement  $k^*$ . ■

As becomes clear from the proof of Proposition 2, the model assumption that each agent's marginal utility from her own payoff exceeds her marginal (dis-)utility from her opponent's payoff is indeed critical. For 'excessive' externalities,  $|\delta_i| > 1$ , externality-ignoring utilitarianism in the manner of Theorem 1 will not generally lead to ex post Pareto-efficient allocations. However, I have presented several examples of economic environments for which the assumption of 'moderate' externalities,  $|\delta_i| < 1$ , is reasonable.

Evidently, Proposition 2 holds for more general (e.g., multi-dimensional) sets of *payoff types* - a property of the AGV-mechanism which is well-known for environments without externality assessments (see, e.g., [Mas-Colell, Whinston, and Green, 1995](#), ch.23).

The following two propositions give proof of the necessary conditions of Theorem 1. These propositions successively constrain the domain of sensitive allocation rules and budget-balanced transfer schemes that allow for strong Bayesian implementation. They stipulate externality robustness in the sense that externality assessments are left inoperative from a strategic point of view. The following Lemma eases the exposition.

**Lemma 1** *Suppose the partially differentiable allocation rule  $k^* : \Theta \times \Delta \rightarrow K$  is strongly Bayesian implementable with an ex post budget-balanced mechanism. Then  $k^*$  satisfies*

$$(7) \quad (1 - \delta_i) \frac{\partial v_i(k^*(\theta, \delta))}{\partial \delta_i} = \left[ \frac{d\pi_i(k^*(\theta, \delta) \mid \theta_i)}{dk} + \frac{d\pi_{-i}(k^*(\theta, \delta) \mid \theta_{-i})}{dk} \right] \frac{\partial k^*(\theta, \delta)}{\partial \theta_i}$$

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 in a budget-balanced way only if the transfer to each agent  $i$  satisfies

$$(8) \quad \mathbb{E}_{\theta_{-i}, \delta_{-i}}[t_i(\theta, \delta)] = \alpha_i + \mathbb{E}_{\theta_{-i}, \delta_{-i}}[\pi_{-i}(k^*(\theta, \delta) | \theta_{-i})]$$

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 Lemma 1, the following proposition implies that the desired mechanism may not condition on externality types, such that externality assessments are not directly strategically operative.

**Proposition 3** *A sensitive allocation rule  $k^* : \Theta \times \Delta \rightarrow K$  is strongly Bayesian implementable with an ex post budget-balanced mechanism only if it is independent from externality types:  $k^* = k^*|_{\Theta}$ .*

**Proof.** Let  $k^* : \Theta \times \Delta \rightarrow K$  be the sensitive allocation rule that corresponds to a sensitive social welfare measure  $W : \mathbb{R}^4 \rightarrow \mathbb{R}$ . Ease notation by writing  $k^* = k^*(\theta, \delta)$ . It has to be shown that  $\partial k^* / \partial \delta_i = 0$  for all  $(\theta, \delta) \in \Theta \times \Delta$  and all  $i \in \{1, 2\}$ .

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

$$(9) \quad W_j = W_j(\pi_1(k^* | \theta_1), \delta_1 \pi_2(k^* | \theta_2), \pi_2(k^* | \theta_2), \delta_2 \pi_1(k^* | \theta_1)).$$

Then the conditions of Definition 1 imply that  $k^*$  satisfies the FOC

$$(10) \quad 0 = \frac{dV(k^*)}{dk} = [W_1 + \delta_2 W_4] \frac{d\pi_1(k^* | \theta_1)}{dk} + [W_3 + \delta_1 W_2] \frac{d\pi_2(k^* | \theta_2)}{dk},$$

where  $W_1 + \delta_2 W_4 = \partial W / \partial \pi_1 > 0$  and  $W_3 + \delta_1 W_2 = \partial W / \partial \pi_2 > 0$  by Definition 1(i).

By Lemma 1,  $k^*$  satisfies also

$$(11) \quad (1 - \delta_1) \frac{\partial v_1(k^*)}{\partial \delta_1} = \left[ \frac{d\pi_1(k^* | \theta_1)}{dk} + \frac{d\pi_2(k^* | \theta_2)}{dk} \right] \frac{\partial k^*}{\partial \theta_1},$$

$$(12) \quad (1 - \delta_2) \frac{\partial v_2(k^*)}{\partial \delta_2} = \left[ \frac{d\pi_1(k^* | \theta_1)}{dk} + \frac{d\pi_2(k^* | \theta_2)}{dk} \right] \frac{\partial k^*}{\partial \theta_2}.$$

Substituting (10) into (11) and (12) yields

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

$$(14) \quad (1 - \delta_2) \frac{\partial v_2(k^*)}{\partial \delta_2} = \left[ 1 - \frac{W_3 + \delta_1 W_2}{W_1 + \delta_2 W_4} \right] \frac{d\pi_2(k^* | \theta_2)}{dk} \frac{\partial k^*}{\partial \theta_2}.$$



On the other hand, as  $\partial k^*/\partial\theta_i \neq 0$  by Definition 1(iii), identities (11) and (12) jointly imply that

$$(15) \quad (1 - \delta_1) \frac{\partial v_1(k^*)}{\partial \delta_1} \frac{\partial k^*}{\partial \theta_2} = (1 - \delta_2) \frac{\partial v_2(k^*)}{\partial \delta_2} \frac{\partial k^*}{\partial \theta_1}.$$

As  $\delta_i < 1$  and  $dv_i/dk \neq 0$  by assumption, identity (15) 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  $(\theta, \delta)$  and all  $i$ . Then each of the factors on the right-hand sides of (13) and (14) is non-zero. In this case, (13) and (14) imply that

$$(16) \quad \frac{(W_3 + \delta_1 W_2)(1 - \delta_1) \frac{\partial v_1(k^*)}{\partial \delta_1}}{\frac{d\pi_1(k^* | \theta_1)}{dk} \frac{\partial k^*}{\partial \theta_1}} = [(W_3 + \delta_1 W_2) - (W_1 + \delta_2 W_4)] \\ = -[(W_1 + \delta_2 W_4) - (W_3 + \delta_1 W_2)] \\ = -\frac{(W_1 + \delta_2 W_4)(1 - \delta_2) \frac{\partial v_2(k^*)}{\partial \delta_2}}{\frac{d\pi_2(k^* | \theta_2)}{dk} \frac{\partial k^*}{\partial \theta_2}}.$$

Rearranging (16), while writing  $\frac{\partial v_i(k^*)}{\partial \delta_i} = \frac{dv_i(k^*)}{dk} \frac{\partial k^*}{\partial \delta_i}$ , yields the identity

$$(17) \quad (W_3 + \delta_1 W_2)(1 - \delta_1) \frac{dv_1(k^*)}{dk} \frac{\partial k^*}{\partial \delta_1} \frac{d\pi_2(k^* | \theta_2)}{dk} \frac{\partial k^*}{\partial \theta_2} \\ = -(W_1 + \delta_2 W_4)(1 - \delta_2) \frac{dv_2(k^*)}{dk} \frac{\partial k^*}{\partial \delta_2} \frac{d\pi_1(k^* | \theta_1)}{dk} \frac{\partial k^*}{\partial \theta_1}.$$

Since the terms  $(W_1 + \delta_2 W_4)$ ,  $(W_3 + \delta_1 W_2)$ , and  $(1 - \delta_i)$  are positive, application of the *sign function* to each side of identity (17) yields

$$(18) \quad \text{sgn} \left( \frac{d\pi_2(k^* | \theta_2)}{dk} \frac{\partial k^*}{\partial \delta_1} \frac{dv_1(k^*)}{dk} \frac{\partial k^*}{\partial \theta_2} \right) = -\text{sgn} \left( \frac{d\pi_1(k^* | \theta_1)}{dk} \frac{\partial k^*}{\partial \delta_2} \frac{dv_2(k^*)}{dk} \frac{\partial k^*}{\partial \theta_1} \right).$$

By Definition 1(iii),  $\text{sgn}(\partial v_1(k^*)/\theta_2) \cdot \text{sgn}(\partial v_2(k^*)/\theta_1) = 1$ , such that (18) can only hold if

$$(19) \quad \text{sgn} \left( \frac{\partial \pi_2(k^* | \theta_2)}{\partial \delta_1} \right) = -\text{sgn} \left( \frac{\partial \pi_1(k^* | \theta_1)}{\partial \delta_2} \right).$$

Equation (19) contradicts identity (4) of Definition 1, unless  $\partial \pi_i(k^* | \theta_i)/\partial \delta_{-i} = 0$  for all  $i$ . Suppose  $\partial \pi_1(k^* | \theta_1)/\partial \delta_2 = 0$ ; then multiplying (13) with  $\partial k^*(\theta, \delta)/\partial \delta_2$  implies that

$$(20) \quad (1 - \delta_1) \frac{\partial v_1(k^*)}{\partial \delta_1} \frac{\partial k^*(\theta, \delta)}{\partial \delta_2} = 0.$$

As  $\delta_i < 1$  and  $dv_i/dk \neq 0$  by assumption, (20) yields  $\frac{\partial k^*(\theta, \delta)}{\partial \delta_1} \frac{\partial k^*(\theta, \delta)}{\partial \delta_2} = 0$ , such that  $\frac{\partial k^*(\theta, \delta)}{\partial \delta_1} = 0 = \frac{\partial k^*(\theta, \delta)}{\partial \delta_2}$  due to (15) and the reasoning thereafter. Hence,  $k^* = k^*|_{\Theta}$ . ■

Externality assessments might indirectly become strategically operative if the allocation rule, even if independent from externality types, unfolds redistributive effects (beyond Benthamite utilitarianism). The next proposition states that strong, budget-balanced Bayesian implementation of a sensitive allocation rule is only feasible if the underlying social welfare measure treats agents' private payoffs as perfect substitutes.

**Proposition 4** *A sensitive allocation rule  $k^* : \Theta \rightarrow K$ , which is independent from externality types, is strongly Bayesian implementable with an ex post budget-balanced mechanism only if  $k^*(\theta) = \arg \max_{k \in K} \pi_1(k | \theta_1) + \pi_2(k | \theta_2)$  for all  $\theta \in \Theta$ ; any mechanism that (ordinarily) Bayesian implements this allocation rule is necessarily of AGV-type.*

**Proof.** If  $k^*$  is independent from externality types, identity (7) of Lemma 1 becomes

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

By Definition 1(iii), either  $\partial k^*/\partial \theta_i > 0$  for all  $\theta_i$ , or  $\partial k^*/\partial \theta_i < 0$  for all  $\theta_i$ . Hence,  $k^*(\theta) = \arg \max_{k \in K} \pi_1(k | \theta_1) + \pi_2(k | \theta_2)$  for all  $\theta \in \Theta$ .

