Universitat Pompeu Fabra Barcelona
Department of Economics and Business

Economics Working Paper Series

Working Paper No. 1676

Implementation via transfers with identical but unknown distributions

Mariann Ollár and Antonio Penta

November 2019

Implementation via Transfers with Identical but Unknown Distributions *

Mariann Ollár
University of Edinburgh

Antonio Penta
ICREA, Universitat Pompeu Fabra and Barcelona GSE

November 6, 2019

Abstract

We consider mechanism design environments in which agents commonly know that types are identically distributed across agents, but without assuming that the actual distribution is common knowledge, nor that it is known to the designer *(common knowledge of identicality)*. Under these assumptions, we explore problems of partial and full implementation, as well as robustness. First, we characterize the transfers which are incentive compatible under the assumption of common knowledge of identicality, and provide necessary and sufficient conditions for partial implementation. Second, we characterize the conditions under which full implementation is possible via direct mechanisms, as well as the transfer schemes which achieve full implementation whenever it is possible. Finally, we study the robustness properties of the implementing transfers with respect to misspecifications of agents' preferences and with respect to lower orders beliefs in rationality.

KEYWORDS: Moment Conditions, Robust Full Implementation, Rationalizability, Interdependent Values, Identical but Unknown Distributions, Uniqueness, Strategic Externalities, Canonical Transfers, Loading Transfers, Equal-externality Transfers.

JEL: D62, D82, D83

1 Introduction

Many economic models assume that agents believe that the types of others are drawn from the same distribution. This is a natural way to represent situations in which agents regard each other as ex-ante symmetric from an informational viewpoint, or more broadly that they come from a common population. Standard modeling techniques, however, not only impose that the distribution of types is *identical* across agents, but also that it is common knowledge among them – and, in mechanism design, also known to the designer. But if *identicality* is a natural way to capture a basic qualitative property of these environments, *common knowledge* of the distribution is a

*We are grateful to Eddie Dekel, Philippe Jehiel, George Mailath, and Rakesh Vohra for their comments. We also thank seminar audiences at the Univ. of Edinburgh, Bar-Ilan Univ, Tel-Aviv Univ. and the Hebrew Uni. of Jerusalem, Carnegie-Mellon, Penn State, ICEF (Moscow) and participants to the 2019 Warwick Economic Theory Workshop (Warwick Univ.) and to the Workshop on New Directions in Mechanism Design (Stony Brook, 2019). Antonio Penta acknowledges the financial support of the Spanish Ministry of Economy and Competitiveness, through the Severo Ochoa Programme for Centres of Excellence in R&D (SEV-2015-0563), and of the European Research Council (ERC), ERC Starting Grant #759424.

different kind of assumption: not only is it strong and unlikely to be satisfied; it is also well-known to heavily affect the results.[1]

A large and growing literature has taken up Wilson's (1987) call for a "[...] repeated weakening of common knowledge assumptions [...]", and developed a robust approach to mechanism design.[2] In this paper we pursue the objectives of the *Wilson doctrine* in settings with informationally symmetric agents. To do this, we maintain common knowledge that types are identically distributed, but without assuming that the actual distribution is common knowledge, nor that it is known to the designer. Under these assumptions, we explore questions of *partial* and *full* implementation: First, we characterize the transfers which are incentive compatible under the assumption of identicality, and provide necessary and sufficient conditions for partial implementation. Second, we characterize the conditions under which full implementation via direct mechanisms is possible under common knowledge of identicality, as well as the transfers which achieve it whenever possible. Finally, we study the robustness of the implementing transfers with respect to the possibility that agents are 'slightly faulty' (e.g., Eliaz (2002) – or equivalently, that their preferences are slightly misspecified), and with respect to lower orders of rationality (which is closely connected to recent work on level-k implementation by de Clippel et al. (2018)).

While direct mechanisms and transfers are without loss for partial implementation, insisting on these mechanisms does restrict the environments in which full implementation is possible. But there are many advantages which come from this restriction. First, classical results on full implementation typically involve unrealistically complicated mechanisms, which have been criticized for providing limited economic insight (see, e.g., Jackson (1992)). Our insistence on using the same class of mechanisms as is typical in the partial implementation literature allows for an easier comparison with that literature, which favors the interpretability of the results and hence addresses Jackson's concern for economic 'relevance'. This approach also enables us to uncover what features of an incentive compatible transfer scheme – namely, as we show, the structure of its *strategic externalities* – may or may not be problematic from the full implementation viewpoint. With this understanding, our approach develops constructive insights on how failures of full implementation can be overcome maintaining the same fundamental structure as the mechanisms for partial implementation, which have a clear economic interpretation.

A central role throughout our analysis is played by the *canonical direct mechanism*, which characterizes the mechanisms for belief-free implementation (to fix ideas, in efficient implementation, the transfers in the canonical direct mechanism are the VCG transfers). Our first result shows that a transfer scheme t is incentive compatible under common knowledge of identicality if and only if it can be written as the sum of the canonical transfers and an extra component which satisfies a certain condition for all beliefs consistent with identicality. As it turns out, such condition implies that, for all beliefs consistent with the model, both the first- and the second-order derivatives of agents' optimization problem given t coincide with those given the canonical transfers. It follows that, when only common knowledge of identicality is maintained, partial implementation is possi-

[1] On the impact of common knowledge assumptions in game theory, see, for instance Rubinstein (1989); Carlsson and Van Damme (1993); Kajii and Morris (1997); Morris and Shin (1998, 2003); Weinstein and Yildiz (2007,b, 2011, 2016); Penta (2012, 2013); Penta and Zuazo-Garin (2017).

[2] This literature was spurred by the seminal works in belief-free settings by Bergemann and Morris (2005, 2009a,b, 2011) for static mechanisms, and further developed by Müller (2016, 2018) and Penta (2015) for dynamic ones. Settings with partial restrictions on agents' beliefs are considered by Lopomo et al. (2011); Artermov et al. (2013); Guo and Yannelis (2017); Ollár and Penta (2017). See also Yamashita (2015), Borgers and Smith (2014), Wolitzky (2016) and Carroll (2015) for alternative approaches.

ble if and only if it can be achieved by the canonical transfers. The canonical transfers therefore are all which needs to be considered to achieve partial implementation under common knowledge of identicality. This, however, is not to say that partial implementation in these settings is as demanding as ex-post incentive compatibility: the latter notion is indeed more demanding; yet, considering the same mechanism suffices for both.

Regarding full implementation, Ollár and Penta (2017) showed that the reason why the canonical transfers fail to achieve full implementation in environments with strong preference interdependence (a result due to Bergemann and Morris (2009a)) is that they induce too strong *strategic externalities*. Ollár and Penta (2017)'s idea was to use information about agents' beliefs, if available, to design incentive compatible transfers which induce small strategic externalities (and hence uniqueness and full implementation) even when preference interdependencies are strong. They thus provided sufficient conditions on agents' beliefs so that the designer could engineer such weakening of strategic externalities.[3] It turns out, however, that if only common knowledge of identicality is maintained, without assuming knowledge of the actual distribution of types, then Ollár and Penta (2017)'s design strategy cannot be pursued: under common knowledge of identicality, any incentive compatible mechanism must display the same total level of strategic externalities as the canonical direct mechanism. Hence, under common belief of identicality, the designer may only pursue a *redistribution* – not a reduction – of the strategic externalities, which in turn are pegged to the level of preference interdependence in the environment. This obviously limits the possibility of achieving full implementation, and requires developing a new design strategy.

Our analysis of full implementation develops such a novel design strategy. The key is to understand how the strategic externalities in the canonical direct mechanism can be optimally re-assigned among the agents, as well as providing constructive insights on how this can be achieved by adding a belief-based component to the canonical transfers. For environments with single-crossing preferences and public concavity, our main result is a full characterization of the conditions for full implementation under common knowledge of identicality, as well as of the transfer scheme which achieves it whenever possible. The transfers in our characterization have a special hierarchical structure: besides preserving, for any player, the total level of strategic externalities he is subject to from his opponents – which, by the results above, is necessary to preserve incentive compatibility when only common belief in identicality is assumed – these transfers *load* all the strategic externalities on the opponent who displays the lowest amount of preference interdependence. The structure of the *loading transfers* enables us to uncover a fairly surprising result: the possibility of full implementation is characterized by the strength of the preference interdependence of the two agents with the least amount of preference interdependence, regardless of the number of the other agents, and of their preferences.

Such a characterization has powerful implications from a broader market design perspective: for instance, if full implementation cannot be achieved for a set of agents (which, by our results, would be due to too strong preference interdependence), adding two more agents whose preferences do not depend much on others' information would suffice to make full implementation possible. At the extreme, whenever the environment includes two agents with private values, common belief in identicality ensures that full implementation is possible via a simple direct mechanism.

[3]For instance, Ollár and Penta (2017) showed that strategic externalities can always be eliminated in common prior models with independent or affiliated types under certain preference restrictions, and hence full implementation be achieved in (interim) dominant strategies.

Besides the *loading transfers*, which as explained have a strongly asymmetric structure, we also consider the *equal-externality transfers*, which evenly redistribute the strategic externalities across the opponents. Such an alternative design strategy is not without loss of generality for full implementation under common knowledge of identicality (there are environments in which full implementation is possible, but not with the equal-externality transfers). Nonetheless, we show that these transfers are still widely applicable, and that their symmetric structure grants them an important robustness property. In particular, while the loading transfers have several desirable robustness properties (for instance, they minimize the sensitivity of implementation with respect to lower-orders of rationality – cf. de Clippel et al. (2018)), we show that the equal-externality transfers minimize the impact on the implemented allocation with respect to the possibility of 'slightly faulty' agents or of misspecification of their preferences (cf., Eliaz (2002)).

The rest of the paper is organized as follows: Section 2 introduces the model and presents some illustrating examples; Sections 3 and 4 provide the characterizations of partial and full implementation, respectively. Section 5 focuses on alternative design strategies for full implementation via transfers. The sensitivity analysis is provided in Section 6.

2 Model

We consider environments with transferable utility with a finite set of agents $I = \{1, ..., n\}$, in which the space of allocations X is a compact and convex subset of a Euclidean space. Agents privately observe their payoff types $\theta_i \in \Theta_i := [\underline{\theta}, \overline{\theta}] \subseteq \mathbb{R}$, are drawn from a closed interval on the real line, common to all agents (the latter assumption is inherent to our main question, which is to study the assumption of identical distributions). We adopt the standard notation $\theta_{-i} \in \Theta_{-i} = \times_{j \neq i} \Theta_j$ and $\theta \in \Theta = \times_{i \in I} \Theta_i$ for profiles. Agent i's valuation function is $v_i : X \times \Theta \to \mathbb{R}$, assumed twice continuously differentiable, and we let $t_i \in \mathbb{R}$ denote the private transfer to agent i: for each outcome $(x, \theta, (t_i)_{i \in I})$, i's utility is equal to $v_i(x, \theta) + t_i$. The tuple $\langle I, (\Theta_i, v_i)_{i \in I} \rangle$ is common knowledge among the agents. If v_i is constant in θ_{-i} for every i, then the environment has private values. If not, it has interdependent values.

An allocation rule is a mapping $d : \Theta \to X$ which assigns to each payoff state the allocation that the designer wishes to implement. We focus on allocation rules that are twice continuously differentiable and responsive, in the sense that for all i and $\theta_i \neq \theta_i'$, there exists $\theta_{-i} \in \Theta_{-i}$ such that $d(\theta_i, \theta_{-i}) \neq d(\theta_i', \theta_{-i})$ (see, e.g., Bergemann and Morris (2009a)).

The model accommodates general externalities in consumption, including both pure cases of private and public divisible goods. The main substantive restrictions are the one-dimensionality of types, and the smoothness of the allocation function, which for instance rules out standard auction applications. We will use the notation $\partial f / \partial x$ for all derivatives, with the understanding that when X is multidimensional, $\frac{\partial v_i}{\partial x}(x, \theta)$ and $\frac{\partial d}{\partial \theta_i}(\theta)$ denote the vectors of partial derivatives and $\frac{\partial v_i}{\partial x}(x, \theta) \cdot \frac{\partial d}{\partial \theta_i}(\theta)$ denotes their inner product.

We assume that agents commonly know that types are identically distributed across agents, but they do not necessarily know (or agree on) the actual distribution, which importantly is unknown to the designer. Hence, for each type θ_i, the designer regards many beliefs $B_{\theta_i}^{id} \subseteq \Delta(\Theta_{-i})$ as possible for type θ_i, namely all those which are consistent with common knowledge that types are

identically distributed.[4] Formally, the designer's assumption about beliefs is represented by belief restrictions $\mathcal{B}^{id} = ((B^{id}_{\theta_i})_{\theta_i \in \Theta_i})_{i \in I}$ such that:[5]

$$B^{id}_{\theta_i} = \{b_{\theta_i} \in \Delta(\Theta_{-i}) : \operatorname*{marg}_{\Theta_j} b_{\theta_i} = \operatorname*{marg}_{\Theta_k} b_{\theta_i} \text{ for all } j,k \neq i\} \text{ for all } i \text{ and } \theta_i. \tag{1}$$

These belief restrictions entail weaker assumptions on agents' beliefs than many standard models in more applied theory and in empirical microeconomics.[6] The belief restrictions in (1) are weaker, for example, than assuming: (i) a joint distribution with identical marginals over agents' types; (ii) a joint distribution with exchangeable random variables; (iv) known independent and identical distributions across agents (as in standard common prior i.i.d. environments); (v) independent and identical but *unknown* distributions; (vi) unobserved heterogeneity but symmetrically distributed values; (vi) environments with pure common values in which the state of the world is unknown to the designer, but commonly known by the agents; etc. Hence, our belief restrictions entail a very weak level of common knowledge in the environment.

We consider direct mechanisms, in which agents report their type and the allocation is chosen according to d. A direct mechanism is thus uniquely determined by a transfer scheme $t = (t_i)_{i \in I}$, $t_i : M \to \mathbb{R}$, which specifies the transfer to each agent i, for all profiles of reports $m \in \Theta$. (To distinguish the report from the state, we maintain the notation m_i even though the message spaces are $M_i = \Theta_i$.) Any transfer scheme induces a game with ex-post payoff functions $U^t_i(m;\theta) = v_i(d(m),\theta) + t_i(m)$. When the transfers are clear from the context, we don't emphasize the dependence of the payoff functions on t. For the analysis of partial implementation, in which each agent expects his opponents to report truthfully, the following notation will be useful: For any θ_i, $b_{\theta_i} \in \Delta(\Theta_{-i})$ and m_i, we let $E^{b_{\theta_i}}(U_i(m_i,\theta_{-i};\theta_i,\theta_{-i}))) := \int_{\Theta_{-i}} U_i(m_i,\theta_{-i};\theta_i,\theta_{-i})\, db_{\theta_i}$. For full implementation instead we will also consider other (non-truthful) reporting strategies for the opponents, and also use the following notation: For every $\theta_i \in \Theta_i$, $\mu \in \Delta(M_{-i} \times \Theta_{-i})$ and $m_i \in M_i$, we let $EU^{\mu}_{\theta_i}(m_i) = \int_{M_{-i} \times \Theta_{-i}} U_i(m_i, m_{-i};\theta_i,\theta_{-i})\, d\mu$ denote agent i's expected payoff from message m_i, if i's type is θ_i and his conjectures are μ, and define $BR_{\theta_i}(\mu) := \arg\max_{m_i \in M_i} EU^{\mu}_{\theta_i}(m_i)$.

2.1 Leading Examples and Preview of Results

In this section we provide some examples to illustrate the key ideas of the paper and their connection with the previous literature. The examples are all based on the following environment: There are three agents, $\{1,2,3\}$, with preferences over the quantity $x \in \mathbb{R}_+$ of public good such that $v_i(x,\theta) = (\theta_i + \gamma_{ij}\theta_j + \gamma_{ik}\theta_k)\, x$ for all i, where we let $j := i+1(\text{mod}3)$ and $k := i+2(\text{mod}3)$. Types $\theta_i \in [0,1]$ are private information to each agent i, and for each i and $j \neq i$, $\gamma_{ij} \in \mathbb{R}$ is a parameter of preference interdependence. The social planner wishes to implement the efficient allocation rule. With production cost $c(x) = x^2/2$, the efficient decision rule is $d(\theta) = \sum_{i=1}^{3} \kappa_i \theta_i$,

[4] For a measurable set E, $\Delta(E)$ denotes the set of probability measures on its Borel σ-algebra.

[5] The notion of a belief restriction is introduced by Ollár and Penta (2017) to model general restrictions on agents' beliefs: a *belief restriction* is a commonly known collection $\mathcal{B} = ((B_{\theta_i})_{\theta_i \in \Theta_i})_{i \in I}$ such that $B_{\theta_i} \subseteq \Delta(\Theta_{-i})$ is non-empty and convex for all i and θ_i, and $B_i : \theta_i \to B_{\theta_i} \subseteq \Delta(\Theta_{-i})$ is continuous for every i. As discussed in Ollár and Penta (2017), special cases of interest include (i) standard Bayesian environments, in which B_{θ_i} is a singleton for all θ_i and i; (ii) common prior environments, in which $\exists p \in \Delta(\Theta)$ such that $B_{\theta_i} = \{p(\cdot|\theta_i)\}$ for all i and θ_i; (iii) belief-free environments, in which $B_{\theta_i} = \Delta(\Theta_{-i})$ for all i and θ_i.

[6] Models with identical distributions of agents' types are often applied to study, for example, information aggregation in voting (e.g., Levy and Razin (2015)), information aggregation in exchanges (e.g., Ollár (2017)) and identification in auctions with symmetric bidders (e.g., Athey and Haile (2007); Hendricks et al. (2003)).

where $\kappa_i \equiv 1 + \gamma_{ji} + \gamma_{ki}$ for all i, which for convenience we assume positive. Given this environment, we consider three sets of assumptions on agents' beliefs: (i) a belief-free setting, (ii) a standard common prior environment, and (iii) a setting in which only common belief in identicality is maintained – the main focus of this paper.

Belief-Free Implementation. If the designer has no information about agents' beliefs, or if he wishes to achieve implementation without relying on any belief restriction, then only the generalized VCG mechanism (cf. Bergemann and Morris (2009a)) can be used.

Example 1 (Belief-Free Implementation)**.** In our example, the VCG transfers are the following:

$$t_i^*(m) = -\kappa_i \left(0.5 m_i^2 + m_i \left(\gamma_{ij} m_j + \gamma_{ik} m_k\right)\right).$$

Given this, as long as $\kappa_i > 0$ for all i, for any profile (θ_{-i}, m_{-i}) of opponents' types and reports, the ex-post best-reply function for type θ_i of player i is

$$BR_{\theta_i}^*(\theta_{-i}, m_{-i}) = proj_{[0,1]}\left(\theta_i + \sum_{j \neq i} \gamma_{ij}(\theta_j - m_j)\right).^7 \tag{2}$$

Observe that, regardless of what γ is, *for any realization* of θ, truthful revelation $(m_i(\theta_i) = \theta_i)$ is a best response to the opponent's truthful strategy $(m_j(\theta_j) = \theta_j)$. This is the well-known ex-post incentive compatibility of the VCG mechanism. Partial implementation of the efficient allocation is thus guaranteed independent of agents' beliefs. Furthermore, if $\sum_{j \neq i} |\gamma_{ij}| < 1$ for all $i \in I$, then equation (2) is a contraction, and its iteration delivers truthful revelation as the only rationalizable strategy. In this case, the VCG mechanism also guarantees full belief-free implementation. Full implementation, however, is only possible if the preference interdepependence is 'small'. For instance, suppose that preference parameters are such that

$$(\gamma_{12}, \gamma_{13}, \gamma_{21}, \gamma_{2,3}, \gamma_{31}, \gamma_{32}) = (0.9, -0.5, 1.2, -0.6, -0.8, 1.6) =: \hat{\gamma}$$

Then, all report profiles are rationalizable, and hence full belief-free implementation fails. □

Hence, *partial* belief-free implementation is always possible in this setting, but *full* belief-free implementation fails if the preference interdependence is too strong (Bergemann and Morris (2009a)). The reason is that if preference interdepedence is strong, then players' best responses in the VCG mechanism are strongly affected by others' strategies. This in turn generates multiplicity of equilibria, and hence failure of full implementation. We thus shift the focus from preference interdependence to the *strategic externalities* of a mechanism, which can be captured by studying how agents' best responses are affected by changes in the opponents' report. This information can be conveniently summarized in a *strategic externality matrix,* whose ij-th entry contains the derivative of player i's best response with respect to j's report, for $j \neq i$, normalized by the concavity of i's payoff function with respect to his own report. In the case of the canonical

[7] For any $y \in \mathbb{R}$, we let $proj_{[0,1]}(y) := \arg\min_{\theta_i \in [0,1]} |\theta_i - y|$ denote the projection of y on the interval $[0,1]$.

mechanism, this amounts to:

$$SE^* = \begin{bmatrix} 0 & \gamma_{12} & \gamma_{13} \\ \gamma_{21} & 0 & \gamma_{23} \\ \gamma_{31} & \gamma_{32} & 0 \end{bmatrix}.$$

Identical and Known Distribution: Reduction of Strategic Externalities. Strategic externalities and preference interdependence coincide in the VCG mechanism. But if the designer has some information about the agents' beliefs, then this necessary coincidence is more relaxed: the strategic externalities can be weakened, so as to ensure uniqueness, even if preference interdependence is strong. This is the main insight from Ollár and Penta (2017).

Example 2 (Known i.i.d. Common Prior)**.** Suppose that types are commonly known to be i.i.d. draws from a uniform distribution over $[0,1]$, and this is known to the designer. Consider the following transfers, which are a special case of Proposition 3 in Ollár and Penta (2017):

$$t_i^{OP}(m) := t_i^* + m_i \kappa_i \left(\sum_{l \neq i} \gamma_{il} (m_l - 0.5) \right) = -\kappa_i \left(\frac{1}{2} m_i^2 + m_i \sum_{l \neq i} \gamma_{il} 0.5 \right). \tag{3}$$

These transfers induce the following best response function:

$$BR_{\theta_i}^{OP}(\mu) = proj_{[0,1]} \left(\theta_i + \sum_{l \neq i} \gamma_{il} \left[E(\theta_l | \theta_i) - 0.5 \right] \right). \tag{4}$$

Under the maintained assumptions, $E(\theta_l|\theta_i) = 0.5$ for all θ_i and $l \neq i$. Hence the term in square brackets cancels out for all types. Truthful revelation therefore is *strictly dominant*, and full implementation is achieved for any γ. Players' best-responses are not affected by other reports, and hence strategic externalities were completely eliminated in this case. □

The result in this example does rely on the restriction on agents' beliefs, and in particular on the knowledge that "$E(\theta_l|\theta_i) = 0.5$ for all θ_i and $l \neq i$". If this *moment condition* were not satisfied, these transfers would achieve neither full nor partial implementation. This moment condition was used in (3) to weaken the strategic externalities of the baseline transfers from Example 1, but in principle others could be used too.[8] Intuitively, the more information the designer has about agents' beliefs, the more freedom he has to choose a convenient moment condition. As shown by Ollár and Penta (2017), common prior models are maximal in the freedom they allow to the designer and, for a large class of environments, as in the example, strategic externalities can be completely eliminated when types are independent or affiliated.

Identical but Unknown Distribution: Redistribution of Strategic Externalities. Now suppose that agents commonly know that their types $\theta_i \in [0,1]$ follow the same distribution over Θ_i. The distribution itself, however, is not necessarily known to the agents and, most importantly, it is unknown to the designer. Transfers from the previous example do not ensure implementation anymore, since agents' beliefs neet not satisfy the moment condition "$E(\theta_l|\theta_i) = 0.5$ for all θ_i and

[8]The idea of modifying ex-post incentive compatible transfers using information about beliefs appears in previous literature as in d'Aspremont, Cremer and Gerard-Varet (1979), Arrow (1979), Cremer and McLean (1988), and more recently in Mathevet (2010); Mathevet and Taneva (2013); Healy and Mathevet (2012); Deb and Pai (2017).

$l \neq i$", and hence incentive compatibility may fail. In fact, as we will show, Ollar and Penta's (2017) idea of *reducing strategic externalities* is incompatible with incentive compatibility under these belief restrictions. The designer is therefore much more limited than in a standard common prior setting, such as that of the previous example. Nonetheless, a novel design strategy, based on a *redistribution of the strategic externalities*, may still be used to achieve full implementation.

Example 3 ($\mathcal{B}^{id}$-Implementation)**.** Suppose that $\gamma = \hat{\gamma}$ as at the end of Example 1, and hence belief-free implementation is not possible. Now consider the following transfers:

$$t_i^e(m) = t_i^*(m) + m_i \kappa_i \frac{\gamma_{ij} - \gamma_{ik}}{2}(m_j - m_k) \text{ for all } i;$$

In this case, the best replies become

$$\begin{aligned} BR_{\theta_i}^e &= proj_{[0,1]}\left(\theta_i + \frac{1}{2}(\gamma_{ij} + \gamma_{ik}) \sum_{l \neq i} E((\theta_l - m_l)|\theta_i) + \frac{1}{2}(\gamma_{ij} - \gamma_{ik}) E(\theta_j - \theta_k)|\theta_i\right) \\ &= proj_{[0,1]}\left(\theta_i + \frac{1}{2}(\gamma_{ij} + \gamma_{ik}) \sum_{l \neq i} E((\theta_l - m_l)|\theta_i)\right) \end{aligned}$$

The simplification in the last line follows from the fact that, under the $\mathcal{B}^{id}$ restrictions, it is common knowledge that $E(\theta_j - \theta_k|\theta_i) = 0$ for all θ_i and i. Because of this simplification, this mechanism is incentive compatible for all beliefs consistent with $\mathcal{B}^{id}$: if $m_l = \theta_l$ for all θ_l and $l \neq i$, then the best response is $m_i = \theta_i$ for all i. Moreover, it can be shown that these best-replies induce a contraction, which ensures that truthful revelation is the only rationalizable profile for all agents. Transfers $(t_i^e)_{i \in I}$ therefore achieve both partial and full $\mathcal{B}^{id}$-implementation.

Next consider the following, more complex, transfers:

$$\begin{bmatrix} t_1^l(m) \\ t_2^l(m) \\ t_3^l(m) \end{bmatrix} = \begin{bmatrix} t_1^*(m) + m_1 \kappa_1 \gamma_{13}(m_3 - m_2) \\ t_2^*(m) + m_2 \kappa_2 \gamma_{23}(m_3 - m_1) \\ t_3^*(m) + m_3 \kappa_3 \gamma_{32}(m_2 - m_1) \end{bmatrix}.$$

It can be shown that these transfers too are incentive compatible under the $\mathcal{B}^{id}$-restrictions, that is, they are based on moment conditions which are commonly known among the agents. Moreover, these transfers too, induce contractive best replies, and hence achieve full implementation.

To understand the logic behind these transfers, it is useful to look at the induced SE-matrices when $\gamma = \gamma^*$, and compare them to the SE-matrix of the VCG transfers:

$$SE^* = \begin{bmatrix} 0 & 0.9 & -0.5 \\ 1.2 & 0 & -0.6 \\ -0.8 & 1.6 & 0 \end{bmatrix}, SE^e = \begin{bmatrix} 0 & 0.2 & 0.2 \\ 0.3 & 0 & 0.3 \\ 0.4 & 0.4 & 0 \end{bmatrix}, SE^l = \begin{bmatrix} 0 & 0.4 & 0 \\ 0.6 & 0 & 0 \\ 0.8 & 0 & 0 \end{bmatrix}.$$

First notice that both $(t_i^e)_{i \in I}$ and $\left(t_i^l\right)_{i \in I}$ induce SE-matrices such that the sum of the strategic externality within each row is the same as in the baseline VCG mechanism. This is not a coincidence: as one of our results will show, under the $\mathcal{B}^{id}$-restrictions, any incentive compatible transfer scheme would have to preserve, for every agent, the *total externalities* across all of his opponents which are present in the underlying canonical mechanism, which in turn are pinned

down by the total level of preference interdependence. (So, for instance, transfers such as $\left(t_i^{OP}\right)_{i\in I}$ from example 2, whose SE-matrix consists of all zeros, will not be incentive compatible under the $\mathcal{B}^{id}$-restrictions.) In this sense, strategic externalities can only be *redistributed*, not reduced.

Second, the SE-matrix of the $(t_i^e)_{i\in I}$ transfers are such that the externalities that any agent i is subject to is constant across all of his opponents. In this sense, the $(t_i^e)_{i\in I}$ transfers induce an *equal redistribution* of the total strategic externalities for every player. With the $\left(t_i^l\right)_{i\in I}$ transfers instead, for every i, the total strategic externalities are all *loaded* on the opponent $l \neq i$ who is subject to the lowest total strategic externalities (that is $l = 2$ for $i = 1$, and $l = 1$ for $i = 2, 3$).

But while both matrices induce a contraction and have the same row-sums – which implies that, in both mechanisms, the same strategies survive the first round of elimination of never best-responses – the square of the SE^l-matrix exhibits lower row-sums than that of the SE^e-matrix:

$$\left(SE^e\right)^2 = \begin{bmatrix} 0.14 & 0.08 & 0.06 \\ 0.12 & 0.18 & 0.06 \\ 0.12 & 0.08 & 0.2 \end{bmatrix}, \left(SE^l\right)^2 = \begin{bmatrix} 0.24 & 0 & 0 \\ 0 & 0.24 & 0 \\ 0 & 0.32 & 0 \end{bmatrix}.$$

Recursively, this also extends to all powers $k \geq 2$, which implies that, from the second round of elimination on, the set of rationalizable reports shrinks more under $\left(t_i^l\right)_{i\in I}$ than under $(t_i^e)_{i\in I}$.□

Our main results for full implementation show that, in a general class of environments, a suitable generalization of the *loading transfers* in the example characterizes the mechanisms which achieve full B^{id}-implementation: under these belief restrictions, full implementation is possible if and only if it is achieved by the loading transfers. This in turn enables us to characterize the environments in which full implementation is possible. We also show that the loading transfers induce the fastest contraction among all implementing mechanisms, and that they are the 'most robust' with respect to lower order beliefs in rationality. The *equal-externality* transfers, instead, are 'most robust' if one considers the possibility of mispecifications of agents' preferences.

3 Partial Implementation and Moment Conditions

In this Section we formalize our notion of partial implementation, and we characterize both the conditions under which it is possible and the transfers which achieve it. For later reference, it will be useful to introduce the *canonical transfers*, $t^* = (t_i^*(\cdot))_{i\in I}$, such that for each $i \in I$ and $m \in \Theta$,

$$t_i^*(m) = -v_i(d(m), m) + \int_{\underline{\theta}_i}^{m_i} \frac{\partial v_i}{\partial \theta_i}(d(s_i, m_{-i}), s_i, m_{-i})\, ds_i, \tag{5}$$

and we refer to the pair (d, t^*) as the *canonical direct mechanism*. Note that, under the maintained assumptions, the canonical direct mechanism induces payoff functions which are twice continuously differentiable. As shown by Ollár and Penta (2017), the canonical transfers characterize the ex-post incentive compatible transfers in general environments with interdependent valuations, up to a constant which does not depend on i's own report:[9]

[9]The term 'canonical mechanism' is traditionally used to refer to Maskin's mechanism for full implementation. That mechanism is not 'direct' and it induces an integer game to eliminate undesirable equilibria. We call (d, t^*) the canonical *direct* mechanism, since special cases of this mechanisms are pervasive in the partial implementation literature. For example, in auctions (Myerson (1981), Dasgupta and Maskin (2000),Segal (2003), Li (2017)), in pivot

Definition 1. *A direct mechanism is* ex-post incentive compatible (ep-IC) *if,* $U_i(\theta;\theta) \geq U_i(\theta_i', \theta_{-i};\theta)$ *for all* θ *and for all* θ_i'.