Suppose there exists a budget-balanced transfer scheme  $T^* = (t_1^*, t_2^*) : \Theta \times \Delta \rightarrow \mathbb{R}^2$  that Bayesian implements  $k^*(\theta) = \arg \max_{k \in K} \pi_1(k | \theta_1) + \pi_2(k | \theta_2)$ . Notice that one can always write

$$(21) \quad t_1^*(\hat{\theta}, \hat{\delta}) = \mathbb{E}_{\theta_2} [\pi_2(k^*(\hat{\theta}_1, \theta_2) | \theta_2)] - \mathbb{E}_{\theta_1} [\pi_1(k^*(\theta_1, \hat{\theta}_2) | \theta_1)] + s_1(\hat{\theta}, \hat{\delta}),$$

$$(22) \quad t_2^*(\hat{\theta}, \hat{\delta}) = \mathbb{E}_{\theta_1} [\pi_1(k^*(\theta_1, \hat{\theta}_2) | \theta_1)] - \mathbb{E}_{\theta_2} [\pi_2(k^*(\hat{\theta}_1, \theta_2) | \theta_2)] + s_2(\hat{\theta}, \hat{\delta}),$$

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

$$(23) \quad \mathbb{E}_{\theta_{-i}, \delta_{-i}} [t_i^*(\theta, \delta)] = \mathbb{E}_{\theta_{-i}} [\pi_{-i}(k^*(\theta) | \theta_{-i})] - \mathbb{E}_{\theta_i, \theta_{-i}} [\pi_i(k^*(\theta) | \theta_i)] + \mathbb{E}_{\theta_{-i}, \delta_{-i}} [s_i(\theta, \delta)].$$

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

$$(24) \quad \mathbb{E}_{\theta_{-i}, \delta_{-i}} [t_i^*(\theta, \delta)] = \alpha_i + \mathbb{E}_{\theta_{-i}} [\pi_{-i}(k^*(\theta) | \theta_{-i})]$$

for all  $(\theta_i, \delta_i) \in \Theta_i \times \Delta_i$  and some constant  $\alpha_i \in \mathbb{R}$ . Jointly, identities (23) and (24) imply that  $\mathbb{E}_{\theta_{-i}, \delta_{-i}} [s_i(\theta, \delta)] = \alpha_i + \mathbb{E}_{\theta_i, \theta_{-i}} [\pi_i(k^*(\theta) | \theta_i)]$  for all  $(\theta_i, \delta_i)$ , so that  $\mathbb{E}_{\theta_{-i}, \delta_{-i}} [s_i(\theta, \delta)]$  must be constant on  $\Theta_i \times \Delta_i$ . Hence, the mechanism  $(k^*, T^*)$  is of AGV-type. ■

Propositions 2 to 4 give proof of Theorem 1. The next result emphasizes the critical role of budget balance when it comes to Bayesian implementation of social welfare judgments in the presence of asymmetric information about agents' externality assessments.

**Theorem 2** *If one waives budget balance, any twice continuously differentiable allocation rule  $k^* : \Theta \times \Delta \rightarrow K$  satisfying  $\min_{\theta_i, \delta_i} \frac{\partial}{\partial \theta_i} \mathbb{E}_{\theta_{-i}, \delta_{-i}} [v_i(k^*(\theta, \delta))] > 0$  for all  $i$  can be strongly Bayesian implemented. If  $k^* = k^*|_{\Theta}$ , then  $\frac{\partial}{\partial \theta_i} \mathbb{E}_{\theta_{-i}} [v_i(k^*(\theta, \delta))] \geq 0$  is sufficient.*

**Proof.** The straight forward proof for allocation rules satisfying  $k^* = k^*|_{\Theta}$  is put to the Appendix. Let  $k^* : \Theta \times \Delta \rightarrow K$ , with  $k^* \neq k^*|_{\Theta}$ , be a twice continuously differentiable allocation rule satisfying  $\beta_i > 0$  for  $\beta_i = \min_{\theta_i, \delta_i} \frac{\partial}{\partial \theta_i} \mathbb{E}_{\theta_{-i}, \delta_{-i}} [v_i(k^*(\theta, \delta))]$ . For functions  $p_i : \Delta_i \rightarrow \mathbb{R}$  define the transfer scheme  $T^* = (t_1^*, t_2^*) : \Theta \times \Delta \rightarrow \mathbb{R}^2$  by

$$(25) \quad \begin{aligned} t_i^*(\hat{\theta}, \hat{\delta}) &= p_i(\hat{\delta}_i) - \hat{\delta}_i \frac{\partial p_i(\hat{\delta}_i)}{\partial \hat{\delta}_i} + \int_{\theta_i^{\min}}^{\hat{\theta}_i} \mathbb{E}_{\theta_{-i}, \delta_{-i}} [v_i(k^*(s, \hat{\theta}_{-i}, \hat{\delta}_i, \hat{\delta}_{-i}))] ds \\ &\quad + \frac{\partial p_{-i}(\hat{\delta}_{-i})}{\partial \hat{\delta}_{-i}} + \frac{\partial}{\partial \hat{\delta}_{-i}} \int_{\theta_{-i}^{\min}}^{\hat{\theta}_{-i}} \mathbb{E}_{\theta_i, \delta_i} [v_{-i}(k^*(\hat{\theta}_i, s, \hat{\delta}_i, \hat{\delta}_{-i}))] ds \\ &\quad - \hat{\delta}_i \frac{\partial}{\partial \hat{\delta}_i} \int_{\theta_i^{\min}}^{\hat{\theta}_i} \mathbb{E}_{\theta_{-i}, \delta_{-i}} [v_i(k^*(s, \hat{\theta}_{-i}, \hat{\delta}_i, \hat{\delta}_{-i}))] ds \\ &\quad - \mathbb{E}_{\theta_{-i}, \delta_{-i}} [\pi_i(k^*(\hat{\theta}, \hat{\delta}) | \hat{\theta}_i)] - \mathbb{E}_{\theta_i, \delta_i} [\pi_i(k^*(\hat{\theta}, \hat{\delta}) | \hat{\theta}_i)]. \end{aligned}$$

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

$$(26) \quad \frac{\left[ \frac{\partial}{\partial \delta_i} \mathbb{E}_{\theta_{-i}, \delta_{-i}} [v_i(k^*(\theta, \delta))] \right]^2}{\frac{\partial}{\partial \theta_i} \mathbb{E}_{\theta_{-i}, \delta_{-i}} [v_i(k^*(\theta, \delta))]} < \frac{\partial^2}{\partial \delta_i^2} \left[ p_i(\delta_i) + \int_{\theta_i^{\min}}^{\theta_i} \mathbb{E}_{\theta_{-i}, \delta_{-i}} [v_i(k^*(s, \theta_{-i}, \delta_i, \delta_{-i}))] ds \right].$$

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

$$(27) \quad c_i = \gamma_i - \min_{\theta_i, \delta_i} \frac{\partial^2}{\partial \delta_i^2} \int_{\theta_i^{\min}}^{\theta_i} \mathbb{E}_{\theta_{-i}, \delta_{-i}} [v_i(k^*(s, \theta_{-i}, \delta_i, \delta_{-i}))] ds$$

for some constant  $\gamma_i$  satisfying  $\beta_i \cdot \gamma_i > \max_{\theta_i, \delta_i} \left[ \frac{\partial}{\partial \delta_i} \mathbb{E}_{\theta_{-i}, \delta_{-i}} [v_i(k^*(\theta, \delta))] \right]^2$ .<sup>30</sup> For an extensive proof of this claim as well as a derivation of  $T^*$ , see the Appendix. ■

Notice that the assumption of  $\min_{\theta_i, \delta_i} \frac{\partial}{\partial \theta_i} \mathbb{E}_{\theta_{-i}, \delta_{-i}} [v_i(k^*(\theta, \delta))] > 0$  in Theorem 2 is fairly weak; as implied by condition (54) in the proof of Lemma 1, any Bayesian implementable allocation rule  $k^*$  necessarily satisfies  $\frac{\partial}{\partial \theta_i} \mathbb{E}_{\theta_{-i}, \delta_{-i}} [v_i(k^*(\theta, \delta))] \geq 0$ .

The next and final result of this section sheds light on the critical role of information about agents' externality assessments.

<sup>30</sup>The latter maximum value exists as  $v_i$  and  $k^*$  are twice continuously differentiable and  $K$  is compact.

**Theorem 3** *Suppose externality types are common knowledge. Then any differentiable allocation rule  $k^* : \Theta \times \Delta \rightarrow K$  satisfying  $\frac{\partial}{\partial \theta_i} \mathbb{E}_{\theta_{-i}} [v_i(k^*(\theta, \delta))] \geq 0$  for all  $(\theta_i, \delta) \in \Theta_i \times \Delta$  and all  $i \in \{1, 2\}$  can be strongly Bayesian implemented with an ex post budget-balanced mechanism.*

**Proof.** Let  $k^* : \Theta \times \Delta \rightarrow K$  be an allocation rule satisfying  $\frac{\partial}{\partial \theta_i} \mathbb{E}_{\theta_{-i}} [v_i(k^*(\theta, \delta))] \geq 0$  for all  $(\theta_i, \delta)$  and all  $i$ . For agents  $i \in \{1, 2\}$  of *commonly known* externality types  $\delta = (\delta_1, \delta_2) \in \Delta$ , define the function  $S_i : \Theta \times \Delta \rightarrow \mathbb{R}$  by

$$(28) \quad S_i(\hat{\theta}, \delta) = \int_{\theta_i^{\min}}^{\hat{\theta}_i} v_i(k^*(s, \hat{\theta}_{-i}, \delta)) ds - \pi_i(k^*(\hat{\theta}, \delta) | \hat{\theta}_i) - \delta_i \pi_{-i}(k^*(\hat{\theta}, \delta) | \hat{\theta}_{-i}).$$

Then the budget-balanced transfer scheme  $T^* = (t_1^*, t_2^*) : \Theta \times \Delta \rightarrow \mathbb{R}^2$  defined by

$$(29) \quad t_1^*(\hat{\theta}, \delta) = \frac{1}{1 - \delta_1} \left[ S_1(\hat{\theta}, \delta) - \mathbb{E}_{\theta_1} [S_1(\theta_1, \hat{\theta}_2, \delta)] \right] \\ + \frac{1}{1 - \delta_2} \left[ -S_2(\hat{\theta}, \delta) + \mathbb{E}_{\theta_2} [S_2(\hat{\theta}_1, \theta_2, \delta)] \right],$$

$$(30) \quad t_2^*(\hat{\theta}, \delta) = \frac{1}{1 - \delta_1} \left[ -S_1(\hat{\theta}, \delta) + \mathbb{E}_{\theta_1} [S_1(\theta_1, \hat{\theta}_2, \delta)] \right] \\ + \frac{1}{1 - \delta_2} \left[ S_2(\hat{\theta}, \delta) - \mathbb{E}_{\theta_2} [S_2(\hat{\theta}_1, \theta_2, \delta)] \right]$$

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

As implied by condition (54) in the proof of Lemma 1, the sufficient condition of Theorem 3 is also necessary.

By Theorem 3, it is not externality assessments per se that constrain the implementability of allocation rules, but rather the asymmetry of information about them. However, in light of the above quoted ‘Wilson doctrine’, Theorem 3 is merely of theoretical relevance.

For the sake of completeness, I should briefly comment on the feasibility of budget-balanced *ex post* implementation. [Jehiel et al. \(2006\)](#) have shown that only constant allocation rules are ex post implementable, irrespective of budget balance. Their model framework entails the one of the present study, the only exception being that, in their model, agents do not internalize the distributive effects of transfers. By the following argument, their result applies nevertheless to the model framework of Section 3: Consider a social choice rule  $(k^*, t_1, t_2)$ , consisting of an allocation rule  $k^*$  and transfers  $(t_1, t_2)$  in the manner of Section 3. Expand their model framework by allowing for monetary transfers  $(t'_1, t'_2)$  the distributive effects of which are not internalized by agents. By [Jehiel et al. \(2006\)](#), the social choice rule  $(k^*, t_1, t_2)$  can be ex post implemented with some

transfer scheme  $(t'_1, t'_2)$  only if  $(k^*, t_1, t_2)$  is constant, which requires  $k^*$  to be constant. This implication holds in particular for the case of  $(t'_1, t'_2) = (0, 0)$ . As a sensitive allocation rule  $k^*$  is non-constant by definition, the (unfortunate) conclusion is that there exists no sensitive allocation rule that could be ex post implemented. Hence, while budget-balanced Bayesian implementation of sensitive allocation rules can (and must) be externality-robust, the assumption that payoff-type distributions are common knowledge remains critical.

## 5 Bargaining with Side-Payments

This section applies the results obtained above 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 process, with ‘ends’ to those allocations that are ‘feasible’ under that process. In particular, how is the feasibility of means and ends restricted when assuming that agents are privately informed about how they value shares of ‘pie’ and how they assess the externalities, tangible or intangible, that their opponent’s share might impose on them?

The bargaining literature can be broadly separated into two strands, one focusing on means, the other one on ends. The ‘means’-strand, starting with [Rubinstein \(1982\)](#), starts out from bargaining rules and takes ends as equilibrium outcomes of the respective non-cooperative game.<sup>31</sup> The ‘ends’-strand, starting earlier with [Nash \(1950\)](#), is often referred to as ‘axiomatic bargaining’ and asks for reasonable, axiomatized properties that an allocation rule, the *bargaining solution*, should possess.<sup>32</sup> Naturally, these properties are preference-contingent, which makes preference revelation a critical issue. Of course, these strands of literature have not been disjoint. For instance, [Myerson \(1979\)](#) has shown that, in a general setting which comprises private information about externality assessments, there exists a unique bargaining solution that is Bayesian incentive-compatible: it maximizes the *generalized Nash product* of [Harsanyi and Selten \(1972\)](#).<sup>33</sup>

---

<sup>31</sup>See [Ausubel, Cramton, and Deneckere \(2002\)](#) for a survey on non-cooperative bargaining under incomplete information.

<sup>32</sup>See [Thomson \(1994\)](#) for a survey.

<sup>33</sup>[Harsanyi and Selten \(1972\)](#) propose maximization of the *generalized Nash product* as an axiomatic solution to bargaining under incomplete information. Notice that the *generalized Nash product* takes type distributions explicitly into account. The welfare judgment it entails thus depends on what bargainers’ types *could have been* and not merely on what agents’ types *are ex post*. Consequently, the result of [Myerson \(1979\)](#) hinges on the assumption that type distributions are common knowledge—an assumption in conflict with the ‘Wilson doctrine’. As will be shown below, bargaining procedures can at least be externality-robust if one allows for side payments.

In the following, I discuss strategic bargaining from a mechanism design perspective. I ask which bargaining solutions are strongly Bayesian incentive-compatible if utility is transferable by means of side payments between agents.<sup>34</sup>

Consider two agents, 1 and 2, who bargain over the division of a ‘pie’ of size 1. Modify the model framework of Section 3 by assuming that, for all  $k \in [0, 1]$ ,  $v_1(k) = v(k)$  and  $v_2(k) = v(1 - k)$ , where  $v : [0, 1] \rightarrow [0, 1]$  is twice continuously differentiable and satisfies  $v(0) = 0$ ,  $v(1) = 1$ ,  $v' > 0$ , and  $v'' < 0$ . Let  $h_1(k) = h_2(k) = 0$  for all  $k$ . From their shares  $k$  and  $1 - k$ , respectively, and transfers  $t_1$  and  $t_2$ , agents 1 and 2 draw ex post utilities

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

By Theorem 1, the only *sensitive* sharing rule, or bargaining solution, that can be *strongly* Bayesian implemented through budget-balanced transfers is the one associated with externality-ignoring utilitarianism:  $k^*(\theta) = \arg \max_{k \in [0, 1]} \theta_1 v(k) + \theta_2 v(1 - k)$ . The respective transfer scheme is 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

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

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 left strategically inoperative. That is, negotiations must focus on private payoffs, irrespective of externalities. When letting  $s = 0$ , as bargainers’ externality assessments might not be common knowledge, the transfer scheme indicates that agents make mutual concessions which amount to the expected externalities they impose on one another under the sharing rule  $k^*$ .

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 likes or dislikes whom how much (and is thus more or less altruistic or spiteful).

The results of Sections 3 and 4 preclude the most prominent solutions to axiomatic bargaining from being *strongly* Bayesian implemented without incentive costs; if at all, they are Bayesian implementable through budget-balanced transfers only for very specific type distributions.

**Proposition 5** *The bargaining solutions of Nash (1950), Kalai (1977), and Kalai and Smorodinsky (1975), all of these either externality-sensitive or externality-ignoring, can-*

---

<sup>34</sup>The results are also informative for “pure” bargaining (i.e., if utility is not transferable), since side payments can be zero if the bargaining solution is incentive-compatible on its own.

not be strongly Bayesian implemented through budget-balanced transfers. The opposite would hold if externality assessments were common knowledge.<sup>35</sup>

**Proof.** Notice first that condition (7) of Lemma 1 implies that a partially differentiable bargaining solution  $k^* : \Theta \times \Delta \rightarrow [0, 1]$  which *does not* maximize the sum of private payoffs is strongly Bayesian implementable in a budget-balanced way *only if* the following holds for all  $(\theta, \delta)$ :<sup>36</sup>

$$(31) \quad \text{sgn} \left( \frac{\partial k^*}{\partial \theta_1} \frac{\partial k^*}{\partial \theta_2} \right) = - \text{sgn} \left( \frac{\partial k^*}{\partial \delta_1} \frac{\partial k^*}{\partial \delta_2} \right).$$

The externality-sensitive *Nash solution* is given by

$$(32) \quad k^*(\theta, \delta) = \arg \max_{k \in [0, 1]} [\theta_1 v(k) + \delta_1 \theta_2 v(1 - k)] \cdot [\theta_2 v(1 - k) + \delta_2 \theta_1 v(k)].$$

The externality-sensitive *Kalai solution*, in the manner of Rawls (1971), requires to maximize the minimum of agents' ex post utilities. Consider its externality-sensitive egalitarian version:  $k^* = k^*(\theta, \delta)$  such that  $u_1(k^*) = u_2(k^*)$ . This is equivalent to  $k^*$  satisfying<sup>37</sup>

$$(33) \quad 0 = (\theta_1 - \delta_1 \theta_2) v(1 - k^*) - (\theta_2 - \delta_2 \theta_1) v(k^*) = F(k^*, \theta, \delta).$$

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:  $\frac{u_1(k^*)}{u_2(k^*)} = \frac{u_1(1)}{u_2(0)}$ . This is equivalent to  $k^*$  satisfying<sup>38</sup>

$$(34) \quad 0 = \theta_2(\theta_1 - \delta_1 \theta_2) v(1 - k^*) - \theta_1(\theta_2 - \delta_2 \theta_1) v(k^*) = G(k^*, \theta, \delta).$$

The respective externality-ignoring versions of (32) to (34) are obtained when letting  $\delta_1 = \delta_2 = 0$  in each of them. These externality-ignoring bargaining solutions violate condition (31), since then  $\frac{\partial k^*}{\partial \delta_i} = 0$ , whereas  $\frac{\partial k^*}{\partial \theta_i} \neq 0$ .

---

<sup>35</sup>The Proposition presumes  $\Delta_i \subset [0, 1]$  for Nash and  $\Delta_i \subset [-1, \frac{\theta_i^{\min}}{\theta_i^{\max}}]$  for Kalai; in case of Kalai-Smorodinsky, the first part of the Proposition presumes  $\Delta_i \subset [-1, \frac{\theta_i^{\min}}{\theta_i^{\max}}]$ , whereas the second part presumes  $\Delta_i \subset [0, \frac{\theta_i^{\min}}{2\theta_i^{\max}}]$ .