Lemma 1 (Lemma 1 in Ollár and Penta (2017))**.** *If* (d,t) *is differentiable and ex-post incentive compatible, then for all* i*, there exists a differentiable function* $\tau_i : M_{-i} \to \mathbb{R}$ *such that, for all* m

$$t_i(m) = t_i^*(m) + \tau_i(m_{-i}). \tag{6}$$

Moreover, if (d,t) *is differentiable, satisfies* (6)*, and if the resulting payoff functions are such that* $\partial^2 U_i(m_i, \theta_{-i};\theta)/\partial^2 m_i < 0$ *for all* m_i*, then* (d,t) *is ex-post incentive compatible.*

It is well-known that ex-post incentive compatibility characterizes the possibility of achieving partial implementation when the designer relies on no information on agents' beliefs (Bergemann and Morris (2005)). Lemma 1 therefore implies that the canonical transfers characterize the mechanisms which may achieve partial implementation in the belief-free sense.

In the present context, the designer knows that agents 'commonly believe in identicality', and hence our analysis of partial implementation relies on the following less demanding notion of incentive compatibility:[10]

Definition 2. *A direct mechanism is* $\mathcal{B}^{id}$*-incentive compatible (*$\mathcal{B}^{id}$*-IC) if for all* $i \in I$*, for all* $\theta_i, \theta_i' \in \Theta_i$*, and for all* $b_{\theta_i} \in B_{\theta_i}^{id}$*,* $E^{b_{\theta_i}}(U_i(\theta_i, \theta_{-i};\theta_i,\theta_{-i})) \geq E^{b_{\theta_i}}(U_i(\theta_i', \theta_{-i};\theta_i,\theta_{-i}))$*. If the inequality holds strictly for all* i*,* θ_i*,* $b_{\theta_i} \in B_{\theta_i}^{id}$ *and* $\theta_i' \neq \theta_i$*, then we say that it is* strictly $\mathcal{B}^{id}$*-IC. If* (d,t) *is* $\mathcal{B}^{id}$*-IC, then we say that the transfers* t $\mathcal{B}^{id}$*-partially implement the allocation function* d*.*

3.1 Partial Implementation: Characterization

In this Section we characterize properties of the transfers which partially implement a given allocation function $d : \Theta \to X$, and study necessary and sufficient conditions for $\mathcal{B}^{id}$-partial implementation. The following Lemma plays an important role in our analysis:

Lemma 2 ($\mathcal{B}^{id}$-IC Transfers: Necessary and Sufficient Conditions)**.**
[**Necessity:**] *If* (d,t) *is twice differentiable and* $\mathcal{B}^{id}$*-IC, then for all* i*, and for all* $m \in M \equiv \Theta$*,*

$$t_i(m) = \underbrace{t_i^*(m) + \tau_i(m_{-i})}_{\substack{\text{belief-free transfers} \\ \text{(ep-IC characterization)}}} + \underbrace{\int_{\underline{\theta}_i}^{m_i} K_i(s_i, m_{-i})\, ds_i}_{\text{belief-based component}} \tag{7}$$

where $\tau_i : M_{-i} \to \mathbb{R}$ *and* $K_i : M \to \mathbb{R}$ *are differentiable functions and* K_i *is such that:*

$$\mathbb{E}^{b_{\theta_i}}(K_i(\theta_i, \theta_{-i})) = 0 \text{ for all } \theta_i \text{ and for all } b_{\theta_i} \in B_{\theta_i}^{id}.^{11} \tag{8}$$

[**Sufficiency:**] *If* (d,t) *is twice differentiable,* t *satisfies (7) and (8), and the resulting payoffs are such that* $E^{b_{\theta_i}}\left(\partial^2 U_i(m_i, \theta_{-i};\theta)/\partial^2 m_i\right) < 0$ *for all* m_i *and* $b_{\theta_i} \in \mathcal{B}_{\theta_i}^{id}$*, then* (d,t) *is* $\mathcal{B}^{id}$*-IC.*

mechanisms (Milgrom (2004), Jehiel and Lamy (2017)), in public goods problems (Laffont and Maskin (1980),Green and Laffont (1977)), etc. Lemma 1 in Ollár and Penta (2017) generalized the earlier results in the papers above. The term *canonical direct mechanism* was first used with this acceptation in Ollár and Penta (2017).

[10] Similar to Bergemann and Morris (2005), one could define $\mathcal{B}^{id}$-Partial Implementation as (partial) Bayesian implementation on all type spaces consistent with the $\mathcal{B}^{id}$-restrictions. By arguments similar to Bergemann and Morris (2005), it can be shown such a notion is equivalent to the incentive compatibility condition in Def. 2.

[11] For any $f : \Theta \to \mathbb{R}$, $\theta_i \in \Theta_i$ and $b_{\theta_i} \in B_{\theta_i}^{id}$, we let $E^{b_{\theta_i}}(f(\theta_i, \theta_{-i})) := \int_{\Theta_{-i}} f_i(\theta_i, \theta_{-i})\, db_{\theta_i}$.

Equation (7) implies that, as far as $\mathcal{B}^{id}$-IC is concerned, it is without loss of generality to design transfers starting out with the canonical transfers, and then adding a *belief-based* term $K_i : M \to \mathbb{R}$. This result therefore extends the characterization of ex-post incentive compatible transfers in Lemma 1 to the belief restrictions $\mathcal{B}^{id}$. The sense in which the extra component is 'belief-dependent' is clarified by the condition in equation (8), which has to be satisfied for all beliefs consistent with $\mathcal{B}^{id}$. We will expand on the conceptual significance of this condition in Section 3.2. In the meantime, note that any twice continuously differentiable mechanism is $\mathcal{B}^{id}$-IC only if the truthful profile satisfies the first- and second-order conditions of agents' optimization problem, for all interior types and for all beliefs consistent with $\mathcal{B}^{id}$. That is, the associated payoff function must be such that, for all $\theta_i \in (\underline{\theta}, \bar{\theta})$ and $b_{\theta_i} \in B^{id}_{\theta_i}$, (i) $E^{b_{\theta_i}}\left(\partial U_i\left(\theta_i, \theta_{-i}; \theta_i, \theta_{-i}\right)/\partial m_i\right) = 0$ and (ii) $E^{b_{\theta_i}}\left(\partial^2 U_i\left(\theta_i, \theta_{-i}; \theta_i, \theta_{-i}\right)/\partial^2 m_i\right) \leq 0$. But if t partially implements d, then by Lemma 2 it can be written as in (7), and hence – letting U^* denote the payoff function of the canonical direct mechanism – for any $\theta_i \in (\underline{\theta}, \bar{\theta})$ and $b_{\theta_i} \in B^{id}_{\theta_i}$, we have:

$$E^{b_{\theta_i}}\left(\partial U_i\left(\theta_i, \theta_{-i}; \theta_i, \theta_{-i}\right)/\partial m_i\right) = E^{b_{\theta_i}}\left(\partial U_i^*\left(\theta_i, \theta_{-i}; \theta_i, \theta_{-i}\right)/\partial m_i\right) + E^{b_{\theta_i}}\left(K_i\left(\theta_i, \theta_{-i}\right)\right), \text{ and}$$
$$E^{b_{\theta_i}}\left(\partial^2 U_i\left(\theta_i, \theta_{-i}; \theta_i, \theta_{-i}\right)/\partial^2 m_i\right) = E^{b_{\theta_i}}\left(\partial^2 U_i^*\left(\theta_i, \theta_{-i}; \theta_i, \theta_{-i}\right)/\partial^2 m_i\right) + E^{b_{\theta_i}}\left(\partial K_i\left(\theta_i, \theta_{-i}\right)/\partial m_i\right).$$

Condition (8) in Lemma 2 implies that the second term on the right-hand side of the first equation is zero, and hence the first-order conditions of any $\mathcal{B}^{id}$-IC mechanism coincide with those of the canonical direct mechanism. Furthermore, it can be shown that any K_i function which satisfies condition (8) also ensures that the second term of right-hand side of the second equation is zero, for all beliefs $b_{\theta_i} \in \mathcal{B}^{id}$. Hence, the first- and second-order conditions are met in (d, t) if and only if it they are met in the canonical direct mechanism. This proves the following results:

Theorem 1 (Partial Implementation: Characterization)**.** *Under the maintained assumptions,*

1. *d is partially $\mathcal{B}^{id}$-implementable if and only if it is partially $\mathcal{B}^{id}$-implemented by t^*.*

2. *(a) If the allocation rule d is partially $\mathcal{B}^{id}$-implementable, then:*
   $E^{b_{\theta_i}}\left(\partial^2 U_i^*\left(\theta_i, \theta_{-i}; \theta_i, \theta_{-i}\right)/\partial^2 m_i\right) \leq 0$ for all i, θ_i, and for all $b_{\theta_i} \in B^{id}_{\theta_i}$.

   (b) If $E^{b_{\theta_i}}\left(\partial^2 U_i^*\left(\theta_i, \theta_{-i}; \theta_i, \theta_{-i}\right)/\partial^2 m_i\right) < 0$ for all i, θ_i and for all $b_{\theta_i} \in B^{id}_{\theta_i}$, then d is partially $\mathcal{B}^{id}$-implementable.

This is our main result on partial implementation. It shows that, under the $\mathcal{B}^{id}$-restrictions, there is no reason to consider transfers other than the canonical ones. (As we will see, this will not be the case for full implementation: full implementation may fail under the canonical transfers, but be achieved by other transfers). Besides its intrinsic interest, this result also simplifies the task of identifying which conditions on the environment are necessary or sufficient for partial implementation: it suffices to study properties of the payoff functions induced by the canonical mechanism, $U_i^*\left(m; \theta\right)$, which only depend on the allocation function and on the agents' preferences. Since, by construction, the canonical transfers satisfy the first-order conditions, sufficiency hinges on the second-order conditions of agents' optimization problem at the truthful profile.

Note that, if the expectation operators were removed from these conditions, so that the second-order conditions are satisfied in the ex-post sense, then these conditions would correspond to ep-IC. It is clear, however, that there is a gap between the two: it may be that the canonical mechanism

(d, t^*) satisfies the second-order conditions in expectation, for all beliefs consistent with the $\mathcal{B}^{id}$ (as in part 2 of Theorem 1), but not in the ex-post sense. This clarifies that the result in Theorem 1 does not imply that $\mathcal{B}^{id}$-IC is possible if and only if ep-IC is possible, but only that in both cases it suffices to consider the same mechanism.

3.2 Incentive Compatible Transfers via Moment Conditions

Further intuition on the belief-based components in Lemma 2 can be gathered by looking at the special case in which the K_i function can be written as $K_i(m) = L_i(m_{-i}) - f_i(m_i)$, for some $L_i : \Theta_{-i} \to \mathbb{R}$ and $f_i : \Theta_i \to \mathbb{R}$. In this case, the expected value condition (8) can be written as

$$\int_{\Theta_{-i}} L_i(\theta_{-i})\, db_{\theta_i} = f_i(\theta_i) \text{ for all } \theta_i \text{ and for all } b_{\theta_i} \in B^{id}_{\theta_i}. \tag{9}$$

If a collection $(L_i, f_i)_{i \in I}$ of functions $L_i : \Theta_{-i} \to \mathbb{R}$ and $f_i : \Theta_i \to \mathbb{R}$ satisfies (9) for every i, then it means that under the belief restrictions $\mathcal{B}^{id}$, agents commonly believe that, for every i, his expectation of moment $L_i(\theta_{-i})$ of others' types varies with θ_i according to f_i. Hence, this condition expresses commonly known assumptions on agents' conditional expectations on a moment of others' types. Based on this observation, Ollár and Penta (2017) introduced the following notion:

Definition 3. *A moment condition is represented by a collection $(L_i, f_i)_{i \in I}$ such that $L_i : \Theta_{-i} \to \mathbb{R}$ and $f_i : \Theta_i \to \mathbb{R}$. It is* consistent with $\mathcal{B}^{id}$ *if it satisfies (9) for all i; it is a* linear moment condition *if L_i is linear for every i.*

Setting $K_i(\theta) = L_i(\theta_{-i}) - f_i(\theta_i)$ in the statement of Lemma 2, eq.(7) specializes to

$$t_i(m) = \underbrace{t_i^*(m) + \tau_i(m_{-i})}_{\substack{\text{characterization} \\ \text{of ep-IC transfers}}} + \underbrace{L_i(m_{-i})\, m_i - \int^{m_i} f_i(s_i)\, ds_i}_{\text{moment condition-based term}}. \tag{10}$$

This is precisely the class of transfers for which Ollár and Penta (2017) provide sufficient conditions for full implementation.[12] By Lemma 2, there may exist incentive compatible transfers which cannot be written as in equation (10), since not all functions $K_i : \Theta \to \mathbb{R}$ in that Lemma are equivalent to moment conditions in the sense of Definition 3. Nonetheless, understanding the set of moment conditions which are commonly known under given belief restrictions is a useful way of looking at the possibilities that the designer has to device incentive compatible transfers under these easy-to-interpret belief-based components. Being concerned with full implementation under general belief restrictions, and particularly on sufficient conditions, Ollár and Penta (2017) did not characterize the set of available moment conditions. That task can be difficult in general, but such a characterization is possible for the belief restrictions considered in this paper, and it provides particularly clean insights into the set of transfers which are available to the designer:

Lemma 3 (Moment Conditions under $\mathcal{B}^{id}$: Characterization)**.** *The moment condition $(L_i, f_i)_{i \in I}$ is consistent with $\mathcal{B}^{id}$ if and only if*

[12] In particular, Ollár and Penta (2017) show that if the belief-restrictions admit moment conditions with certain properties, then this design strategy ensures full implementation. They also illustrate the usefulness of those sufficient conditions in common prior environments and in settings in which only the conditional averages are common knowledge. (Note that, under the $\mathcal{B}^{id}$ restrictions we consider in this paper, the conditional averages of types are neither common knowledge nor known to the designer.)

1. *$f_i(\theta_i) = c$ for some $c \in \mathbb{R}$, for all θ_i;*

2. *L_i is constant at identical types and agrees with c: $L_i(\theta) = c$ for all θ s.t. $\theta_i = \theta_j$ for all i, j;*

3. *L_i is additively separable across players: there exist real functions L_{ij} such that $L_i(\theta_{-i}) = \sum_{j \neq i} L_{ij}(\theta_j)$ for all $\theta_{-i} \in \Theta_{-i}$.*

An interesting question is how our analysis would change if, beyond common knowledge of identicality, one also assumed common knowledge of independence across different players. This can be formalized by replacing the B^{id}-restrictions with the stronger belief restrictions $\mathcal{B}^{iid}$, which also require beliefs $b_{\theta_i} \in \Delta(\Theta_{-i})$ in condition (1) to be the independent product of an identical distribution over $[\underline{\theta}, \overline{\theta}]$. It can be shown that results analogous to Lemma 2 obtain for $\mathcal{B}^{iid}$-restrictions, as well as a characterization analogous to Lemma 3, with the only difference that part 3 of Lemma 3 is not required. Intuitively, the stronger information that the designer has about agents beliefs in $\mathcal{B}^{iid}$, compared to $\mathcal{B}^{id}$, allows a richer set of moment conditions which can be used to design incentive compatible transfers. Interestingly, however, such extra freedom does not really expand the possibility of implementation: it can be shown that, under the $\mathcal{B}^{iid}$-restrictions, the characterizations of both partial and full implementation is the same as in Theorems 1 and 2.

4 Full Implementation

Our notion of full implementation is based on the solution concept of $\mathcal{B}^{id}$-rationalizability, a special case of Battigalli and Siniscalchi (2003)'s Δ-rationalizability.[13] $\mathcal{B}^{id}$-rationalizability is defined by an iterated deletion procedure in which, for each type θ_i, a report survives the k-th round of deletion if and only if it can be justified by conjectures (joint distributions over opponents' types and strategies) which are consistent with the belief restrictions $\mathcal{B}^{id}$, and with the previous rounds of the deletion procedure. For every i and θ_i, the set of conjectures that are consistent with common belief in identicality is defined as $C^{id}_{\theta_i} := \left\{\mu_i \in \Delta(M_{-i} \times \Theta_{-i}) : marg_{\Theta_{-i}}\mu_i \in B^{id}_{\theta_i}\right\}$.