<sup>36</sup>Specifically, multiplying (7) for  $i = 1$  with (7) for  $i = 2$  and then applying the sign function to both sides of the resulting identity yields the condition:  $\text{sgn} \left( \frac{\partial k^*}{\partial \theta_1} \frac{\partial k^*}{\partial \theta_2} \right) = \text{sgn} \left( \frac{\partial v_1(k^*)}{\partial \delta_1} \frac{\partial v_2(k^*)}{\partial \delta_2} \right)$ . In the present context, this condition is equivalent to (31).

<sup>37</sup>Condition (33) is well-defined on  $\Theta \times \Delta$  if and only if  $\delta_i^{\max} < \frac{\theta_i^{\min}}{\theta_i^{\max}}$  for all  $i$ : Equation (33) has a solution  $k^*$  if and only if either  $(\theta_1 - \delta_1 \theta_2), (\theta_2 - \delta_2 \theta_1) > 0$ , or  $(\theta_1 - \delta_1 \theta_2), (\theta_2 - \delta_2 \theta_1) < 0$ ; however, the latter condition would imply that  $(1 - \delta_1 \delta_2) \theta_i < 0$ , thereby contradicting the assumptions on  $\Theta \times \Delta$ .

<sup>38</sup>As before, condition (34) is well-defined on  $\Theta \times \Delta$  if and only if  $\delta_i^{\max} < \frac{\theta_i^{\min}}{\theta_i^{\max}}$  for all  $i$ .



Assuming non-negative externality types, the Nash solution is not strongly Bayesian implementable through budget-balanced transfers, due to Proposition 1(iv) and Theorem 1, whereas the second part of the Proposition is implied by Theorem 3.

It will be made clear in the Appendix that the bargaining solutions (33) and (34) both satisfy the following conditions:  $\text{sgn}(\frac{\partial k^*}{\partial \delta_1} \frac{\partial k^*}{\partial \delta_2}) = -1 = \text{sgn}(\frac{\partial k^*}{\partial \theta_1} \frac{\partial k^*}{\partial \theta_2})$  for all  $(\theta, \delta)$  with  $\delta_i < \frac{\theta_i^{\min}}{\theta_i^{\max}}$ , which violates condition (31). Furthermore,  $\frac{\partial k^*}{\partial \theta_2} < 0 < \frac{\partial k^*}{\partial \theta_1}$  if  $\Delta_i \subset [-1, \frac{\theta_i^{\min}}{\theta_i^{\max}}]$  in case of (33) and  $\Delta_i \subset [0, \frac{\theta_i^{\min}}{2\theta_i^{\max}}]$  in case of (34). The latter condition implies in particular that  $\frac{\partial v_i(k^*)}{\partial \theta_i} > 0$  for all  $i$ , such that  $\frac{\partial}{\partial \theta_i} \mathbb{E}_{\theta_{-i}} [v_i(k^*(\theta, \delta))] > 0$ ; hence, Theorem 3 gives proof of the second part of the Proposition. ■

## 6 Holistic Social Welfare Measures

Up to this point, I have restricted attention to the social welfare judgment inherent to the allocation rule. To which extent does the result of Theorem 1 expand to social welfare judgments that are *holistic* in the sense that they incorporate the distributive effects of a transfer scheme? With Theorem 1 at hand, it is easy to answer this question.

Consider a differentiable function  $W : \mathbb{R}^4 \rightarrow \mathbb{R}$  and define  $V : K \times \mathbb{R}^2 \rightarrow \mathbb{R}$  by

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

where  $\pi_i(k, t_i | \theta_i) = \theta_i v_i(k) + h_i(k) + t_i$ . Suppose  $W$  is an *ex post social welfare measure* in that it is invariant to changes in type distributions. Assume also that  $W$  satisfies

$$(35) \quad \frac{\partial W}{\partial \pi_1} = [W_1 + \delta_2 W_4] > 0 \quad \text{and} \quad \frac{\partial W}{\partial \pi_2} = [W_3 + \delta_1 W_2] > 0.$$

The social choice rule  $(k^*, t_1^*, t_2^*)$ , with allocation rule  $k^* : \Theta \times \Delta \rightarrow K$  and transfer scheme  $(t_1^*, t_2^*) : \Theta \times \Delta \rightarrow \mathbb{R}^2$ , is budget-balanced and maximizes  $V$  if and only if  $t_2^* = -t_1^*$  and

$$(36) \quad (k^*, t_1^*) = \arg \max_{(k, t_1) \in K \times \mathbb{R}} V(k, t_1, -t_1).$$

Assuming  $W$  allows for interior solutions,  $(k^*, t_1^*)$  satisfies the first-order conditions

$$(37) \quad 0 = \frac{\partial V(k^*, t_1^*, -t_1^*)}{\partial k} = [W_1 + \delta_2 W_4] \frac{d\pi_1(k^* | \theta_1)}{dk} + [W_3 + \delta_1 W_2] \frac{d\pi_2(k^* | \theta_2)}{dk},$$

$$(38) \quad 0 = \frac{\partial V(k^*, t_1^*, -t_1^*)}{\partial t_1} = [W_1 + \delta_2 W_4] - [W_3 + \delta_1 W_2],$$

where  $\pi_i(k^* | \theta_i) = \theta_i v_i(k) + h_i(k)$ . Conditions (35), (37), and (38) jointly imply that the socially efficient allocation rule is necessarily 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 social welfare judgments yield redistributive tax tariffs that are Bayesian incentive-compatible?

By Theorem 1, the optimal allocation rule  $k^*$  can be Bayesian implemented in a budget-balanced way if and only if transfers are of AGV-type. As AGV-type transfers vary with changes in the distribution of types, 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,  $(k^*, t^*, -t^*)$ , that is interior solution to the maximization of a differentiable ex post social welfare measure  $W$  satisfying condition (35) is Bayesian incentive-compatible if and only if  $W$  is consistent with externality-ignoring utilitarianism. The respective mechanism is of AGV-type.*

Theorem 4 applies in particular to the welfare measures listed in Proposition 1.

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

**Proposition 6** *A budget-balanced social choice rule,  $(k^*, t^*, -t^*)$ , satisfying Rawls' maximin principle, inclusive or exclusive of externalities, is not Bayesian incentive-compatible.*

**Proof.** Consider the maximin principle inclusive of externalities and let

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

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

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

where  $\pi_i(k^* | \theta_i) = \theta_i v_i(k) + h_i(k)$ . 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

$$u_1 = u_2 = \frac{1 - \delta_1 \delta_2}{2 - \delta_1 - \delta_2} [\pi_1(k^* | \theta_1) + \pi_2(k^* | \theta_2)].$$

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 so as to Bayesian implement  $k^*$ . As  $t^*$  is not of AGV-type,  $(k^*, t^*, -t^*)$  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. ■

## 7 Conclusion

How agents assess the (in-)tangible externalities that others might impose on them can strongly influence strategic interaction. I have explored ex post Pareto-efficient Bayesian implementation for agents whose externality assessments and private payoffs, exclusive of externalities, are all subject to asymmetric information. Under reasonable assumptions, ex post Pareto-efficient allocations are Bayesian implementable with a budget-balanced mechanism if and only if the social welfare judgment underlying the choice of allocations is that of externality-ignoring utilitarianism. This restriction is caused by the asymmetry of information about agents' externality assessments, as common knowledge of externality assessments allows for budget-balanced Bayesian implementation of (nearly) any allocation rule.

The ex post Pareto-efficient, budget-balanced mechanism corresponding to externality-ignoring utilitarianism necessarily takes the form of the renowned 'expected externality mechanism' due to [Arrow \(1979\)](#) and [d'Aspremont and Gérard-Varet \(1979\)](#). This mechanism is externality-robust in that it leaves externality assessments strategically inoperative. Externality robustness turns out to be not just a desirable property in order to avoid unrealistic common knowledge assumptions, as urged by [Wilson \(1987\)](#): externality robustness is rather necessary from an incentive compatibility point of view. Otherwise, agents would internalize the distributive effects of the mechanism itself, and counterbalancing the associated adverse incentives would come at costs, embodied in the violation of budget balance. As they result from the welfare judgment inherent to an allocation rule, I have called these the incentive costs of welfare judgments.

In the range of conflict resolution, the central result provides a rationale for the common-sense approach most people would adopt when arbitrating between conflicting parties: namely, to *not* condition the arbitration process or the final resolution on the extent to which the opponents despise each other, but to rather "focus on the issue" and to base arbitration merely on how it would affect the opponents' *material* wealth: 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. The central result implies in particular that, even when allowing for side payments, the most prominent bargaining solutions, namely those of [Nash \(1950\)](#), [Kalai \(1977\)](#), and [Kalai and Smorodinsky \(1975\)](#), are not Bayesian incentive-compatible.

From a more general perspective, 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.

# Appendix

## Proof of Proposition 1

Suppose in the following that  $h_i \equiv 0$  and  $\Delta_i \subset [0, 1)$  for all  $i$ . Hence,  $\pi_i = \theta_i v_i$ . Obviously, the social welfare measures (iv) and (v) satisfy the Pareto-property as well as condition (i) of Definition 1. In the following, it is shown that they also satisfy the identities (3) and (4). For this purpose, ease notation by letting  $\pi_i = \pi_i(k^* | \theta_i)$  and  $v_i = v_i(k^*)$ .