Definition 4 ($\mathcal{B}^{id}$-rationalizability)**.** *Fix a direct mechanism. For every $i \in I$, let $R_i^{id,0} = \Theta_i \times M_i$ and for each $k = 1, 2, ...$, let $R_{-i}^{id,k-1} = \times_{j \neq i} R_j^{id,k-1}$,*

$$R_i^{id,k} = \left\{(\theta_i, m_i) : m_i \in BR_{\theta_i}(\mu_i) \text{ for some } \mu_i \in C^{id}_{\theta_i} \cap \Delta\left(R_{-i}^{id,k-1}\right)\right\}, \text{ and } R_i^{id} = \bigcap_{k \geq 0} R_i^{id,k}.$$

The set of $\mathcal{B}^{id}$-rationalizable messages for type θ_i is defined as $R_i^{id}(\theta_i) := \left\{m_i : (\theta_i, m_i) \in R_i^{id}\right\}$.

Definition 5 (Full Implementation)**.** *The transfer scheme $t = (t_i)_{i \in I}$ fully implements d under common knowledge of identicality if $R_i^{id}(\theta_i) = \{\theta_i\}$ for all θ_i and all i.*[14]

First we note that $\mathcal{B}^{id}$-Rationalizability is in general a weak solution concept, and hence our notion of implementation is a demanding one. On the other hand, sufficient conditions for full

[13] Battigalli and Siniscalchi (2003)'s concept allows for general restrictions on players' first-order beliefs on others' types and strategies. Within mechanism design, Ollár and Penta (2017) focused on the case in which belief restrictions are only on others' types; Lipnowski and Sadler (2017) instead adopted restrictions on beliefs about others' behavior for their concept of peer-confirming equilibrium, although not in an implementation setting.

[14] A weaker notion of implementability would allow non-truthful reports, provided that they all induce the same allocation as the true type profile. It can be shown that the two notions coincide for responsive allocation rules.

$\mathcal{B}^{id}$-implementation guarantee full implementation with respect to any (non-empty) refinement of $\mathcal{B}^{id}$-Rationalizability, and hence the weakness of the solution concept strengthens our results. Furthermore, it can be shown that $\mathcal{B}^{id}$-rationalizability characterizes the set of all Bayes-Nash equilibrium strategies, taking the union over all type spaces that are consistent with $\mathcal{B}^{id}$. Definition 5 therefore can be seen as a shortcut to analyze standard questions of Bayesian implementation for all beliefs consistent with the $\mathcal{B}^{id}$ restrictions.[15]

Second, our notion of implementation requires that truthful revelation is the only rationalizable report in the direct mechanism, given the belief restrictions $\mathcal{B}^{id}$. As it is well-known, insisting on direct mechanisms does make the task of achieving full implementation harder. We should thus expect our characterization results to be in general more demanding than those which would be obtained without restricting the space of mechanisms in this way. But there are many advantages from restricting the space of mechanisms. From a conceptual viewpoint, classical results on full implementation typically involve unrealistically complicated mechanisms, which have been criticized for instance in an influential paper by Jackson (1992). Our insistence on implementation via transfer schemes which only elicit payoff-relevant information imposes a more demanding criterion to favor the interpretability of the results, and can thus be seen as pushing Jackson's concern for 'relevance' a bit further.[16] Our approach also allows an easier comparison with the partial implementation literature, since it makes transparent what features of an incentive compatible transfer scheme may be problematic for full implementation, and shows how failure to achieve full implementation can be overcome without changing the fundamental structure of the mechanism.

For later reference, we introduce a class of environments which satisfy a standard single-crossing condition, and in which the concavity of agents' valuation functions is public information:

Definition 6 (SC-PC)**.** *An environment satisfies single crossing and public concavity (SC-PC) if:*

1. For all i and (x,θ), $\frac{\partial^2 v_i}{\partial x \partial \theta_i}(x,\theta) > 0$ and $\partial d/\partial \theta_i > 0$
2. For all i and j, $\partial^2 v_i/\partial^2 x$ and $\partial^2 v_i/\partial x \partial \theta_j$ are constant in θ, and $\frac{\partial^2 d}{\partial \theta_i \partial \theta_j}(\theta) = 0$ for all θ.

These conditions generalize properties of standard quadratic-linear environments with single crossing preferences, which are common both in the theoretical and in the empirical literature for the convenient property that they imply linear best replies. Special cases of our conditions are common in models of social interactions, markets with network externalities, supply function competition, divisible good auctions, markets with adverse selection, provision of public goods.[17] Compared to these applications, Definition 6 also accommodates more general dependence on x, as long as the concavity and the cross derivatives are public information.

[15] By the same arguments, Bergemann and Morris (2009a) and Ollár and Penta (2017) study full implementation, respectively in belief-free settings and under general belief-restrictions, using corresponding versions of Δ-rationalizability. (For earlier versions of these results on Δ-rationalizability, see Battigalli and Siniscalchi (2003).)

[16] In contrast to our static constructions, the more complex or multistage mechanisms suggested by the full implementation literature (e.g., Maskin (1999); Moore and Repullo (1988)) are fragile to perturbations of information (cf. Aghion et al. (2012)). Bergemann and Morris (2009a) also restrict the analysis to direct mechanisms, but in the special case of quasilinear environments considered in this paper, their results only provide conditions to check whether a given mechanism (d,t) is (belief-free) implementable, but they do not provide insights on how transfers should be designed to achieve implementation, if at all possible.

[17] Quadratic-linear models are frequent in the literature of networks (e.g., Ballester et al. (2006), Bramoulle and Kranton (2007), Bramoulle et al. (2014), Galeotti, Golub and Goyal (2019)), social interactions models (Blume et al., (2015)), markets with network externalities (e.g., Fainmesser et al., (2015)), divisible good auctions (e.g., Wilson (1979)) and public goods (e.g., Duggan and Roberts (2002)).

The important consequence of these assumptions is that, in the canonical direct mechanisms, all second order derivatives $\frac{\partial^2 U_i^*}{\partial m_i \partial m_j} = -\frac{\partial^2 v_i}{\partial x \partial \theta_j} \cdot \frac{\partial d}{\partial \theta_i}$ are constant in (θ, m) and s.t. $\partial^2 U_i^*/\partial^2 m_i \neq 0$. We can thus define the (normalized) *canonical externalities* as real numbers $\xi_{ij} := \frac{\partial^2 U_i^*/\partial m_i \partial m_j}{\partial^2 U_i^*/\partial^2 m_i}$. For each i, let $\xi_i := \sum_{j \neq i} \xi_{ij}$, and relabel agents if necessary so that $|\xi_1| \leq |\xi_2| \leq \ldots \leq |\xi_n|$.

4.1 Redistribution of Strategic Externalities

In order to achieve full $\mathcal{B}^{id}$-implementation, the truthful profile must be a mutual (strict) best response for all types θ_i and for all beliefs $b_{\theta_i} \in \Delta(\Theta_{-i})$. Strict $\mathcal{B}^{id}$-IC therefore is a necessary condition for full $\mathcal{B}^{id}$-implementation. It follows that t fully $\mathcal{B}^{id}$-implements d only if it satisfies the expected value condition (8) of Lemma 2. Beyond this partial implementation requirement, however, we will show that full implementation imposes more stringent restrictions on the mechanism, and specifically on the strategic externalities that it induces.

To this end, for any transfer scheme t, and for every $(m, \theta) \in M \times \Theta$, we define the *strategic externality matrix*, $SE^t(m, \theta) \in \bar{\mathbb{R}}^{n \times n}$, in which the entry in row i and column j is equal to $SE^t(m, \theta)_{ij} = \frac{\partial^2 U_i^t(m,\theta)/\partial m_i \partial m_j}{\partial^2 U_i^t(m,\theta)/\partial^2 m_i} \in \bar{\mathbb{R}}$ if $i \neq j$ and $SE^t_{ij} = 0$ if $i = j$. (Recall that $U_i^t(m, \theta)$ denotes i's payoff function induced by transfers t.) For example, in SC-PC settings, the canonical transfers t^* induce the following matrix of strategic externalities: for all (m, θ),

$$SE^*(m, \theta) = \begin{bmatrix} 0 & \xi_{12} & \ldots & \xi_{1n} \\ \xi_{21} & 0 & \ldots & \xi_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \xi_{n1} & \xi_{n2} & \ldots & 0 \end{bmatrix}.$$

The next result shows that strategic externalities are key for the design of transfers for full implementation. In particular, it shows that whether a $\mathcal{B}^{id}$-IC transfer scheme t also achieves full implementation, depends on the properties of a matrix which is closely related to $SE^t(m, \theta)$. Such a matrix is obtained by focusing on the absolute values of the largest externalities across the domain, normalized by the smallest concavity in the domain. Formally, let $|SE^t(m, \theta)|$ be such that $|SE^t|_{ii} = 0$ for each i and $|SE^t|_{ij} := \frac{\max_{(m,\theta) \in \Theta \times \Theta} |\partial^2 U_i^t(m,\theta)/\partial m_i \partial m_j|}{\min_{(m,\theta) \in \Theta \times \Theta} |\partial^2 U_i^t(m,\theta)/\partial^2 m_i|}$ for each i and $j \neq i$.

Lemma 4 (Eigenvalues and Full $\mathcal{B}^{id}$-Implementation)**.** *A $\mathcal{B}^{id}$-IC direct mechanism (d, t) achieves full $\mathcal{B}^{id}$-implementation if all the eigenvalues of $|SE^t|$ are smaller than one in absolute value. This condition is also necessary in SC-PC environments, when t is based on a linear moment condition.*

This lemma states that the key condition for a $\mathcal{B}^{id}$-IC transfer scheme to achieve full implementation is that the spectral radius of the associated matrix of strategic externalities is less than one. Intuitively, the reason is that eigenvalues in general describe the properties of the iterated of a matrix. In the case of strategic externality matrices, this amounts to describing the iteration of best replies which are implicit in the rationalizability operator. The condition that the spectral radius is smaller than one is sufficient to induce contractive best replies, and hence a unique rationalizable probile.[18] Incentive Compatibility – which is assumed in the Lemma – in turn ensures that such a unique profile is actually the truthful revelation profile.

[18]The *spectral radius* of a square matrix is the largest absolute value of its eigenvalues. Other known conditions in the literature, such as diagonal dominance, are easier to check but only sufficient for contractiveness.

Since, as discussed, $\mathcal{B}^{id}$-IC is a necessary condition for full $\mathcal{B}^{id}$-implementation, at a minimum any transfer scheme which achieves full implementation needs to satisfy all the conditions we discussed in Section 3. Note, however, that those conditions only restricted the first and second-order conditions of the agents' optimization problem at the truthful profile (which is all that matters for partial implementation). In contrast, the result in Lemma 4 refers to properties of the $|SE^t|$-matrix, which in turn depend on the properties of the strategic externalities at all profiles (m, θ). Hence, before being able to apply the design strategy suggested by Lemma 4, it is important to first understand which restrictions, if any, are imposed by $\mathcal{B}^{id}$-IC on the strategic externalities at the non-truthful profiles. This is the objective of the next Lemma:

Lemma 5. *If (d, t) is $\mathcal{B}^{id}$-IC, then, for all θ and $(m_i, \bar{m}_{-i})$ s.t. $\bar{m}_j = \bar{m}_k$ for all $j, k \neq i$:*

1. $\partial^2 U_i\left(m_i, \bar{m}_{-i}; \theta\right) / \partial^2 m_i = \partial^2 U_i^*\left(m_i, \bar{m}_{-i}; \theta\right) / \partial^2 m_i$ *and*
2. $\sum_{j \neq i} \partial^2 U_i\left(m_i, \bar{m}_{-i}; \theta\right) / \partial m_i \partial m_j = \sum_{j \neq i} \partial^2 U_i^*\left(m_i, \bar{m}_{-i}; \theta\right) / \partial m_i \partial m_j$.

These conditions are also sufficient in SC-PC, when t is based on a linear moment condition.

In words, these conditions say that for any agent i and for any state θ, at any profile in which i's opponents report (not necessarily truthfully) the same type, then both the concavity in own-action (condition 1), and the sum of the strategic externalities of all the opponents (condition 2), induced by any $\mathcal{B}^{id}$-IC transfer scheme, must be the same as those of the canonical direct mechanism.[19]

Hence, the overall design strategy that emerges from combining these two lemmas is that the designer should seek to minimize the absolute value of the eigenvalues of the $|SE^t|$-matrix, subject to the constraints imposed by $\mathcal{B}^{id}$-IC (and, particularly, by Lemma 5). Such constraints imply that the designer may only *redistribute*, not reduce, the total strategic externalities induced by the canonical direct mechanism. When the conditions in Lemmas 4 and 5 are both necessary and sufficient (as is the case in SC-PC environments and transfers based on linear moment conditions), such a design strategy is characterized by a particularly tight mathematical structure:

Proposition 1 (Redistribution of Strategic Externalities)**.** *Consider an SC-PC environment, and let t be a transfer scheme based on a linear moment condition. Then, (d, t) achieves full $\mathcal{B}^{id}$-implementation if and only if it satisfies the following conditions:*

1. *All the eigenvalues of the $|SE^t|$-matrix are smaller than one in absolute value.*
2. $\partial^2 U_i^t / \partial^2 m_i = \partial^2 U_i^* / \partial^2 m_i$ *and* $\sum_{j \neq i} \partial^2 U_i^t / \partial m_i \partial m_j = \sum_{j \neq i} \partial^2 U_i^* / \partial m_i \partial m_j$.

Note that, since in SC-PC environments the matrix SE^* of strategic externalities of the canonical direct mechanism is constant in (m, θ), the conditions in points 1 and 2 above essentially require that, in order to preserve $\mathcal{B}^{id}$-IC, a transfer scheme should induce a matrix of strategic externalities which preserves, row by row, the same row-sums as in the SE^*-matrix.

[19] The proof of Lemma 5 follows from Lemma 2, and on a more involved characterization of the K_i functions which satisfy the expected value condition (8), which must be such that $K_i(m) = \sum_{k=0}^{\infty} m_i^k \sum_{j \neq i} H_{ij}^k(m_j)$, where $\{H_{ij}^k\}_{j \neq i, k \in \mathbb{N}}$ are polynomials $H_{ij}^k : M_j \to \mathbb{R}$ such that $\sum_{j \neq i} H_{ij}^k(m_j) = 0$ whenever $m_l = m_j$ for all $j, l \neq i$.

4.2 Full Implementation via Transfers: Characterization

In this section we restrict attention to SC-PC environments, which as discussed are especially important from the viewpoint of the applied theoretical literature. Similar to what we did for partial implementation, we seek to identify a transfer scheme $\hat{t}$ which can be used to identify whether or not full $\mathcal{B}^{id}$-Implementation is possible. Intuitively, because of the characterization in Proposition 1, such a transfer scheme should minimize the spectral radius of the associated $|SE|$-matrix, within the set of all $\mathcal{B}^{id}$-IC transfer schemes, i.e. subject to preserving the same row-sums as in the SE^*-matrix: for such a transfer scheme, if the spectral radius is larger than one, then full $\mathcal{B}^{id}$-implementation would be impossible, because any $\mathcal{B}^{id}$-IC transfer scheme would have at least as high a spectral radius; on the other hand, if the spectral radius of the $|SE^{\hat{t}}|$-matrix is smaller than one, then full implementation is possible, and it is achieved by $\hat{t}$.

We define next the *loading transfers*, which will be shown to provide the answer to this question. As illustrated in Example 3, the logic of the construction is to redistribute the strategic externalities so that, in the resulting mechanism, they are all concentrated on the two agents with the smallest canonical externalities (given the relabeling above, these are agents 1 and 2). Formally, the *loading transfers* $\left(t_i^l\right)_{i\in I}$ are defined as follows: for each $i \in I$ and $m \in M_i \times M_{-i}$,

$$t_i^l(m) = \underbrace{t_i^*(m)}_{\text{canonical transfers}} + \underbrace{L_i^l(m_{-i})\, m_i}_{\substack{\text{redistribution of} \\ \text{canonical externalities}}}, \tag{11}$$

where $L_i^l : M_{-i} \to \mathbb{R}$ is such that

$$L_i^l(m_{-i}) = \begin{cases} \left[-\sum_{\substack{k\neq 1 \\ k\neq 2}} \frac{\partial^2 v_1}{\partial x \partial \theta_k} m_2 + \sum_{\substack{k\neq 1 \\ k\neq 2}} \frac{\partial^2 v_1}{\partial x \partial \theta_k} m_k\right] \frac{\partial d}{\partial \theta_1} \text{ if } i = 1 \\ \left[-\sum_{\substack{k\neq 1 \\ k\neq j}} \frac{\partial^2 v_j}{\partial x \partial \theta_k} m_1 + \sum_{\substack{k\neq 1 \\ k\neq j}} \frac{\partial^2 v_j}{\partial x \partial \theta_k} m_k\right] \frac{\partial d}{\partial \theta_j} \text{ if } i \neq 1 \end{cases} \tag{12}$$

First note that the transfers in (11) take the form of (10), with $f_i^l(m_i) = 0$. That is, these transfers are based on a *moment condition* in the sense of Definition 3. $\left(L_i^l, f_i^l\right)_{i\in I}$ thus defined satisfies the conditions of Lemma 3, with $c = 0$, and hence it is consistent with $\mathcal{B}^{id}$, which in turn implies that the loading transfers are $\mathcal{B}^{id}$-IC (this follows from Lemma 2, with $K_i(\theta) = L_i^l(\theta) - f_i^l(\theta_i)$, and from the SC-PC assumption, which ensures that the concavity condition in the sufficiency part of Lemma 2 is satisfied).

Second, letting $U_i^l(m;\theta)$ denote the payoff function which results from these transfers, it can be checked that $\partial_{i1}^2 U_i^l = \sum_{j\neq i} \partial_{ij}^2 U_i^*$ for all $i \neq 1$; $\partial_{12}^2 U_1^l = \sum_{j\neq 1} \partial_{1j}^2 U_1^*$ and otherwise $\partial_{ij}^2 U_i^l = 0$. That is, the total canonical externalities are all loaded onto the two agents with the smallest canonical externalities: for all $i \neq 1$, the sum of canonical externalitites for i are all loaded onto agent 1; whereas the sum of canonical externalities for agent 1 are loaded onto 2.

$$SE^l = \begin{bmatrix} 0 & \xi_1 & \dots & 0 \\ \xi_2 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \xi_n & 0 & \dots & 0 \end{bmatrix}.$$

Theorem 2 (Full Implementation: Characterization)**.** *In SC-PC environments:*

1. *d is fully $\mathcal{B}^{id}$-implementable if and only if it is fully $\mathcal{B}^{id}$-implemented by t^l.*

2. *d is fully $\mathcal{B}^{id}$-implementable if and only if the canonical externalities are such that $|\xi_1\xi_2| < 1$.*

Part 1 of the theorem follows from the fact (shown in the proof) that among the class of $\mathcal{B}^{id}$-IC transfers, the loading transfers are those with the smallest associated spectral radius. Hence, as it turns out, the best way of minimizing the spectral radius of the strategic externality matrix, subject to the constraint (imposed by $\mathcal{B}^{id}$-IC) of preserving the same rowsums as in the SE^*-matrix, is to concentrate all the strategic externalities of any player i on the opponent with the smallest $|\xi_j|$: that is on agent 2 for $i = 1$, and on agent 1 for all $i \neq 1$. This is the reason why the loading transfers characterize the possibility of full implementation under the $\mathcal{B}^{id}$-restrictions.

Part 2 follows from the fact that the condition $|\xi_1\xi_2| < 1$ is equivalent to stating that the eigenvalues of $|SE^l|$ are less than one in absolute value, which by Lemma 4 is both necessary and sufficient for these transfers to achieve full implementation (that is because t^l are both $\mathcal{B}^{id}$-IC, and based on a linear moment condition). The condition $|\xi_1\xi_2| < 1$ implies that the possibility of achieving full $\mathcal{B}^{id}$-implementation is completely determined by the canonical externalities of the two agents with the smallest canonical externalities (or, equivalently, by the two agents with the smallest level of preference interdependence).[20] Thus, full implementation is possible if and only if the combined effect of these two agents' canonical externalities are not too large, and that is regardless of the strength of the preference interdependence of others, or of their number.

As we mentioned in the introduction, this result also has interesting implications from a broader market design perspective: for instance, if full implementation cannot be achieved for a set of agents, then all is needed to achieve full implementation is to add to the the set two agents with small preference interdependence. At the extreme, whenever an implementation problem involves at least two agents with private values, or whenever two such agents can be added to the group, then full implementation is possible via a simple direct mechanism.

4.3 Beyond SC-PC Environments

The implementation results of Theorem 2 can also be extended to a larger class of environments, beyond SC-PC. The next proposition shows that, under the $\mathcal{B}^{id}$-restrictions, whenever a preference environment is 'close enough' to an SC-PC environment which satisfies the condition in point 2 of Theorem 2, then full implementation is also possible in the non SC-PC environment, applying the same design principle as in the loading transfers discussed above.

Definition 7 (Approximately SC-PC Environment)**.** *We say that an environment $\mathcal{E} = (d, v)$ is α-close to an SC-PC environment $\mathcal{E}^{PC} = \left(d^{PC}, v^{PC}\right)$, for all i,*

- *the canonical direct mechanism in $\mathcal{E}$ is strictly concave: $\partial^2_{ii} U^*_i(m;\theta) < 0$ for all m and θ,*

- *the externalities in the canonical direct mechanism in $\mathcal{E}$ are α-close to the externalities of $\left(d^{PC}, t^{*,PC}\right)$: $\left|\partial^2_{ij} U^*_i(m;\theta) - \partial^2_{ij} U^{*(PC)}_i(m;\theta)\right| < \alpha$ for all m and θ, for all i and $j \neq i$.*

[20] That is because $|\xi_1\xi_2| < 1$ is equivalent to $|\sum_{j\neq 1} \frac{\partial^2 v_1}{\partial x \partial \theta_j} \cdot \sum_{j\neq 2} \frac{\partial^2 v_2}{\partial x \partial \theta_j}| < \frac{\partial^2 v_1}{\partial x \partial \theta_i} \cdot \frac{\partial^2 v_2}{\partial x \partial \theta_2}$.

The condition in the second point in this definition is satisfied if, for instance: (i) the valuation functions are ϵ-close: for all i and j, $|\frac{\partial^2 v_i}{\partial^2 x}(x,\theta) - \frac{\partial^2 v_i^{PC}}{\partial^2 x}(x,\theta)| < \epsilon$ and $|\frac{\partial^2 v_i}{\partial x \partial \theta_j}(x,\theta) - \frac{\partial^2 v_i^{PC}}{\partial x \partial \theta_j}(x,\theta)| < \epsilon$; and (ii) the allocation rules are ϵ-close: $|\frac{\partial^2 d}{\partial \theta_i \partial \theta_j}(\theta) - \frac{\partial^2 d^{PC}}{\partial \theta_i \partial \theta_j}(\theta)| < \epsilon$ for all i, j and θ.

Theorem 3 (Full $\mathcal{B}^{id}$-Implementation in Approximately SC-PC Environments)**.** *Consider an SC-PC environment $\mathcal{E}^{PC} = \left(d^{PC}, v^{PC}\right)$ which admits full $\mathcal{B}^{id}$-implementation. Then, there exists $\alpha_0 > 0$ such that for all $0 \leq \alpha < \alpha_0$, any environment $\mathcal{E} = (d, v)$ which is α-close to $\mathcal{E}^{PC}$ also admits Full $\mathcal{B}^{id}$-implementation.*

Moreover, there exists $\alpha_0 > 0$ such that for all $0 \leq \alpha < \alpha_0$, in an α-close environment (d, v), the following transfers ensure Full $\mathcal{B}^{id}$-implementation: for each i,

$$t_i(m) = t_i^*(m) + L_i^{l,PC}(m_{-i})\, m_i, \tag{13}$$

where t_i^* denotes the canonical transfers in $\mathcal{E}$, and $L_i^{l,PC}$ is obtained applying (12) to $\mathcal{E}^{PC}$.

This result follows from the fact that the spectral radius of a matrix is a continuous operator, and hence if the general conditions of Lemma 4 hold in an environment, then they also hold in a 'nearby' environment. It follows that the loaded transfers of an SC-PC environment also ensures full implementation in all nearby environments (SC-PC or not).

5 Further Design Strategies for Full Implementation

In this Section we consider alternative transfer schemes to the loading transfers, which have especially relevant structure and properties. The next Lemma provides general sufficient conditions which will be useful for the following discussion:

Lemma 6. *The $\mathcal{B}^{id}$-IC transfer scheme t achieves full $\mathcal{B}^{id}$-implementation if either: (i) it ensures* limited strategic externalities from other agents – *that is, if $\sum_{j \neq i} |SE^t|_{ij} < 1$ for all i; or (ii) it ensures* limited strategic impact on other agents – *that is, if $\sum_{j \neq i} |SE^t|_{ji} < 1$ for all i.*

The condition in the first point of this Lemma resembles the design principle in Ollár and Penta (2017), in that it requires 'not too strong' *strategic externalities*.[21] Formally, it is a row-wise condition on the $|SE^t|$-matrix. The second condition instead is a column-wise restriciton on $|SE^t|$, which can be interpreted as requiring that any agent's *strategic impacts* on others is not too strong. Both claims in the lemma follow from showing that either condition suffices to ensure that the externality matrix satisfies the eigenvalue condition of Lemma 4.

5.1 Beyond Loading: Even Redistribution of Externalities

We introduce next a transfer scheme which, as illustrated by the t^e transfers in Example 3, pursues a uniform redistribution of the strategic externalities. As it will be shown, such alternative design principle is still widely applicable and has desirable robustness properties.

We define the *equal-externality transfers* $t^e = (t_i^e)_{i \in I}$ as follows: for each i and m,

$$t_i^e(m) := \underbrace{t_i^*(m)}_{\text{canonical transfers}} + \underbrace{L_i^e(m_{-i})\, m_i}_{\substack{\text{redistribution of} \\ \text{canonical externalities}}}, \tag{14}$$

[21] Unlike here, the analysis in Ollár and Penta (2017) was limited to transfers based on linear moment conditions.

where $L_i^e : M_{-i} \to \mathbb{R}$ is such that

$$L_i^e(m_{-i}) = \sum_{j \neq i} \left[\left(-\frac{\partial^2 v_i}{\partial x \partial \theta_j} + \frac{1}{n-1} \sum_{k \neq i} \frac{\partial^2 v_i}{\partial x \partial \theta_k} \right) m_j \right] \frac{\partial d}{\partial \theta_i}.$$

Similar to the loading transfers, also these transfers are $\mathcal{B}^{id}$-IC in SC-PC environments, and are based on a linear moment condition in the sense of Definition 3.[22] Moreover, letting $U_i^e(m;\theta)$ denote the payoff function which results from these transfers, we have that $\partial_{ij}^2 U_i^e = \frac{1}{n-1} \sum_{j \neq i} \partial_{ij}^2 U_i^*$ for all i and $j \neq i$, and $\partial_{ii}^2 U_i^e = \partial_{ii}^2 U_i^*$ for all i. Hence, these transfers redistribute the total externalities of the canonical direct mechanism evenly across all of i's opponents. This can be easily seen from the strategic externality matrix they induce:

$$SE^e = \begin{bmatrix} 0 & \frac{\xi_1}{n-1} & \cdots & \cdots & \frac{\xi_1}{n-1} \\ \frac{\xi_2}{n-1} & 0 & \frac{\xi_1}{n-1} & \cdots & \frac{\xi_2}{n-1} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \frac{\xi_2}{n-1} & \cdots & \frac{\xi_1}{n-1} & 0 & \frac{\xi_2}{n-1} \\ \frac{\xi_2}{n-1} & \cdots & \cdots & \frac{\xi_2}{n-1} & 0 \end{bmatrix}.$$

While Theorem 2 ensures that, in SC-PC environments, the loading transfers achieve full $\mathcal{B}^{id}$-implementation whenever such implementation is possible, the next result provides easy-to-check conditions under which full implementation can be achieved via the equal-externality transfers t^e:

Proposition 2. *Under SC-PC, the transfers in (14) achieve full $\mathcal{B}^{id}$-implementation if*

$$\textit{either (i)} \left| \sum_{j \neq i} \frac{\partial^2 v_i}{\partial x \partial \theta_j} \right| < \frac{\partial^2 v_i}{\partial x \partial \theta_i} \textit{ for all } i\textit{; or (ii)} \sum_{j \neq i} \left| \frac{\partial^2 v_j}{\partial x \partial \theta_i} \Big/ \frac{\partial^2 v_j}{\partial x \partial \theta_j} \right| < 1 \textit{ for all } i. \tag{15}$$

The proof of this proposition follows directly from the more general result in Lemma 6. To appreciate the conditions in (15), it is useful to compare them to the following known sufficient for full implementation via the *canonical* transfers: Under SC-PC, the canonical transfers achieve (belief-free) full implementation if

$$\sum_{j \neq i} \left| \frac{\partial^2 v_i}{\partial x \partial \theta_j} \right| < \frac{\partial^2 v_i}{\partial x \partial \theta_i} \text{ for all } i. \tag{16}$$

Condition (16) requires that the sum of preference interdependencies, across all of opponents' of agent i, to be small relative to the dependence of i's marginal utility on his own type. When this condition is satisfied, the resulting strategic externalities in the canonical direct mechanism are small, and belief-free full implementation follows from the results in Ollár and Penta (2017) and Bergemann and Morris (2009a). Relative to this belief-free benchmark, the $\mathcal{B}^{id}$-restrictions enable the designer to redistribute the strategic externalities, and hence to weaken Condition (16) to part (i) of Proposition 2, in which the absolute value is moved outside of the summation. This means that, by relying on the $\mathcal{B}^{id}$-restrictions, preference interdependencies with opposite signs

[22] One way to verify that the t^e transfers are $\mathcal{B}^{id}$-IC is to notice that the associated moment condition satisfies the conditions of Lemma 3, with $c = 0$.

can be leveraged, to obtain full implementation: Under $\mathcal{B}^{id}$, it is the total amount of *net* preference interdependence that matters, not the total amount of *absolute* interdependence.

The second condition in (15) instead focuses on the total *impact* that agent i's type has on other agents' preferences. Rather than pointing at the way player i's preferences depend on others' information, it measures the total impact of i's information on others' preferences. The reason why this alternative condition is also sufficient is related to the idea of *Limited Strategic Impact* which we discussed above and is a consequence of part (ii) of Lemma 6 above.

Example 4. *In the setting of the leading examples, consider preferences $v_i : X \times \Theta \to \mathbb{R}$ which satisfy the following condition:*

$$\left(\frac{\partial^2 v_i}{\partial x \partial \theta_j}\right)_{\substack{i=1,2,3\\ j=1,2,3}} = \begin{bmatrix} 1 & \frac{7}{6} & -\frac{5}{6} \\ -\frac{1}{6} & 1 & \frac{3}{6} \\ -\frac{4}{6} & -\frac{4}{6} & 1 \end{bmatrix}$$

It is immediate to check that condition (16) does not hold, and one can also show that belief-free full implementation is not possible in this setting. In fact, for agent 3 *condition (i) is also violated. Condition (ii), however, holds and implementation via the equal-externality transfers is possible. (Of course, implementation via the loading transfers is possible too.)*

The next result formalizes the sense in which – while still not as applicable as the loading transfers (which, by Theorem 2, achieve full implementation whenever possible) – the logic of the equal-externality transfers is still widely applicable:

Proposition 3. *Under SC-PC, one of the conditions in Lemma 6 are satisfied by* some *$\mathcal{B}^{id}$-IC transfer scheme t , then the equal-externality transfers $(t_i^e)_{i\in I}$ achieve full $\mathcal{B}^{id}$-implementation.*

Corollary 1. *If Condition* (16) *holds, then both t^* and t^e ensure full $\mathcal{B}^{id}$-implementation.*

Hence, whenever there is an implementing transfer scheme which satisfies the easy-to-check conditions of Lemma 6, then the equal-externality transfers t^e also achieve full $\mathcal{B}^{id}$-implementation. There are, however, environments in which the canonical transfers t^* achieve full $\mathcal{B}^{id}$-implementation, but the equal-externality transfers t^e do not:

Example 5. Consider 4 agents and an SC-PC environment for which the canonical direct mechanism and the corresponding balancing transfers induce the following externality matrix:

$$SE^* = SE^l = \begin{bmatrix} 0 & 0.1 & 0 & 0 \\ 0.2 & 0 & 0 & 0 \\ 6 & 0 & 0 & 0 \\ 6 & 0 & 0 & 0 \end{bmatrix} \text{ and } SE^e = \begin{bmatrix} 0 & \frac{1}{30} & \frac{1}{30} & \frac{1}{30} \\ \frac{2}{30} & 0 & \frac{2}{30} & \frac{2}{30} \\ 2 & 2 & 0 & 2 \\ 2 & 2 & 2 & 0 \end{bmatrix},$$

In this example, the $|SE^*|$-matrix has specctral ratio less than 1, however the $|SE^e|$-matrix has an eigenvalue larger than 2. Here the canonical transfers coincide with the loading transfers, and so achieve full implementation, but the equal-externality transfers do not.[23] □

[23]For cases in which, contrary to this example, the canonical transfers fail full implementation but the transfers with uniformly redistributed externalities work well, see Examples 3 and 4.