### Proof of Proposition 1(iv)

Let  $V(k) = [\pi_1(k | \theta_1) + \delta_1 \pi_2(k | \theta_2)] \cdot [\pi_2(k | \theta_2) + \delta_2 \pi_1(k | \theta_1)]$ . By assumption,  $k^* : \Theta \times \Delta \rightarrow K$  satisfies the FOC

$$(39) \quad 0 = \frac{dV(k^*)}{dk} = \left( \frac{d\pi_1}{dk} + \delta_1 \frac{d\pi_2}{dk} \right) (\pi_2 + \delta_2 \pi_1) + \left( \frac{d\pi_2}{dk} + \delta_2 \frac{d\pi_1}{dk} \right) (\pi_1 + \delta_1 \pi_2).$$

Define  $x_1 = \pi_1 + \delta_1 \pi_2$  and  $x_2 = \pi_2 + \delta_2 \pi_1$ . Notice that  $x_1, x_2 > 0$ . Then (39) can be rewritten so as to obtain

$$(40) \quad 0 = (x_1 + \delta_1 x_2) \frac{d\pi_2}{dk} + (x_2 + \delta_2 x_1) \frac{d\pi_1}{dk},$$

where  $x_1 + \delta_1 x_2 > 0$  and  $x_2 + \delta_2 x_1 > 0$ , since  $\delta_1, \delta_2 \geq 0$ . Implicit differentiation of (39) with respect to  $\theta_1$  yields  $\partial k^* / \partial \theta_1 = -X_1 / [d^2 V(k^*) / dk^2]$ , 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  $d^2 V(k^*) / dk^2 < 0$  by the SOC,  $\text{sgn}(\partial k^* / \partial \theta_1) = \text{sgn}(X_1)$ . Having assumed that  $h_i \equiv 0$ , one can make use of the identities  $v_1 \frac{d\pi_1}{dk} = \pi_1 \frac{dv_1}{dk}$  and (40) 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  $\delta_1, \delta_2 \in [0, 1)$  and  $\pi_i, x_i > 0$ , letting  $\delta_1 = \delta_2 = 0$  yields the lower bound  $Y_1 > x_1 x_2 + \pi_1(-x_2) = (x_1 - \pi_1)x_2 = \delta_1 \pi_2 x_2 \geq 0$ . Hence,  $\text{sgn}(\partial k^* / \partial \theta_1) = \text{sgn}(X_1) = \text{sgn}(dv_1 / dk)$ , while, by assumption,  $\text{sgn}(\partial k^* / \partial \theta_i) \neq 0$  and  $\text{sgn}(dv_i / dk) \neq 0$  for all  $i$ . Hence,  $1 = \text{sgn}^2(\partial k^* / \partial \theta_1) =$

$\text{sgn}(\partial v_1/\partial k) \text{sgn}(\partial k^*/\partial \theta_i) = \text{sgn}(\partial v_1/\partial \theta_1)$ . By symmetry,  $1 = \text{sgn}(\partial v_2/\partial \theta_2)$ . Hence,  $1 = \text{sgn}(\partial v_1/\partial \theta_1 \cdot \partial v_2/\partial \theta_2) = \text{sgn}(\partial v_1/\partial \theta_2) \text{sgn}(\partial v_2/\partial \theta_1)$ , as required.

On the other hand, implicit differentiation of the FOC (39) with respect to  $\delta_1$  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 SOC,  $\text{sgn}(\partial k^*/\partial \delta_1) = \text{sgn}(Z_1)$ . By making use of (40),  $Z_1$  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,  $\text{sgn}(\partial k^*/\partial \delta_1) = \text{sgn}(Z_1) = \text{sgn}(\pi_2/dk)$ , such that

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

By symmetry,  $\text{sgn}(\partial \pi_1/\partial \delta_2) = \text{sgn}^2(d\pi_1/dk) \in \{0, 1\}$ . As  $x_1 + \delta_1 x_2 > 0$  and  $x_2 + \delta_2 x_1 > 0$ , the FOC (39) implies that  $d\pi_1/dk = 0$  if and only if  $d\pi_2/dk = 0$ . Hence, as required,  $\text{sgn}(\partial \pi_1/\partial \delta_2) = \text{sgn}(\partial \pi_2/\partial \delta_1)$ . Altogether,  $W$  is sensitive.  $\blacksquare$

## Proof of Proposition 1(v)

Let  $V(k) = [[\pi_1(k | \theta_1) + \delta_1 \pi_2(k | \theta_2)]^{-\rho} + [\pi_2(k | \theta_2) + \delta_2 \pi_1(k | \theta_1)]^{-\rho}]^{-\frac{1}{\rho}}$ , with  $\rho \in (-1, \infty) \setminus \{0\}$ . By assumption,  $k^* : \Theta \times \Delta \rightarrow K$  satisfies the FOC

$$(41) \quad 0 = \frac{dV(k^*)}{dk} = [V(k^*)]^{1+\rho} \left[ (\pi_1 + \delta_1 \pi_2)^{-\rho-1} \left( \frac{d\pi_1}{dk} + \delta_1 \frac{d\pi_2}{dk} \right) + (\pi_2 + \delta_2 \pi_1)^{-\rho-1} \left( \frac{d\pi_2}{dk} + \delta_2 \frac{d\pi_1}{dk} \right) \right].$$

Define  $x_1 = \pi_1 + \delta_1 \pi_2$  and  $x_2 = \pi_2 + \delta_2 \pi_1$ . Notice that  $x_1, x_2 > 0$ . By (41),

$$(42) \quad 0 = (x_1^{-\rho-1} + \delta_2 x_2^{-\rho-1}) \frac{d\pi_1}{dk} + (x_2^{-\rho-1} + \delta_1 x_1^{-\rho-1}) \frac{d\pi_2}{dk},$$

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 (41) with respect to  $\theta_1$  yields  $\partial k^*/\partial \theta_1 = -X_1[V(k^*)]^{1+\rho}/[d^2V(k^*)/dk^2]$ , where

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

Since  $d^2V(k^*)/dk^2 < 0$  by the SOC,  $\text{sgn}(\partial k^*/\partial \theta_1) = \text{sgn}(X_1)$ . Having assumed that  $h_i \equiv 0$ , one can make use of the identities  $v_1 \frac{d\pi_1}{dk} = \pi_1 \frac{dv_1}{dk}$  and (42) to rewrite  $X_1$  as

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

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

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

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  $Y_1 = Y_2$ .<sup>39</sup> Hence, as required,

$$\begin{aligned} (43) \quad 1 &= \text{sgn}^2 \left( \frac{\partial k^*}{\partial \theta_1} \right) \text{sgn}^2 \left( \frac{\partial k^*}{\partial \theta_2} \right) \\ &= \text{sgn} \left( \frac{\partial k^*}{\partial \theta_1} \right) \text{sgn} \left( \frac{dv_1}{dk} \right) \text{sgn}(Y_1) \text{sgn} \left( \frac{\partial k^*}{\partial \theta_2} \right) \text{sgn} \left( \frac{dv_2}{dk} \right) \text{sgn}(Y_2) \\ &= \text{sgn} \left( \frac{\partial v_1}{\partial \theta_2} \right) \text{sgn} \left( \frac{\partial v_2}{\partial \theta_1} \right), \end{aligned}$$

where the first equality of (43) holds due to the assumption that  $\partial k^*/\partial \theta_i \neq 0$  for all  $i$ . On the other hand, implicit differentiation of the FOC (41) with respect to  $\delta_1$  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 SOC,  $\text{sgn}(\partial k^*/\partial \delta_1) = \text{sgn}(Z_1)$ . By making use of (42),  $Z_1$  can be written as

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

<sup>39</sup>For  $\delta_1 = \delta_2 = 0$ , one observes that  $Y_i = -\rho(\pi_1 \pi_2)^{-\rho-1}$ . Hence,  $\text{sgn}(\partial k^*/\partial \theta_i) = -\text{sgn}(dv_i/dk)$  for  $\rho > 0$ . For this reason, I let Definition 1 require the weaker property of  $\text{sgn}(\partial v_1/\partial \theta_2) \text{sgn}(\partial v_2/\partial \theta_1) = 1$ , instead of  $\text{sgn}(\partial k^*/\partial \theta_i) \text{sgn}(dv_i/dk) = 1$  for all  $i$ .

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

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

By symmetry,  $\text{sgn}(\partial \pi_1/\partial \delta_2) = \text{sgn}^2(d\pi_1/dk) \in \{0, 1\}$ . 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 (42) implies that  $d\pi_1/dk = 0$  if and only if  $d\pi_2/dk = 0$ . Hence, as required,  $\text{sgn}(\partial \pi_1/\partial \delta_2) = \text{sgn}(\partial \pi_2/\partial \delta_1)$ . Altogether,  $W$  is sensitive.  $\blacksquare$

## Proof of Lemma 1

Suppose the sensitive allocation rule  $k^* : \Theta \times \Delta \rightarrow \mathbb{R}$  is strongly Bayesian implemented by the ex post budget-balanced transfer scheme  $T = (t_1, t_2) : \Theta \times \Delta \rightarrow \mathbb{R}^2$ . Define

$$(44) \quad \bar{v}_i(\hat{\theta}_i, \hat{\delta}_i) = \mathbb{E}_{\theta_{-i}, \delta_{-i}} [v_i(k^*(\hat{\theta}_i, \hat{\delta}_i, \theta_{-i}, \delta_{-i}))],$$

$$(45) \quad \bar{h}_i(\hat{\theta}_i, \hat{\delta}_i) = \mathbb{E}_{\theta_{-i}, \delta_{-i}} [h_i(k^*(\hat{\theta}_i, \hat{\delta}_i, \theta_{-i}, \delta_{-i}))],$$

$$(46) \quad \bar{\pi}_{-i}(\hat{\theta}_i, \hat{\delta}_i) = \mathbb{E}_{\theta_{-i}, \delta_{-i}} [\pi_{-i}(k^*(\hat{\theta}_i, \hat{\delta}_i, \theta_{-i}, \delta_{-i}) | \theta_{-i})],$$

$$(47) \quad \bar{t}_i(\hat{\theta}_i, \hat{\delta}_i) = \mathbb{E}_{\theta_{-i}, \delta_{-i}} [t_i(\hat{\theta}_i, \hat{\delta}_i, \theta_{-i}, \delta_{-i})],$$

$$(48) \quad \bar{t}_{-i}(\hat{\theta}_i, \hat{\delta}_i) = \mathbb{E}_{\theta_{-i}, \delta_{-i}} [t_{-i}(\hat{\theta}_i, \hat{\delta}_i, \theta_{-i}, \delta_{-i})],$$

where  $\pi_i(k | \theta_i) = \theta_i v_i(k) + h_i(k)$ . For  $i \in \{1, 2\}$ , 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  $(\theta_i, \delta_i)$  and if agent  $-i$  reports her type truthfully:

$$(49) \quad U_i(\hat{\theta}_i, \hat{\delta}_i | \theta_i, \delta_i) = \theta_i \bar{v}_i(\hat{\theta}_i, \hat{\delta}_i) + \bar{h}_i(\hat{\theta}_i, \hat{\delta}_i) + \bar{t}_i(\hat{\theta}_i, \hat{\delta}_i) + \delta_i \bar{\pi}_{-i}(\hat{\theta}_i, \hat{\delta}_i) + \delta_i \bar{t}_{-i}(\hat{\theta}_i, \hat{\delta}_i).$$