5.2 Environments with Symmetric Aggregators

Next, we examine full $\mathcal{B}^{id}$-implementability in a special case of our environments, which satisfy a (still weak) assumption of symmetry in agents' preferences. We show that, under this mild assumption of symmetry, transfers with uniformly redistributed externalities are indeed without loss of generality, in the sense that they achieve full implementation whenever it is possible.

Definition 8 (Symmetric Aggregators in Valuations)**.** *An environment has symmetric aggregators in valuations if for all i, there exist $w : X \times \mathbb{R} \to \mathbb{R}$ and $h_i : \Theta \to \mathbb{R}$ strictly increasing in θ_i such that $v_i(x, \theta) = w(x, h_i(\theta))$, $\partial h_i(\theta)/\partial\theta_i = \partial h_j(\theta)/\partial\theta_j$ and $\sum_{k\neq i}(\partial h_i(\theta)/\partial\theta_k) = \sum_{k\neq j}(\partial h_j(\theta)/\partial\theta_k)$ for all i, j and θ.*

Proposition 4 (Full $\mathcal{B}^{id}$-Implementation with Symmetric Aggregators)**.** *Consider an SC-PC environment with symmetric aggregators in valuations.*

1. *Full $\mathcal{B}^{id}$-Implementation is possible if and only if it is achieved by transfers $(t_i^e)_{i\in I}$.*
2. *Full $\mathcal{B}^{id}$-Implementation is possible if and only if $\left|\sum_{k\neq i}\frac{\partial^2 v_i}{\partial x\partial\theta_k}\right| < \frac{\partial^2 v_i}{\partial x\partial\theta_i}$ for all i.*

Next, Figure 1 summarizes the relations between different transfer design strategies. This figure summarizes the implications of the counterexamples above; and the results on full implementability under identical distributions (via the loading transfers in Theorem 2, via designing diagonal dominance in Lemma 6, via the equal-externality transfers in Proposition 2, and in environments with symmetric aggregators in Proposition 4).

6 Sensitivity Results

6.1 *Slightly Faulty* Players and Preference Misspecification

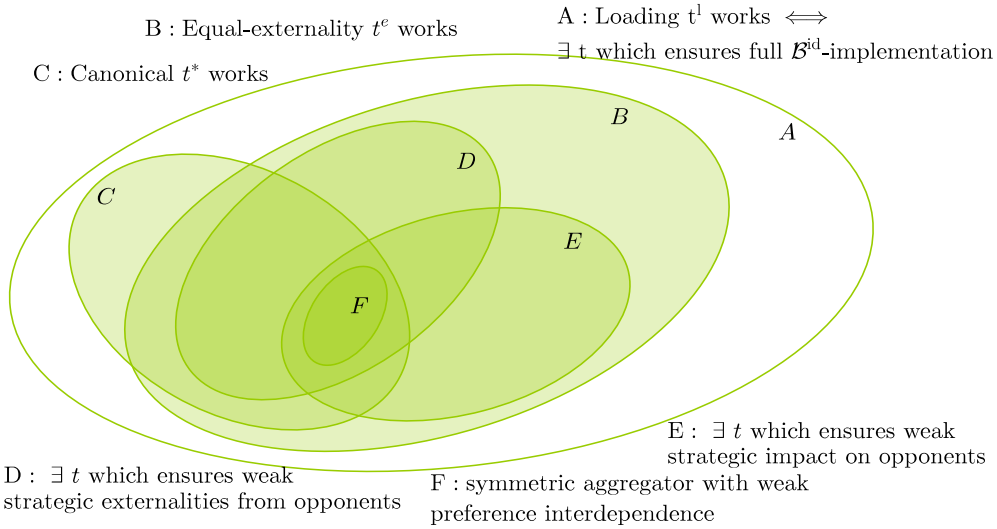
In many settings, it may be desirable to ensure that the implementing mechanism does not rely too heavily on the fact that agents' behavior would coincide exactly with that entailed by the maintained assumptions on their preferences and rationality. In this section we explore the implications of this kind of desiderata on the design of transfers for full implementation, by requiring the implementing mechanism to minimize the impact of an ε-mistake in an agents' report. Such 'mistakes' can be interpreted as either stemming from agents' *slightly faulty* behavior (similar to Eliaz (2002)), or due to a misspecification of agent's preferences in the model.[24]

Formally, let $F \subseteq I$ be an arbitrary set of possibly *slightly faulty* agents, in the sense that they may report messages up to $\varepsilon > 0$ away from their optimal response. For any $\varepsilon > 0$, $\theta_i \in \Theta_i$ and $\mu_i \in \Delta(\Theta_{-i} \times M_{-i})$, let

$$BR^{\varepsilon}_{\theta_i}(\mu_i) = \{m_i \in \Theta_i : |m_i - m_i'| \leq \varepsilon \text{ for some } m_i' \in BR_{\theta_i}(\mu_i)\}.$$

[24] Robustness with respect of the possibility of *slightly faulty* agents is somewhat in the spirit of the analysis in Eliaz (2002). There are, however, several differences: first, Eliaz (2002) considers a complete information environment, whereas our environments features incomplete information and interdependent values; second, in Eliaz (2002)'s model the possibility of agent's mistakes affects the very notion of incentive compatibility, which is strengthened to a condition intermediate between Nash and dominant-strategy incentive compatibility; third, Eliaz (2002) does not restrict the space mechanisms, and in that paper implementation is achieved through a modulo game.

Figure 1: Relations of Environments which Admit Full $\mathcal{B}^{id}$-Implementing Transfer Schemes



This diagram illustrates full implementability via different transfer schemes. The relations between different design strategies are illustrated on sets of environments (pairs of allocation rule and value functions (d, v)), which satisfy single crossing and public concavity (Assumption SC-PC), and which admit full $\mathcal{B}^{id}$-implementation for some transfer scheme.

Definition 9 (F_ε-rationalizability)**.** *Fix a direct mechanism (d,t), $\varepsilon \geq 0$ and $F \subseteq I$. For any $\theta_i \in \Theta_i$ and $\mu_i \in \Delta(\Theta_{-i} \times M_{-i})$, let*

$$BR_{\theta_i}^{F_\varepsilon}(\mu_i) = \begin{cases} BR_{\theta_i}^{\varepsilon}(\mu_i) & \text{if } i \in F \\ BR_{\theta_i}(\mu_i) & \text{if } i \notin F \end{cases}$$

For every $i \in I$, let $R_i^{F_\varepsilon,0} = \Theta_i \times M_i$ and for each $k = 1, 2, ...$, let $R_{-i}^{F_\varepsilon,k-1} = \times_{j \neq i} R_i^{F_\varepsilon,k-1}$,

$$R_i^{F_\varepsilon,k} = \left\{(\theta_i, m_i) : m_i \in BR_{\theta_i}^{F_\varepsilon}(\mu_i) \text{ for some } \mu_i \in C_{\theta_i}^{id} \cap \Delta\left(R_{-i}^{F_\varepsilon,k-1}\right)\right\}, \text{ and } R_i^{F_\varepsilon} = \bigcap_{k \geq 0} R_i^{F_\varepsilon,k}.$$

The set of F_ε-rationalizable messages for type θ_i is defined as $R_i^{F_\varepsilon}(\theta_i) := \left\{m_i : (\theta_i, m_i) \in R_i^{F_\varepsilon}\right\}$.

$R_i^{F_\varepsilon}$ represents our model of strategic interaction when players consider the possibility that agents in F may be ε-faulty. Implicit in this notion is the idea that all agents share a common bound ε on the mistakes which could be made by the faulty agents. That is because, at each iteration of the F_ε-rationalizability procedure, players commonly believe that agents in $I \setminus F$ best-respond to their conjectures, whereas agents in F may play any report which is within ε from their optimal report. The next definition formalizes our notion of robustness to 'slightly faulty' agents:

Definition 10 (Sensitivity to ε-Faulty Agents)**.** *Fix a direct mechanism (d,t). For any $k=1,...,n$, let $N(k) := \{F \subseteq I : |F| = k\}$, and $\eta^t(\varepsilon,k) := \sup_{F\in N(k)} \sup_{i\in I} \sup_{\theta_i\in\Theta_i} \sup_{m_i\in R_i^{F_\varepsilon}(\theta_i)} |m_i - \theta_i|$ be t's sensitivity to k agents who are ε-faulty, and let $\eta^t(\varepsilon) = (\eta^t(\varepsilon,1),...,\eta^t(\varepsilon,n))$*

In words, the idea is that the designer does not know how many or which agents might be potentially faulty, and the criterion with which he/she assesses the robustness of the mechanism is the worst-case scenario across all possible configurations of sets of faulty agents. The measure of the fragility of the mechanism is therefore provided by the largest misreport consistent with $R_i^{F_\varepsilon}$, across all agents and all configurations of the set of faulty agents. The next result shows that, in SC-PC environments with symmetric aggregators, the equal-externality transfers introduced in the previous section are more robust than the loading transfers this sense:

Proposition 5. *[Sensitivity to ε-Faulty Players]Under SC-PC and Symmetric Aggregators, for all $\varepsilon > 0$, $\eta^{t^e}(\varepsilon) \leq \eta^{t^l}(\varepsilon)$, moreover for all $n < k$, $\eta^{t^e}(\varepsilon,k) < \eta^{t^l}(\varepsilon,k)$.*

The intuition behind this result is simple: as explained, the loading transfers induce a very hierarchical strategic structure, in which the contractiveness of the mechanism is completely determined by the two agents with smallest preference interdependence. But loading all strategic externalities on these agents also makes the mechanism especially vulnerable to the possibility that precisely those agents turn out to be faulty. In that case, the loading transfers would perform very poorly. To avoid this risk, and not knowing which of the agents may potentially be faulty, the safest solution for the designer is to redistribute the strategic externalities uniformly across all players, so that no player becomes especially critical for the mechanism.

The proof of Proposition 5 relies on the following technical lemma, which characterizes the set of possible misreports at each iteration of the F_ε-rationalizability procedure:

Lemma 7. *Under SC-PC and linear moment conditions, for given ε and F, the largest set of reports in $R_i^{F_\varepsilon}$ is the largest element of the vector $\left[\ I - |SE^t|\ \right]^{-1} \boldsymbol{\varepsilon}^F$, where $\boldsymbol{\varepsilon}^F = (\varepsilon_i)_{i\in I}$ is the vector such that $\varepsilon_i = \varepsilon$ if $i \in F$ and $\varepsilon_i = 0$ if $i \notin F$.*

6.2 Lower Orders of Rationality and Robust Level-k Implementation

In this Section we consider a different notion of robustness, namely with respect to lower order beliefs in rationality. To this end, it is useful to introduce notation for the set of reports which survive the k-th round of $\mathcal{B}^{id}$-rationalizability (def. 4) for a given type θ_i: $R_i^{id,k}(\theta_i) := \left\{m_i : (\theta_i, m_i) \in R_i^{id,k}\right\}$. To stress the dependence of this set on the specific transfer scheme t, when needed, we will write $R_i^{id,k}(\theta_i|t)$. The properties of the loading transfers discussed in Section 4 – namely, that they maximize the speed of the contraction induced by iterating the best responses, among the class of all $\mathcal{B}^{id}$-IC transfers, – also ensure the following result:

Theorem 4. *Let t be any $\mathcal{B}^{id}$-IC transfer scheme. Then: $R_i^{id,k}\left(\theta_i|t^l\right) \subseteq R_i^{id,k}(\theta_i|t)$ for all k.*

This result is interesting because it points at a different notion of robustness, with respect to lower order beliefs in rationality: the loading transfers are the most efficient at minimizing the possible misreports which could arise due to failures of common belief in rationality. This is an important property, because common belief in rationality (which is implicit in the notion of rationalizability) is often very demanding, and need not be satisfied in a given environment. If

the designer is concerned with agents' sharing lower orders of mutual belief in rationality, then he would not only consider the sets $R_i^{id}(\theta_i)$, but also $\left(R_i^{id,k}(\theta_i)\right)_{k\in\mathbb{N}}$. Hence, among two fully implementing transfers (i.e., both such that $R^{id}(\theta_i)=\{\theta_i\}$), he should prefer the one which also induces the smaller $R_i^{id,k}(\theta_i)$ at every k. The loading tranfers are optimal in this respect.

This notion of robustness is connected to the literature on level-k implementation, which has explicitly considered designing mechanisms for players who don't share common knowledge of rationality. In particular, in an important recent paper, de Clippel et al. (2018) have studied a notion of level-k implementation which, for the class of environments and the direct mechanisms we consider, can be described as follows: Let $p\in\Delta(\Theta)$ denote a common prior, and $\Theta=\times_{i\in I}\Theta_i$. For any direct mechanism (d,t), let Σ_i denote the set of strategies $\sigma_i:\Theta_i\to M_i$ ($M_i=\Theta_i$ in the direct mechanism). Each player is characterized by an anchor, $\alpha_i:\Theta_i\to\Delta(M_i)$, which specifies the message chosen in the mechanism by the non-strategic level-0 type. As usual, let α and α_{-i} denote the profiles of anchors (with independent randomization across players).

de Clippel et al. (2018) introduce the following solution concept for level-k implementation:

$$S_i^1(\alpha)=\left\{\sigma_i\in\Sigma_i:\forall\theta_i,\sigma_i(\theta_i)\in\arg\max_{m_i}\int_{\Theta_{-i}}U_i^t(m_i,\alpha_{-i}(\theta_{-i}),\theta_i,\theta_{-i})\,dp(\theta_{-i}|\theta_i)\right\}$$

$$\forall k\geq 2,\ \ S_i^k(\alpha)=\left\{\sigma_i\in\Sigma_i:\begin{array}{c}\exists\sigma_{-i}\in S_{-i}^{k-1}(\alpha)\ \text{ s.t. }\ \forall\theta_i,\\ \sigma_i(\theta_i)\in\arg\max_{m_i}\int_{\Theta_{-i}}U_i^t(m_i,\alpha_{-i}(\theta_{-i}),\theta_i,\theta_{-i})\,dp(\theta_{-i}|\theta_i)\end{array}\right\}$$

Definition 11 (Level-k Implementation (de Clippel et al. (2018))). *A direct mechanism (d,t) achieves level-k implementation if $S^k(\mu|\alpha)=\{\sigma^*\}$ for every k.*

Compared to the previous literature on level-k implementation, de Clippel et al. (2018)'s notion is more robust in that it doesn't rely on the designer's knowledge of the agents' levels of sophistication: implementation is required to be achieved for all k. Their results are also more general than previous analysis in that they provide results for various anchors α. Their analysis, however, maintains the classical assumption of a commonly known prior $p\in\Delta(\Theta)$.[25] But this notion of implementation can be easily adapted to our belief restrictions, $\mathcal{B}^{id}=((B_{\theta_i}^{id})_{\theta_i\in\Theta_i})_{i\in I}$, by replacing the solution concept in the above definition to the following weaker version:

$$\hat{S}_i^1(\alpha)=\left\{\sigma_i\in\Sigma_i:\begin{array}{c}\forall\theta_i,\exists b_i\in B_{\theta_i}^{id}\\ \sigma_i(\theta_i)\in\arg\max_{m_i}\int_{\Theta_{-i}}U_i^t(m_i,\alpha_{-i}(\theta_{-i}),\theta_i,\theta_{-i})\,db_i(\theta_{-i})\end{array}\right\}$$

$$\forall k\geq 2,\ \ \hat{S}_i^k(\alpha)=\left\{\sigma_i\in\Sigma_i:\begin{array}{c}\exists\sigma_{-i}\in\hat{S}_{-i}^{k-1}(\alpha)\ \text{ s.t. }\ \forall\theta_i,\exists b_i\in B_{\theta_i}^{id}\\ \sigma_i(\theta_i)\in\arg\max_{m_i}\int_{\Theta_{-i}}U_i^t(m_i,\alpha_{-i}(\theta_{-i}),\theta_i,\theta_{-i})\,db_i(\theta_{-i})\end{array}\right\}$$

As de Clippel et al. (2018) remark, the behavioral anchors are completely arbitrary, they may be mechanism specific and may differ across agents. In their setting, however, it is still the case that anchors $\alpha_j:T_j\to M_j$ are common knowledge among the agents, and also known to the designer. A natural strengthening of the implementation requirement would be to allow for different players to have different views about others' anchors, or be uncertain over them, or on others' views about anchors, and so on. And, most importantly, without requiring that the designer knows each

[25] Kneeland (2018) studied level-k implementation both in common prior and belief-free settings. Unlike de Clippel et al. (2018), however, she restricts anchors to be type-independent and equal to the uniform distribution, and she allows different SCFs (selected from a multi-valued social chocie rule) to be implemented for different level-k's.

player's anchor, nor his beliefs about others', at any order. If we let possible anchors in each player i's mind to be any $\alpha_{-i} : T_{-i} \to \Delta(M_{-i})$ – i.e., also allowing for possible correlations – then we obtain the following solution concept for robust level-k implementation:

$$RL_i^1 = \bigcup_{\alpha_{-i} \in \Delta(M_{-i})^{T_{-i}}} \hat{S}_i^1(\alpha) \text{ and}$$

$$\forall k \geq 2, \; RL_i^k = \left\{ \sigma_i \in \Sigma_i : \begin{array}{c} \exists \sigma_{-i} \in RL_{-i}^{k-1} \text{ s.t. } \forall \theta_i, \exists b_i \in B_{\theta_i}^{id} \\ \sigma_i(\theta_i) \in \arg\max_{m_i} \int_{\Theta_{-i}} U_i^t(m_i, \alpha_{-i}(\theta_{-i}), \theta_i, \theta_{-i}) \, db_i(\theta_{-i}) \end{array} \right\}$$

Definition 12 (Robust Level-k $\mathcal{B}^{id}$-Implementation)**.** *A direct mechanism (d,t) achieves robust level-k $\mathcal{B}^{id}$-implementation if $RL^k = \{\sigma^*\}$ for every k.*

It is easy to verify that, for every k, $\sigma_i \in RL_i^k$ if and only if $\sigma_i(\theta_i) \in R_i^{id,k}(\theta_i)$ for every θ_i. Hence, if one wishes to obtain full implementation for every k – i.e., level-k implementation à la de Clippel et al. (2018), but in the much more robust specification for what concerns agents' anchors – then one needs to obtain implementation in $\mathcal{B}$-dominant strategies, because it requires $R_i^{id,1}(\theta_i) = \{\theta_i\}$ for every θ_i. If that can be obtained, as for instance Ollár and Penta (2017) show in SC-PC environments with independent or affiliated common priors, then the result follows for all levels, and hence *interim $\mathcal{B}$-Dominant Strategy Incentive Compatibility* (iDSIC) characterizes this notion of *robust level-k implementation*. But iDSIC is very demanding, and in particular under the $\mathcal{B}^{id}$-restrictions it cannot be satisfied outside of the very special case of private values. It is then natural to ask what is the best that one could obtain, if such stricter notion of implementation cannot be obtained for every k. One possibility is to ensure that, for each k, the $R_i^{id,k}$-sets are as small as possible around the truthful revelation profile. The result in Theorem 4 addresses precisely this question, and implies that the loading transfers introduced above are optimal with respect to this notion of *robust level-k $\mathcal{B}^{id}$-implementation*.

Appendix

Proof of Lemma 2. Assume that t ensures $\mathcal{B}^{id}$-incentive compatibility which, by t's differentiability, means that for all i and θ_i

$$\int_{\Theta_{-i}} \frac{\partial \left(v_i(d(m_i, \theta_{-i}), \theta) + t_i(m_i, \theta_{-i}) \right)}{\partial m_i} db_{\theta_i} \Big|_{m_i = \theta_i} = 0 \text{ for all } b_{\theta_i} \in B_{\theta_i}^{id},$$

or rearranged,

$$\int_{\Theta_{-i}} \frac{\partial t_i(m_i, \theta_{-i})}{\partial m_i} db_{\theta_i} \Big|_{m_i = \theta_i} = - \int_{\Theta_{-i}} \frac{\partial v_i(d(m_i, \theta_{-i}), \theta)}{\partial m_i} db_{\theta_i} \Big|_{m_i = \theta_i} \text{ for all } b_{\theta_i} \in B_{\theta_i}^{id}.$$

The canonical transfer t_i^* also satisfies this equation, thus for the difference between t_i and t_i^*,

$$\mathbb{E}^{b_{\theta_i}} \left(\frac{\partial}{\partial m_i} \left(t_i(m_i, \theta_{-i}) - t_i^*(m_i, \theta_{-i}) \right) \right) \Big|_{m_i = \theta_i} = 0 \text{ for all } (\theta_i, b_{\theta_i}) : b_{\theta_i} \in B_{\theta_i}^{id}.$$

Let the difference between t_i and t_i^* be $D_i(m) := t_i(m) - t_i^*(m)$. By the smoothness assumptions of this Lemma, D_i is differentiable. Consider the part of D_i that is independent from m

and let this part be d_0; let its part that is independent from m_i be $\tau_i(m_{-i}) := D_i(m) - d_0 - \int_{\underline{\theta}_i}^{m_i} \frac{\partial D_i}{\partial m_i}(s_i, m_{-i})\, ds_i$, and further let $G_i(m) := D_i(m) - \tau_i(m_{-i})$ for all m.

Then, the previous displayed equation can be rewritten such that for all fixed θ_i

$$\mathbb{E}^{b_{\theta_i}}\left(\frac{\partial G_i}{\partial m_i}(\theta_i, \theta_{-i})\right) = 0 \text{ for all } b_{\theta_i} \in B^{id}_{\theta_i}.$$

By the definition of $D_i(m)$ and by setting $K_i(m_i, m_{-i}) := \frac{\partial G_i}{\partial m_i}(m_i, m_{-i})$, the transfer t_i takes the form $t_i(m) = t_i^*(m) + \tau_i(m_{-i}) + \int_{\underline{\theta}_i}^{m_i} K_i(s_i, m_{-i})\, ds_i$ for all m. Moreover, if (d,t) is twice differentiable, then by the definition of canonical transfers t^* is twice differentiable too, and therefore K_i is differentiable. Since K_i is differentiable in all its arguments, τ_i is differentiable too, which completes the proof of the necessity part of this Lemma.

If (d,t) is twice differentiable and t satisfies the characterization in (7) and the expected value condition in (8), then

$$\begin{aligned}
E^{b_{\theta_i}}\left(\frac{\partial U_i}{\partial m_i}(\theta;\theta)\right) &= E^{b_{\theta_i}}\left(\frac{\partial v_i}{\partial m_i}(\theta;\theta) + \frac{\partial t_i}{\partial m_i}(\theta;\theta)\right) \\
&= E^{b_{\theta_i}}\left(\frac{\partial v_i}{\partial m_i}(\theta;\theta) + \frac{\partial t_i^*}{\partial m_i}(\theta;\theta)\right) + 0 + E^{b_{\theta_i}}\left(K_i(\theta;\theta)\right) \\
&= E^{b_{\theta_i}}\left(\frac{\partial v_i}{\partial m_i}(\theta;\theta) - \frac{\partial v_i}{\partial m_i}(\theta;\theta)\right) + 0 + 0 = 0,
\end{aligned}$$

and thus the message $m_i = \theta_i$ is an extreme point. For all beliefs in $B^{id}_{\theta_i}$, the corresponding expected utility, by assumption, is strictly concave, therefore this extreme point is a global optimum for all beliefs in $B^{id}_{\theta_i}$, and thus (d,t) is $\mathcal{B}^{id}$-IC which completes the proof of the sufficiency part of this Lemma. ■

Proof of Theorem 1.

Step 1: (Properties of K_is for Partial Implementation) If $K_i : M \to \mathbb{R}$ satisfies condition (8) in Lemma 2, then it's derivative satisfies that $E^{b_{\theta_i}}\left(\partial K_i(\theta_i, \theta_{-i})/\partial m_i\right) = 0$ for all θ_i and for all $b_{\theta_i} \in B^{id}_{\theta_i}$.

To show this step, recall the expected value condition in Equation 8, in Lemma 2, that is, $\mathbb{E}^{b_{\theta_i}}\left(K_i(\theta_i, \theta_{-i})\right) = 0$ for all θ_i and for all $b_{\theta_i} \in B^{id}_{\theta_i}$. Fix $p \in B^{id}_{\theta_i}$. It is a consequence of identicality that if $p \in B^{id}_{\theta_i}$, then $p \in B^{id}_{\theta_i'}$ for all $\theta_i' \in [\underline{\theta}, \overline{\theta}]$, that is $\mathbb{E}^p\left(K_i(\theta_i, \theta_{-i})\right) \equiv 0$, which then implies that for all θ_i, $\mathbb{E}^p\left(\partial K_i(\theta_i, \theta_{-i})/\partial m_i\right) = 0$, and this holds for any $p \in B^{id}_{\theta_i}$, which completes the proof of this Step. □

Step 2: If (d,t) partially implements d, then by Lemma 2, t can be written as in (7), and hence – letting U^* denote the payoff function of the canonical direct mechanism – for any θ_i and $b_{\theta_i} \in \mathcal{B}^{id}$ we have:

$$\begin{aligned}
&E^{b_{\theta_i}}\left(\partial U_i(\theta_i, \theta_{-i}; \theta_i, \theta_{-i})/\partial m_i\right) = E^{b_{\theta_i}}\left(\partial U_i^*(\theta_i, \theta_{-i}; \theta_i, \theta_{-i})/\partial m_i\right) + E^{b_{\theta_i}}\left(K_i(\theta_i, \theta_{-i})\right), \text{ and} \\
&E^{b_{\theta_i}}\left(\partial^2 U_i(\theta_i, \theta_{-i}; \theta_i, \theta_{-i})/\partial^2 m_i\right) = E^{b_{\theta_i}}\left(\partial^2 U_i^*(\theta_i, \theta_{-i}; \theta_i, \theta_{-i})/\partial^2 m_i\right) + E^{b_{\theta_i}}\left(\partial K_i(\theta_i, \theta_{-i})/\partial m_i\right).
\end{aligned}$$

Now, condition (8) and Step 1, respectively, imply that the second terms on the right-hand side of both equations are zero. Hence, the first- and second-order conditions of the agent's optimization problem, for all beliefs in $\mathcal{B}^{id}$ are equivalent in (d,t) and (d,t^*), which prooves this Theorem. ■

Proof of Lemma 3. Setting $K_i := L_i - f_i$ in Step 1 of the Proof of Theorem 2 below, which gives the characterization of $\mathcal{B}^{id}$-consistent K_i functions, implies this Lemma. ■

Proof of Lemma 4. PART 1. Fix θ_i *in* $\left(\underline{\theta}, \overline{\theta}\right)$ and examine the k-th round of eliminations: fix $m_i \in R_i^k(\theta_i)$. Since (d,t) is $\mathcal{B}^{id} - IC$, θ_i is best-reply to truthtelling conjectures. In particular, consider a truthtelling conjecture which is concentrated on R_{-i}^{k-1}, let this conjecture be μ_T. For m_i too, there exists a conjecture which supports m_i as a best reply and is concentrated on R_{-i}^{k-1}. Let this conjecture be μ_L.

Note that any k-th-round best-reply m_i is either inner point or a boundary point. Let $b_l \leq b_u$ be the boundary points of the set of $k-1$-rationalizable messages of θ_i. Let $E^{\mu}U_i(m_i;\theta_i)$ denote the expected utility of type θ_i, given this type's conjecture μ, when reporting m_i.

First, if m_i is an interior point, then we have that

$$0 = \partial_i E^{\mu_L} U_i(m_i;\theta_i) - \partial_i E^{\mu_T} U_i(\theta_i;\theta_i)$$

$$= \underbrace{\partial_i E^{\mu_L} U_i(m_i;\theta_i) - \partial_i E^{\mu_L} U_i(\theta_i;\theta_i)}_{\text{difference due to own action}} + \underbrace{\partial_i E^{\mu_L} U_i(\theta_i;\theta_i) - \partial_i E^{\mu_T} U_i(\theta_i;\theta_i)}_{\text{difference due to external (others') actions}}.$$

Examining these two differences, notice that applying a mean value theorem to each of these two differences gives that there exist s_i and $m_{-i}, s_{-i} \in R_{-i}^{k-1}(\theta_{-i})$ such that

$$-\partial_{ii}^2 E^{\mu_L} U_i(s_i;\theta_i)(m_i - \theta_i) = \sum_{j \neq i} \partial_{ij}^2 U_i(\theta_i, s_{-i};\theta)(m_j - \theta_j).$$

Second, if m_i is boundary such that $m_i = b_l$, then, because m_i is best reply,

$$-\partial_{ii}^2 E^{\mu_L} U_i(s_i;\theta_i)(m_i - \theta_i) \geq \sum_{j \neq i} \partial_{ij}^2 U_i(\theta_i, s_{-i};\theta)(m_j - \theta_j).$$

Third, if m_i is boundary such that $m_i = b_u$, then, because m_i is best reply,

$$-\partial_{ii}^2 E^{\mu_L} U_i(s_i;\theta_i)(m_i - \theta_i) \leq \sum_{j \neq i} \partial_{ij}^2 U_i(\theta_i, s_{-i};\theta)(m_j - \theta_j).$$

Given the sign of $\partial_{ii}^2 U_i$ and the respective signs of $(m_i - \theta_i)$ in the latter two cases, we can summarize that for, either boundary or inner, $m_i \in R_i^k(\theta_i)$ there exist not-yet eliminated messages s_i, s_{-i}, m_{-i} such that

$$|\partial_{ii}^2 E^{\mu_L} U_i(s_i;\theta_i)|\,|(m_i - \theta_i)| \leq |\sum_{j \neq i} \partial_{ij}^2 U_i(\theta_i, s_{-i};\theta)(m_j - \theta_j)|.$$

From this, for each agent j and round k, letting $l_j^k := \max_{\theta_j, m_j \in R_j^k(\theta_j)} |\theta_j - m_j|$, we have

$$|m_i - \theta_i| \leq \frac{\sum_{j \neq i} |\partial_{ij}^2 U_i(\theta_i, s_{-i};\theta)| l_j^{k-1}}{|\partial_{ii}^2 E^{\mu_L} U_i(s_i;\theta_i)|} \leq \left[|SE^t| l^{k-1}\right]_i.$$

Since this inequality holds for all k, we can apply it iteratively, which gives that in the kth round for all $m_i \in R_i^k(\theta_i)$,

$$|m_i - \theta_i| \leq \left[|SE^t| l^{k-1}\right]_i \leq \left[|SE^t||SE^t| l^{k-2}\right]_i \leq \ldots \leq \left[|SE^t|^k \mathbf{1}\right]_i.$$

By the eigenvalue condition, the spectral radius $\rho(|SE^t|) < 1$, and thus $|SE^t|^k \to 0$ in k, and thus for all i, $l_i^k \to 0$ in k, and thus full $\mathcal{B}^{id}$-implementation follows. $\square$

Before we proceed to Part 2, we show in the next step that under a $\mathcal{B}^{id}$-implementing transfer scheme the step-by-step iterative eliminations result in sets of k-rationalizable strategies, whose sizes reflect the canonical externalitites. (This step is again used in the Proof of Theorem 2 below.)