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  $\delta_i, \hat{\delta}_i \in \Delta_i$ :

$$(50) \quad \begin{aligned} U_i(\theta_i, \delta_i) &\geq U_i(\hat{\theta}_i, \delta_i | \theta_i, \delta_i) \\ &= U_i(\hat{\theta}_i, \delta_i) + (\theta_i - \hat{\theta}_i) \bar{v}_i(\hat{\theta}_i, \delta_i), \end{aligned}$$

$$(51) \quad \begin{aligned} U_i(\hat{\theta}_i, \delta_i) &\geq U_i(\theta_i, \delta_i | \hat{\theta}_i, \delta_i) \\ &= U_i(\theta_i, \delta_i) + (\hat{\theta}_i - \theta_i) \bar{v}_i(\theta_i, \delta_i), \end{aligned}$$

$$(52) \quad \begin{aligned} U_i(\theta_i, \delta_i) &\geq U_i(\theta_i, \hat{\delta}_i | \theta_i, \delta_i) \\ &= U_i(\theta_i, \hat{\delta}_i) + (\delta_i - \hat{\delta}_i) [\bar{\pi}_{-i}(\theta_i, \hat{\delta}_i) + \bar{t}_{-i}(\theta_i, \hat{\delta}_i)], \end{aligned}$$

$$(53) \quad \begin{aligned} U_i(\theta_i, \hat{\delta}_i) &\geq U_i(\theta_i, \delta_i | \theta_i, \hat{\delta}_i) \\ &= U_i(\theta_i, \delta_i) + (\hat{\delta}_i - \delta_i) [\bar{\pi}_{-i}(\theta_i, \delta_i) + \bar{t}_{-i}(\theta_i, \delta_i)]. \end{aligned}$$



Without loss of generality, suppose  $\hat{\theta}_i > \theta_i$ . Then (50) and (51) imply that

$$(54) \quad \bar{v}_i(\hat{\theta}_i, \delta_i) \geq \frac{U_i(\hat{\theta}_i, \delta_i) - U_i(\theta_i, \delta_i)}{\hat{\theta}_i - \theta_i} \geq \bar{v}_i(\theta_i, \delta_i).$$

As  $\bar{v}_i$  is continuous on  $\Theta_i$ , letting  $\hat{\theta}_i \downarrow \theta_i$  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  $\theta_i$  yields the identity

$$(55) \quad U_i(\theta_i, \delta_i) = p_i(\delta_i) + \int_{\theta_i^{\min}}^{\theta_i} \bar{v}_i(s, \delta_i) ds,$$

with some function  $p_i : \Delta_i \rightarrow \mathbb{R}$ . Similarly, suppose  $\hat{\delta}_i > \delta_i$ . Then (52) and (53) imply that

$$(56) \quad \bar{\pi}_{-i}(\theta_i, \hat{\delta}_i) + \bar{t}_{-i}(\theta_i, \hat{\delta}_i) \geq \frac{U_i(\theta_i, \hat{\delta}_i) - U_i(\theta_i, \delta_i)}{\hat{\delta}_i - \delta_i} \geq \bar{\pi}_{-i}(\theta_i, \delta_i) + \bar{t}_{-i}(\theta_i, \delta_i).$$

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

$$(57) \quad \frac{\partial U_i(\theta_i, \delta_i)}{\partial \delta_i} = \bar{\pi}_{-i}(\theta_i, \delta_i) + \bar{t}_{-i}(\theta_i, \delta_i).$$

Integrating with respect to  $\delta_i$  in (57) yields the identity

$$(58) \quad U_i(\theta_i, \delta_i) = q_i(\theta_i) + \int_{\delta_i^{\min}}^{\delta_i} \bar{\pi}_{-i}(\theta_i, r) dr + \int_{\delta_i^{\min}}^{\delta_i} \bar{t}_{-i}(\theta_i, r) dr,$$

with some function  $q_i : \Theta_i \rightarrow \mathbb{R}$ . Identity (58) and the assumptions on the functions  $v_i$  imply that  $U_i(\theta_i, \delta_i)$  and, thus,  $p_i$  from (55) must be differentiable. Jointly, identities (55) and (58) imply that

$$(59) \quad \int_{\delta_i^{\min}}^{\delta_i} \bar{t}_{-i}(\theta_i, r) dr = p_i(\delta_i) - q_i(\theta_i) + \int_{\theta_i^{\min}}^{\theta_i} \bar{v}_i(s, \delta_i) ds - \int_{\delta_i^{\min}}^{\delta_i} \bar{\pi}_{-i}(\theta_i, r) dr.$$

Differentiating (59) with respect to  $\delta_i$  yields

$$(60) \quad \bar{t}_{-i}(\theta_i, \delta_i) = \frac{\partial p_i(\delta_i)}{\partial \delta_i} - \bar{\pi}_{-i}(\theta_i, \delta_i) + \frac{\partial}{\partial \delta_i} \int_{\theta_i^{\min}}^{\theta_i} \bar{v}_i(s, \delta_i) ds.$$

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

(61)

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

where the second equality is implied by identity (60), and where the Leibniz integral rule has been used to obtain the second and the last equality.

In order to be Bayesian implementable with a budget-balanced mechanism,  $k^*$  must satisfy identity (61) irrespective of the specific form that the transfer scheme might take. As  $k^*$  is also assumed to be *strongly* Bayesian implementable (in the manner of Definition 2), identity (61) must hold for *any* set of (non-degenerate) type distributions  $\{F_{-i}, G_{-i}\}$ . However, due to the assumptions on the functions  $v_i$ ,  $h_i$ , and  $k^*$ , the argument of  $\mathbb{E}_{\theta_{-i}, \delta_{-i}}[\cdot]$  in (61) 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  $(\theta, \delta) \in \Theta \times \Delta$ . This proves the first part of Lemma 1. For the second part, reconsider identities (55) and (60). Under truthful revelation, they jointly imply that

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

Now suppose  $k^*$  is independent from externality types:  $k^* = k^*|_{\Theta}$ . According to identities (62) and (60), respectively,  $\bar{t}_i(\theta_i, \delta_i)$  and  $\bar{t}_{-i}(\theta_i, \delta_i)$  then satisfy

$$(63) \quad \bar{t}_i(\theta_i, \delta_i) = p_i(\delta_i) - \delta_i \frac{\partial p_i(\delta_i)}{\partial \delta_i} - \theta_i \bar{v}_i(\theta_i, \delta_i) - \bar{h}_i(\theta_i, \delta_i) + \int_{\theta_i^{\min}}^{\theta_i} \bar{v}_i(s, \delta_i) ds,$$

$$(64) \quad \bar{t}_{-i}(\theta_i, \delta_i) = \frac{\partial p_i(\delta_i)}{\partial \delta_i} - \bar{\pi}_{-i}(\theta_i, \delta_i),$$

where, due to  $k^* = k^*|_{\Theta}$ , only the terms containing  $p_i$  effectively depend on  $\delta_i$ . Due to budget balance, identities (63) and (64) imply that  $p_i$  solves the differential equation

$$(65) \quad a_i = p_i(\delta_i) + (1 - \delta_i) \frac{\partial p_i(\delta_i)}{\partial \delta_i},$$

where  $a_i$  is some constant. Differentiating (65) with respect to  $\delta_i$  yields  $\frac{\partial^2 p_i(\delta_i)}{\partial \delta_i^2} = 0$ , such that  $\frac{\partial p_i(\delta_i)}{\partial \delta_i} = -\alpha_i$  for some constant  $\alpha_i$ . Hence, identity (64) reads  $\bar{t}_{-i}(\theta_i, \delta_i) = -\alpha_i - \bar{\pi}_{-i}(\theta_i, \delta_i)$ , implying that  $\bar{t}_i(\theta_i, \delta_i) = \alpha_i + \bar{\pi}_{-i}(\theta_i, \delta_i)$  due to budget balance.  $\blacksquare$

## Proof of Theorem 2 Continued

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  $(\theta_i, \delta_i)$  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

$$(66) \quad \frac{\partial}{\partial \hat{\theta}_i} U_i(\hat{\theta}_i, \hat{\delta}_i | \theta_i, \delta_i) = (\theta_i - \hat{\theta}_i) \frac{\partial}{\partial \hat{\theta}_i} \bar{v}_i(\hat{\theta}_i, \hat{\delta}_i) + (\delta_i - \hat{\delta}_i) \frac{\partial}{\partial \hat{\delta}_i} \bar{v}_i(\hat{\theta}_i, \hat{\delta}_i),$$

$$(67) \quad \frac{\partial}{\partial \hat{\delta}_i} U_i(\hat{\theta}_i, \hat{\delta}_i | \theta_i, \delta_i) = (\theta_i - \hat{\theta}_i) \frac{\partial}{\partial \hat{\delta}_i} \bar{v}_i(\hat{\theta}_i, \hat{\delta}_i) + (\delta_i - \hat{\delta}_i) \frac{\partial^2}{\partial \hat{\delta}_i^2} \left[ p_i(\hat{\delta}_i) + \int_{\theta_i^{\min}}^{\hat{\theta}_i} \bar{v}_i(s, \hat{\delta}_i) ds \right].$$

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 the partial derivatives (66) and (67) read<sup>40</sup>

$$(68) \quad \frac{\partial}{\partial \hat{\theta}_i} U_i(\hat{\theta}_i, \hat{\delta}_i | \theta_i, \delta_i) = (\theta_i - \hat{\theta}_i) B_i + (\delta_i - \hat{\delta}_i) A_i,$$

$$(69) \quad \frac{\partial}{\partial \hat{\delta}_i} U_i(\hat{\theta}_i, \hat{\delta}_i | \theta_i, \delta_i) = (\theta_i - \hat{\theta}_i) A_i + (\delta_i - \hat{\delta}_i) C_i.$$

Suppose  $k^* \neq k^*|_{\Theta}$ . Then,  $B_i > 0$  by assumption. Choose  $p_i(\delta_i) = \frac{1}{2} c_i \delta_i^2$ , with  $c_i$  as defined in (27). Then  $C_i > 0$ , and condition (26) is satisfied:

$$(70) \quad A_i^2 < B_i C_i.$$

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

$$(71) \quad H_i = \begin{pmatrix} -B_i & -A_i \\ -A_i & -C_i \end{pmatrix}.$$

The principal minors of (71), namely  $-B_i < 0$  and  $\det(H_i) = B_i C_i - A_i^2 > 0$ , are alternating in sign, with the first-order principal minor being negative. Hence,  $(\theta_i, \delta_i)$  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 no local maximizer of  $U_i(\hat{\theta}_i, \hat{\delta}_i | \theta_i, \delta_i)$  is located on the boundary of  $\Theta_i \times \Delta_i$ .

<sup>40</sup>Suppose  $k^* = k^*|_{\Theta}$ , and assume  $B_i \geq 0$ . Then,  $A_i = 0$ . When choosing  $p_i = 0$ , then also  $C_i = 0$ . By (68) and (69), truth-telling is then a global maximizer of each agent  $i$ 's expected utility under the transfer scheme  $T^*$ , which gives proof of the second part of Theorem 2.

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 (69) 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  $\Theta_i \times \Delta_i$  qualify as potential further local maximizers.

Suppose  $(\theta_i^{\max}, \delta_i^{\max})$  is a local maximizer. Then  $0 \leq \frac{\partial}{\partial \theta_i} U_i(\theta_i^{\max}, \delta_i^{\max} | \theta_i, \delta_i)$  and  $0 \leq \frac{\partial}{\partial \delta_i} U_i(\theta_i^{\max}, \delta_i^{\max} | \theta_i, \delta_i)$  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, it also implies that  $(\delta_i - \delta_i^{\max}) \geq -(\theta_i - \theta_i^{\max}) \frac{A_i}{C_i}$  and, thus,

$$(72) \quad 0 \leq (\theta_i - \theta_i^{\max}) B_i + (\delta_i - \delta_i^{\max}) A_i \leq (\theta_i - \theta_i^{\max}) \frac{1}{C_i} (B_i C_i - A_i^2) < 0.$$

Suppose  $(\theta_i^{\max}, \delta_i^{\min})$  is a local maximizer. Then  $0 \leq \frac{\partial}{\partial \theta_i} U_i(\theta_i^{\max}, \delta_i^{\min} | \theta_i, \delta_i)$  and  $0 \geq \frac{\partial}{\partial \delta_i} U_i(\theta_i^{\max}, \delta_i^{\min} | \theta_i, \delta_i)$  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, it also implies that  $(\theta_i - \theta_i^{\max}) \geq -(\delta_i - \delta_i^{\min}) \frac{A_i}{B_i}$  and, thus,

$$(73) \quad 0 \geq (\theta_i - \theta_i^{\max}) A_i + (\delta_i - \delta_i^{\min}) C_i \geq (\delta_i - \delta_i^{\min}) \frac{1}{B_i} (B_i C_i - A_i^2) > 0.$$

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

$$(74) \quad 0 \geq (\theta_i - \theta_i^{\min}) B_i + (\delta_i - \delta_i^{\min}) A_i \geq (\theta_i - \theta_i^{\min}) \frac{1}{C_i} (B_i C_i - A_i^2) > 0.$$

Finally, suppose  $(\theta_i^{\min}, \delta_i^{\max})$  is a local maximizer. Then  $0 \geq \frac{\partial}{\partial \theta_i} U_i(\theta_i^{\min}, \delta_i^{\max} | \theta_i, \delta_i)$  and  $0 \leq \frac{\partial}{\partial \delta_i} U_i(\theta_i^{\min}, \delta_i^{\max} | \theta_i, \delta_i)$  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, it also implies that  $(\theta_i - \theta_i^{\min}) \leq -(\delta_i - \delta_i^{\max}) \frac{A_i}{B_i}$  and, thus,

$$(75) \quad 0 \leq (\theta_i - \theta_i^{\min}) A_i + (\delta_i - \delta_i^{\max}) C_i \leq (\delta_i - \delta_i^{\max}) \frac{1}{B_i} (B_i C_i - A_i^2) < 0.$$

Altogether,  $(\theta_i, \delta_i)$  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  $T^* = (t_1^*, t_2^*) : \Theta \times \Delta \rightarrow \mathbb{R}^2$  (strongly) Bayesian implements the twice continuously differentiable allocation rule  $k^* : \Theta \times \Delta \rightarrow K$ . With notation adopted from the proof of Lemma 1, condition (60) of that proof states that  $T^*$  must satisfy the identity

$$(76) \quad \bar{t}_{-i}(\theta_i, \delta_i) = \frac{\partial p_i(\delta_i)}{\partial \delta_i} - \bar{\pi}_{-i}(\theta_i, \delta_i) + \frac{\partial}{\partial \delta_i} \int_{\theta_i^{\min}}^{\theta_i} \bar{v}_i(s, \delta_i) ds,$$

where  $p_i : \Delta_i \rightarrow \mathbb{R}$  is some differentiable function. Conditions (49), (55), and (76) imply that

$$(77) \quad \begin{aligned} p_i(\delta_i) + \int_{\theta_i^{\min}}^{\theta_i} \bar{v}_i(s, \delta_i) ds &= U_i(\theta_i, \delta_i) \\ &= \bar{\pi}_i(\theta_i, \delta_i) + \bar{t}_i(\theta_i, \delta_i) + \delta_i \bar{\pi}_{-i}(\theta_i, \delta_i) + \delta_i \bar{t}_{-i}(\theta_i, \delta_i) \\ &= \bar{\pi}_i(\theta_i, \delta_i) + \bar{t}_i(\theta_i, \delta_i) + \delta_i \frac{\partial p_i(\delta_i)}{\partial \delta_i} + \delta_i \frac{\partial}{\partial \delta_i} \int_{\theta_i^{\min}}^{\theta_i} \bar{v}_i(s, \delta_i) ds. \end{aligned}$$

Hence,  $T^*$  must also satisfy the identity

$$(78) \quad \begin{aligned} \bar{t}_i(\theta_i, \delta_i) &= p_i(\delta_i) - \delta_i \frac{\partial p_i(\delta_i)}{\partial \delta_i} - \bar{\pi}_i(\theta_i, \delta_i) \\ &\quad + \int_{\theta_i^{\min}}^{\theta_i} \bar{v}_i(s, \delta_i) ds - \delta_i \frac{\partial}{\partial \delta_i} \int_{\theta_i^{\min}}^{\theta_i} \bar{v}_i(s, \delta_i) ds. \end{aligned}$$

From identities (76) and (78), the transfer scheme  $T^*$  can be “guessed”. ■

## Proof of Theorem 3 Continued

Consider the functions  $S_i$  defined by (28) and the transfer scheme  $T^*$  defined by (29) and (30). Notice first that, for all  $(\hat{\theta}_1, \theta_2) \in \Theta$  and all  $\delta \in \Delta$ ,

$$(79) \quad \begin{aligned} \mathbb{E}_{\theta_2} [t_1^*(\hat{\theta}_1, \theta_2, \delta) + \delta_1 t_2^*(\hat{\theta}_1, \theta_2, \delta)] &= \mathbb{E}_{\theta_2} [S_1(\hat{\theta}_1, \theta_2, \delta)] \\ &\quad - \mathbb{E}_{\theta_1, \theta_2} [S_1(\theta_1, \theta_2, \delta)], \end{aligned}$$

$$(80) \quad \begin{aligned} \mathbb{E}_{\theta_2} [S_1(\hat{\theta}_1, \theta_2, \delta)] &= \int_{\theta_1^{\min}}^{\hat{\theta}_1} \mathbb{E}_{\theta_2} [v_1(k^*(s, \theta_2, \delta))] ds \\ &\quad - \mathbb{E}_{\theta_2} [\pi_1(k^*(\hat{\theta}_1, \theta_2, \delta) | \hat{\theta}_1)] \\ &\quad - \delta_1 \cdot \mathbb{E}_{\theta_2} [\pi_2(k^*(\hat{\theta}_1, \theta_2, \delta) | \theta_2)], \end{aligned}$$

where Fubini's theorem has been used to obtain equation (80). Under the assumption that agent 2 reveals her payoff type truthfully, agent 1 chooses  $\hat{\theta}_1$  so as to maximize her interim expected utility. By making use of equations (79) and (80),

$$\begin{aligned}
& \mathbb{E}_{\theta_2} \left[ u_1(k^*(\hat{\theta}_1, \theta_2, \delta), t_1^*(\hat{\theta}_1, \theta_2, \delta), t_2^*(\hat{\theta}_1, \theta_2, \delta), \theta_2 \mid \theta_1) \right] \\
&= \mathbb{E}_{\theta_2} \left[ [\pi_1(k^*(\hat{\theta}_1, \theta_2, \delta) \mid \theta_1) + t_1^*(\hat{\theta}_1, \theta_2, \delta)] + \delta_1 \cdot [\pi_2(k^*(\hat{\theta}_1, \theta_2, \delta) \mid \theta_2) + t_2^*(\hat{\theta}_1, \theta_2, \delta)] \right] \\
&= \mathbb{E}_{\theta_2} [\pi_1(k^*(\hat{\theta}_1, \theta_2, \delta) \mid \theta_1)] + \delta_1 \cdot \mathbb{E}_{\theta_2} [\pi_2(k^*(\hat{\theta}_1, \theta_2, \delta) \mid \theta_2)] \\
&\quad + \mathbb{E}_{\theta_2} [S_1(\hat{\theta}_1, \theta_2, \delta)] - \mathbb{E}_{\theta_1, \theta_2} [S_1(\theta_1, \theta_2, \delta)] \\
&= \mathbb{E}_{\theta_2} [\pi_1(k^*(\hat{\theta}_1, \theta_2, \delta) \mid \theta_1)] - \mathbb{E}_{\theta_2} [\pi_1(k^*(\hat{\theta}_1, \theta_2, \delta) \mid \hat{\theta}_1)] \\
&\quad + \int_{\theta_1^{\min}}^{\hat{\theta}_1} \mathbb{E}_{\theta_2} [v_1(k^*(s, \theta_2, \delta))] ds - \mathbb{E}_{\theta_1, \theta_2} [S_1(\theta_1, \theta_2, \delta)].
\end{aligned}$$

Hence, by making use of the Leibniz integral rule,

$$\begin{aligned}
& \frac{\partial}{\partial \hat{\theta}_1} \mathbb{E}_{\theta_2} \left[ u_1(k^*(\hat{\theta}_1, \theta_2, \delta), t_1^*(\hat{\theta}_1, \theta_2, \delta), t_2^*(\hat{\theta}_1, \theta_2, \delta), \theta_2 \mid \theta_1) \right] \\
&= \mathbb{E}_{\theta_2} \left[ \frac{\partial}{\partial \hat{\theta}_1} \pi_1(k^*(\hat{\theta}_1, \theta_2, \delta) \mid \theta_1) \right] + \mathbb{E}_{\theta_2} [v_1(k^*(\hat{\theta}_1, \theta_2, \delta))] - \mathbb{E}_{\theta_2} \left[ \frac{\partial}{\partial \hat{\theta}_1} \pi_1(k^*(\hat{\theta}_1, \theta_2, \delta) \mid \hat{\theta}_1) \right] \\
&= (\theta_1 - \hat{\theta}_1) \cdot \frac{\partial}{\partial \hat{\theta}_1} \mathbb{E}_{\theta_2} [v_1(k^*(\hat{\theta}_1, \theta_2, \delta))].
\end{aligned}$$

By assumption, the expected value in the last line is non-negative. Hence, truth-telling,  $\hat{\theta}_1 = \theta_1$ , maximizes agent 1's interim expected utility. By symmetry,  $\hat{\theta}_2 = \theta_2$ . As the above arguments hold for any set of (non-degenerate) type distributions,  $T^*$  *strongly* Bayesian implements  $k^*$ .  $\blacksquare$

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

Suppose externality types are common knowledge, and assume that the sensitive allocation rule  $k^* : \Theta \times \Delta \rightarrow K$  is strongly Bayesian implemented by the ex post budget-balanced transfer scheme  $T = (t_1, t_2) : \Theta \times \Delta \rightarrow \mathbb{R}^2$ . Define

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

where  $\pi_i(k | \theta_i) = \theta_i v_i(k) + h_i(k)$ . For  $i \in \{1, 2\}$ , denote by  $U_i(\hat{\theta}_i | \theta_i, \delta)$  agent  $i$ 's interim expected utility from reporting  $\hat{\theta}_i$  if her true payoff type is  $\theta_i$  and if agent  $-i$  reports her payoff type truthfully:

$$(81) \quad U_i(\hat{\theta}_i | \theta_i, \delta) = \theta_i \bar{v}_i(\hat{\theta}_i, \delta) + \bar{h}_i(\hat{\theta}_i, \delta) + \bar{t}_i(\hat{\theta}_i, \delta) + \delta_i \bar{\pi}_{-i}(\hat{\theta}_i, \delta) + \delta_i \bar{t}_{-i}(\hat{\theta}_i, \delta)$$

By the same reasoning that has led to equation (55) in the proof Lemma 1, the following must hold for all  $i$  and all  $(\theta_i, \delta)$ :

$$(82) \quad U_i(\theta_i | \theta_i, \delta) = p_i(\delta) + \int_{\theta_i^{\min}}^{\theta_i} \bar{v}_i(s, \delta) ds$$

for some function  $p_i : \Delta \rightarrow \mathbb{R}$ . For ease of notation, write  $t_i = t_i(\theta, \delta)$  and  $\pi_i = \pi_i(k^*(\theta, \delta) | \theta_i)$ . Then, by (82), the transfer scheme  $T$  must satisfy the following identities:

$$(83) \quad \mathbb{E}_{\theta_2}[t_1] + \delta_1 \mathbb{E}_{\theta_2}[t_2] = p_1(\delta) + \int_{\theta_1^{\min}}^{\theta_1} \bar{v}_1(s, \delta) ds - \mathbb{E}_{\theta_2}[\pi_1] - \delta_1 \mathbb{E}_{\theta_2}[\pi_2],$$

$$(84) \quad \mathbb{E}_{\theta_1}[t_2] + \delta_2 \mathbb{E}_{\theta_1}[t_1] = p_2(\delta) + \int_{\theta_2^{\min}}^{\theta_2} \bar{v}_2(s, \delta) ds - \mathbb{E}_{\theta_1}[\pi_2] - \delta_2 \mathbb{E}_{\theta_1}[\pi_1].$$

Due to budget balance, identities (83) and (84) imply that the interim expected transfers must satisfy

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

and

$$\begin{aligned} -(1 - \delta_1) \mathbb{E}_{\theta_2}[t_2] &= p_1(\delta) + \int_{\theta_1^{\min}}^{\theta_1} \bar{v}_1(s, \delta) ds - \mathbb{E}_{\theta_2}[\pi_1] - \delta_1 \mathbb{E}_{\theta_2}[\pi_2], \\ (1 - \delta_2) \mathbb{E}_{\theta_1}[t_2] &= p_2(\delta) + \int_{\theta_2^{\min}}^{\theta_2} \bar{v}_2(s, \delta) ds - \mathbb{E}_{\theta_1}[\pi_2] - \delta_2 \mathbb{E}_{\theta_1}[\pi_1], \end{aligned}$$

whereas ex post budget balance requires that also  $t_1 + t_2 = 0$ . From these conditions, the transfer scheme  $T^*$  can be ‘‘guessed’’.  $\blacksquare$

## Proof of Proposition 5 Continued

Implicit differentiation of (33) yields:  $\frac{\partial k^*}{\partial \delta_1} = \frac{\theta_2 v(1-k^*)}{\partial F / \partial k^*}$ ;  $\frac{\partial k^*}{\partial \delta_2} = \frac{-\theta_1 v(k^*)}{\partial F / \partial k^*}$ ;  $\frac{\partial k^*}{\partial \theta_1} = \frac{-v(1-k^*) - \delta_2 v(k^*)}{\partial F / \partial k^*}$ ; and  $\frac{\partial k^*}{\partial \theta_2} = \frac{\delta_1 v(1-k^*) + v(k^*)}{\partial F / \partial k^*}$ .



Notice that  $\frac{\partial F}{\partial k^*} = -(\theta_1 - \delta_1\theta_2)v'(1 - k^*) - (\theta_2 - \delta_2\theta_1)v'(k^*) < 0$ , since  $(\theta_i - \delta_i\theta_{-i}) > 0$  as  $\delta_i^{\max} < \frac{\theta_i^{\min}}{\theta_{-i}^{\max}}$ . By substituting for (33),  $v(1 - k^*) + \delta_2v(k^*) = \frac{1 - \delta_1\delta_2}{\theta_1 - \delta_1\theta_2}\theta_2v(k^*) > 0$  and  $\delta_1v(1 - k^*) + v(k^*) = \frac{1 - \delta_1\delta_2}{\theta_1 - \delta_1\theta_2}\theta_1v(k^*) > 0$ . Hence,  $\frac{\partial k^*}{\partial \delta_1} < 0 < \frac{\partial k^*}{\partial \delta_2}$  and  $\frac{\partial k^*}{\partial \theta_2} < 0 < \frac{\partial k^*}{\partial \theta_1}$ .

Implicit differentiation of (34) yields:  $\frac{\partial k^*}{\partial \delta_1} = \frac{\theta_2^2 v(1 - k^*)}{\partial G / \partial k^*}$ ;  $\frac{\partial k^*}{\partial \delta_2} = \frac{-\theta_1^2 v(k^*)}{\partial G / \partial k^*}$ ;  
 $\frac{\partial k^*}{\partial \theta_1} = \frac{(\theta_2 - 2\delta_2\theta_1)v(k^*) - \theta_2v(1 - k^*)}{\partial G / \partial k^*}$ ; and  $\frac{\partial k^*}{\partial \theta_2} = \frac{\theta_1v(k^*) - (\theta_1 - 2\delta_1\theta_2)v(1 - k^*)}{\partial G / \partial k^*}$ .

Notice that  $\frac{\partial G}{\partial k^*} = -\theta_2(\theta_1 - \delta_1\theta_2)v'(1 - k^*) - \theta_1(\theta_2 - \delta_2\theta_1)v'(k^*) < 0$ , since  $(\theta_i - \delta_i\theta_{-i}) > 0$  as  $\delta_i^{\max} < \frac{\theta_i^{\min}}{\theta_{-i}^{\max}}$ . By substituting for (34),

$$\begin{aligned}
(85) \quad & (\theta_2 - 2\delta_2\theta_1)v(k^*) - \theta_2v(1 - k^*) \\
&= -(\delta_1\theta_2^2 - 2\delta_1\delta_2\theta_1\theta_2 + \delta_2\theta_1^2) \frac{v(k^*)}{\theta_1 - 2\delta_1\theta_2} \\
&= -[(\delta_1\theta_2 - \delta_2\theta_1)^2 + \delta_1(1 - \delta_1)\theta_2^2 + \delta_2(1 - \delta_2)\theta_1^2] \frac{v(k^*)}{\theta_1 - 2\delta_1\theta_2} \\
&= -[\theta_1v(k^*) - (\theta_1 - 2\delta_1\theta_2)v(1 - k^*)].
\end{aligned}$$

Hence,  $\text{sgn}(\frac{\partial k^*}{\partial \theta_1} \frac{\partial k^*}{\partial \theta_2}) = -1 = \text{sgn}(\frac{\partial k^*}{\partial \delta_1} \frac{\partial k^*}{\partial \delta_2})$ . When assuming  $\Delta_i \subset [0, \frac{\theta_i^{\min}}{2\theta_{-i}^{\max}}]$ , the term in the third line of (85) is negative. In this case,  $\frac{\partial k^*}{\partial \theta_2} < 0 < \frac{\partial k^*}{\partial \theta_1}$ . ■

## 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.
- Becker, Gary S. 1981. *A treatise on the family*. Cambridge, MA: Harvard University Press.

- Bergemann, Dirk and Stephen Morris. 2005. “Robust mechanism design.” *Econometrica* 73 (6):1771–1813.
- . 2009. “Robust implementation in direct mechanisms.” *Review of Economic Studies* 76 (4):1175–1204.
- . 2011. “Robust implementation in general mechanisms.” *Games and Economic Behavior* 71 (2):261–281.
- . 2013. “Robust predictions in games with incomplete information.” *Econometrica* 81 (4):1251–1308.
- 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.
- Boskin, Michael J and Eytan Sheshinski. 1978. “Optimal redistributive taxation when individual welfare depends upon relative income.” *Quarterly Journal of Economics* 92 (4):589–601.
- 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.
- Mas-Colell, Andreu, Michael D Whinston, and Jerry R Green. 1995. *Microeconomic theory*. New York: Oxford University Press.
- 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.