Step 1: (Iterations and Canonical Externalities.) Consider a twice differentiable, $\mathcal{B}^{id}$-IC direct mechanism (d,t). In relation to the canonical direct mechanism, for all θ_i there exist message profiles s^+ and $s^{+'}$ such that the message

$$\underset{R_i^{id,k-1}(\theta_i)}{\text{proj}}\left(\theta_i + \frac{\sum_{j\neq i}\partial_{ij}^2 E^{b_{\theta_i}} U_i^*\left(s^+;\theta_i\right) l_{o,i}^{k-1,+}}{|\partial_{ii}^2 E^{b_{\theta_i}} U_i^*\left(s^{+'};\theta_i\right)|}\right)$$

is in $R_i^{id,k}(\theta_i)$, and there exist message profiles s^- and $s^{-'}$ such that the message

$$\underset{R_i^{id,k-1}(\theta_i)}{\text{proj}}\left(\theta_i - \frac{\sum_{j\neq i}\partial_{ij}^2 E^{b_{\theta_i}} U_i^*\left(s^-;\theta_i\right) l_{o,i}^{k-1,-}}{|\partial_{ii}^2 E^{b_{\theta_i}} U_i^*\left(s^{-'};\theta_i\right)|}\right)$$

is in $R_i^{id,k}(\theta_i)$ too.

To show this Step, fix θ_i *in* $\left(\underline{\theta},\overline{\theta}\right)$ and fix some type $\theta_o \in \left(\underline{\theta},\overline{\theta}\right)$ and some message $m_o \in \left(\underline{\theta},\overline{\theta}\right)$ for i's opponents. Since t defines a $\mathcal{B}^{id} - IC$ mechanism, θ_i is best-reply to truthtelling conjectures. In particular, it is best-reply to the conjecture which, for every $j \neq i$, assigns probability 1 to the event that $\theta_j = \theta_o$ and assigns probability 1 to the event that all opponents are reporting their types. Let this - concentrated truth-reporting - conjecture be μ_T. There exists also a message of i which is best-reply to the conjecture that assigns probability 1 to the event that $\theta_j = \theta_o$ and assigns probability 1 to all opponents reporting m_o regardless of their types. Denote this undominated strategy by m_i and let this - concentrated m_o-reporting - conjecture be μ_L. Note that both μ_T and μ_L are consistent with $\mathcal{B}^{id}$. Consider the message m_i which is best reply to μ_L.

First, if m_i is an interior point, then we have that

$$0 = \partial_i E^{\mu_L} U_i\left(m_i;\theta_i\right) - \partial_i E^{\mu_T} U_i\left(\theta_i;\theta_i\right) = \partial_i E^{\mu_L} U_i^*\left(m_i;\theta_i\right) - \partial_i E^{\mu_T} U_i^*\left(\theta_i;\theta_i\right)$$

$$= \underbrace{\partial_i E^{\mu_L} U_i^*\left(m_i;\theta_i\right) - \partial_i E^{\mu_L} U_i^*\left(\theta_i;\theta_i\right)}_{\text{difference due to own action}} + \underbrace{\partial_i E^{\mu_L} U_i^*\left(\theta_i;\theta_i\right) - \partial_i E^{\mu_T} U_i^*\left(\theta_i;\theta_i\right)}_{\text{difference due to external (others') actions}},$$

where the first equality holds because of the canonical - additive - representation of (d,t) in Lemma 2 and the characterization of the belief-based terms in Step 1 of the Proof of Theorem 1. The conjectures μ_T and μ_L are constructed such that they satisfy identicality on the margins of the messages too.

In this Proof, we simplify the notation of those profiles in which opponents' elements are identical in that instead of $(s_o,\ldots,s_o,\theta_i,s_o,\ldots,s_o)$ we write $\left(\theta_i,s_{-i}^o\right)$.

Examining the two differences above, notice that by the mean value theorem, there exists s_i such that

$$\partial_i E^{\mu_L} U_i^*\left(m_i;\theta_i\right) - \partial_i E^{\mu_L} U_i^*\left(\theta_i;\theta_i\right) = \partial_{ii}^2 E^{\mu_L} U_i^*\left(s_i;\theta_i\right)\left(m_i - \theta_i\right),$$

and there exists s_o such that

$$\partial_i E^{\mu_L} U_i^* (\theta_i;\theta_i) - \partial_i E^{\mu_T} U_i^* (\theta_i;\theta_i) = \sum_{j\neq i} \partial_{ij}^2 U_i^* \left(\theta_i, s_{-i}^o; \theta_i, \theta_{-i}^o\right) (m_o - \theta_o).$$

Note that any k-th-round best-reply m_i is either inner point (as above) or a boundary point. Let $b_l \leq b_u$ be the boundary points of the set of $k-1$-rationalizable messages of θ_i.

Second, if m_i is boundary such that $m_i = b_l$, then, because m_i is best reply,

$$0 \geq \partial_i E^{\mu_L} U_i (m_i;\theta_i) - \partial_i E^{\mu_T} U_i (\theta_i;\theta_i) = \partial_i E^{\mu_L} U_i^* (m_i;\theta_i) - \partial_i E^{\mu_T} U_i^* (\theta_i;\theta_i),$$

which, following the steps as above, gives that there exists s_i and s_o such that

$$0 \geq \partial_{ii}^2 E^{\mu_L} U_i^* (s_i;\theta_i)(m_i - \theta_i) + \sum_{j\neq i} \partial_{ij}^2 U_i^* \left(\theta_i, s_{-i}^o; \theta_i, \theta_{-i}^o\right) (m_o - \theta_o).$$

This gives that $m_i = b_l$ only if there exists profiles such that

$$\theta_i - \frac{\sum_{j\neq i} \partial_{ij}^2 U_i^* \left(\theta_i, s_{-i}^o; \theta_i, \theta_{-i}^o\right) (m_o - \theta_o)}{\partial_{ii}^2 E^{\mu_L} U_i^* (s_i;\theta_i)} \leq b_l = m_i.$$

Third, if m_i is boundary such that $m_i = b_u$, then, because m_i is best reply,

$$0 \leq \partial_i E^{\mu_L} U_i (m_i;\theta_i) - \partial_i E^{\mu_T} U_i (\theta_i;\theta_i) = \partial_i E^{\mu_L} U_i^* (m_i;\theta_i) - \partial_i E^{\mu_T} U_i^* (\theta_i;\theta_i),$$

which gives that, for some profile,

$$\theta_i - \frac{\sum_{j\neq i} \partial_{ij}^2 U_i^* \left(\theta_i, s_{-i}^o; \theta_i, \theta_{-i}^o\right) (m_o - \theta_o)}{\partial_{ii}^2 E^{\mu_L} U_i^* (s_i;\theta_i)} \geq b_u = m_i.$$

We summarize these three cases and note that, for every θ_i, one can set θ_o and m_o such that $m_o - \theta_o = l_{i,o}^{k-1+}$, which gives that there exists s_o and s_i such that

$$m_i = \underset{R_i^{id,k-1}(\theta_i)}{\text{proj}} \left(\theta_i - \frac{\sum_{j\neq i} \partial_{ij}^2 U_i^* \left(\theta_i, s_{-i}^o; \theta_i, \theta_{-i}^o\right) l_{i,o}^{k-1,+}}{|\partial_{ii}^2 U_i^* \left(s_i, m_{-i}^o; \theta_i, \theta_{-i}^o\right)|} \right) \in R_i^{id,k} (\theta_i),$$

Now, for every θ_i, it is also possible to set θ_o and m_o such that $m_o - \theta_o = -l_{i,o}^{k-1,-}$. Considering the corresponding k-th round best reply m_i being interior or boundary, and following the previous steps we have that there exists s_o' and s_i' such that

$$m_i = \underset{R_i^{id,k-1}(\theta_i)}{\text{proj}} \left(\theta_i + \frac{\sum_{j\neq i} \partial_{ij}^2 U_i^* \left(\theta_i, s_{-i}^{o'}; \theta_i, \theta_{-i}^o\right) l_{i,o}^{k-1,-}}{|\partial_{ii}^2 U_i^* \left(s_i', m_{-i}^o; \theta_i, \theta_{-i}^o\right)|} \right) \in R_i^{id,k} (\theta_i),$$

which, completes the proof of this Step. □

Part 2. In SC-PC environments, for transfer schemes with linear moment conditions, the resulting payoff functions are such that the second order derivatives are constants. Let G_i be t's moment condition-based part and since it is linear for all i it can be written as $G_i (m) = L_i (m_{-i})\, m_i =$

$\sum_{j\neq i}\lambda_j m_j m_i$.[26] Then, the second order partial derivatives are such that $\partial^2_{ij}U^t_i(m;\theta)=\partial^2_{ij}U^*_i+\lambda_j$ and they are therefore constant in θ and m, for all i and j.

Let $l^k\in\mathbb{R}^n$ be the vector of the largest distance between kth round rationalizable strategies of the agents. Then, one can show that in every kth round of iterations, the set of rationalizable sets are not only bounded, but - by applying Step 1 and by the assumption SC-PC - are precisely given based on $|SE^t|$ such that:

$$R^k_i(\theta_i)=\left[\theta_i\pm\left[|SE^t|^k l^{k-1}\right]_i\right]\cap R^{k-1}_i(\theta_i),$$

where in this special case, $SE^t_{ij}=\partial^2_{ij}U^t_i/\partial^2_{ii}U^t_i$ if $j\neq i$. This means that the size of kth round rationalizable sets, $R^k_i(\theta_i)\to 0$ if and only if the spectral radius $\rho(|SE^t|)<1$. ■

Proof of Lemma 5.

Step 1: (Belief-Based Components under $\mathcal{B}^{id}$: Characterization) A differentiable function $K_i: M\to\mathbb{R}$ satisfies the expected value condition in (8) if and only if it can be written as

$$K_i(m)=\sum_{k=0}^{\infty}m_i^k\sum_{j\neq i}H^k_{ij}(m_j)$$

where $\left\{H^k_{ij}\right\}_{j\neq i,k\in\mathbb{N}}$ are polynomials $H^k_{ij}:M_j\to\mathbb{R}$ such that

$$\text{for all } m_{-i} \text{ for which } m_l=m_j \text{ for all } j,l\neq i:\sum_{j\neq i}H^k_{ij}(m_j)=0.$$

To show this step, assume, that K_i satisfies the expected value condition in (8) under $\mathcal{B}^{id}$. Since K_i is a continuous function, it can be approximated by Bernstein polynomials such that $K_i(m)=\lim_{n\to\infty}\sum_{v=0}^{n}K_i(m/n)\,b_{v,n}(m)$. Since K_i is bounded, this polynomial expression can be reorganized into a power series of m_i and thus there exist polynomials $H_k:M_{-i}\to\mathbb{R}$ such that $K_i(m)=\sum_{k=0}^{\infty}H_k(m_{-i})\,m_i^k$.

In the next two sub-steps, we show that, since K_i satisfies the expected value condition in (8) under $\mathcal{B}^{id}$, these H_ks are additively separable and at identical profiles, they are 0.

Step 1a: (Each H_k is additively separable.) From the polynomial format and since K_i satisfies the expected value condition, we have that for all k, $\mathbb{E}^{b_{\theta_i}}(H_k(\theta_{-i}))=0$ for all beliefs $b_{\theta_i}\in B^{id}_{\theta_i}$ for all θ_i. Fix a type θ_i. Assume, by way of contradiction, that H_k is not separable in its variables. More specifically and without loss of generality, assume that H_k is not separable in its first argument and, to avoid confusions in indexing, refer to this agent as j. This step relies on comparing two constructed joint distributions which both represent identical distributions but one of them represents perfectly correlated random variables, while the other one represents independence; that is, the jth random variable is independent from the other $n-2$ variables while these $n-2$ variables are again perfectly correlated.[27]

By the assumed non-separability, there exist $\theta^1\in[\underline{\theta},\overline{\theta}]$ and $\theta^2\in[\underline{\theta},\overline{\theta}]$ such that $\theta^1\neq\theta^2$ and

$$H_k\left(\theta^1,\theta^2,\ldots,\theta^2\right)-H_k\left(\theta^2,\theta^2,\ldots,\theta^2\right)\neq H_k\left(\theta^1,\theta^1,\ldots,\theta^1\right)-H_k\left(\theta^2,\theta^1,\ldots,\theta^1\right). \tag{17}$$

[26] To ease notation, we may write $\partial^2_{ij}U^t_i(m;\theta)$ for $\partial^2U^t_i(m;\theta)/(\partial m_i\partial m_j)$.

[27] This proof can be seen as a proof by coupling, a proof technique typically applied for topics that involve Markov chains and other finite or nonfinite discrete probabilities, but here applied for distributions over continuous support.

Consider the following two joint distributions over Θ_{-i}. Let p^{corr} be such that it prescribes perfect correlation for all agents in $I \setminus \{i\}$, and let p^{indep} be such that it prescribes perfect correlations for all agents in $I \setminus \{i\}$ except for j, where j's type is independent of the others' types. Let these two joint distributions further be such that on all their margins, they are equal and concentrated on the two specific values θ^1 and θ^2 such that for all $k \neq i$, $\text{marg}_{\Theta_k} p^{corr} = \text{marg}_{\Theta_k} p^{indep}$, and on θ^1: $\text{marg}_{\Theta_k} p^{corr}\left(\{\theta_k = \theta^1\}\right) = \text{marg}_{\Theta_k} p^{indep}\left(\{\theta_k = \theta^1\}\right) = 0.5$, and on θ^2: $\text{marg}_{\Theta_k} p^{corr}\left(\{\theta_k = \theta^2\}\right) = \text{marg}_{\Theta_k} p^{indep}\left(\{\theta_k = \theta^2\}\right) = 0.5$. Observe that both p^{corr} and p^{indep} are available under the belief restrictions $\mathcal{B}^{id}$, formally, $p^{corr} \in B^{id}_{\theta_i}$ and $p^{indep} \in B^{id}_{\theta_i}$. For ease of notations, let p be a probability measure over $[\underline{\theta}, \overline{\theta}]$ such that $p\left(\{\theta_k = \theta^1\}\right) = p\left(\{\theta_k = \theta^2\}\right) = 0.5$ and let f_p be p's distribution function.

Consider the perfectly correlated joint distribution p^{corr}, and observe that

$$\mathbb{E}^{p^{corr}}\left(H_k\left(\theta_{-i}\right)\right) = \int_{\Theta_{-i}} H_k\left(\theta_{-i}\right) dp^{corr} = \int_{\underline{\theta}}^{\overline{\theta}} H_k\left(\theta, \theta, \ldots, \theta\right) f_p d\theta =$$

$$= 0.5 H_k\left(\theta^1, \theta^1, \ldots, \theta^1\right) + 0.5 H_k\left(\theta^2, \theta^2, \ldots, \theta^2\right).$$

Consider the joint distribution, with independence from θ_j, p^{indep}, and observe that

$$\begin{aligned}\mathbb{E}^{p^{indep}}\left(H_k\left(\theta_{-i}\right)\right) = \int^{\Theta_{-i}} H_k\left(\theta_j, \theta_{-j,-i}\right) dp^{indep} = \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{\theta}}^{\overline{\theta}} H_k\left(\theta_j, \theta, \theta, \ldots, \theta\right) f_p \cdot f_p d\theta_j d\theta = \\ = 0.25 H_k\left(\theta^1, \theta^1, \ldots, \theta^1\right) + 0.25 H_k\left(\theta^1, \theta^2, \ldots, \theta^2\right) + 0.25 H_k\left(\theta^2, \theta^1, \ldots, \theta^1\right) + \\ + 0.25 H_k\left(\theta^2, \theta^2, \ldots, \theta^2\right) \neq \\ \neq 0.5 H_k\left(\theta^1, \theta^1, \ldots, \theta^1\right) + 0.5 H_k\left(\theta^2, \theta^2, \ldots, \theta^2\right).\end{aligned}$$

The last negation follows from Equation (17), which recall was the consequence of non-separability, and this negation implies that $\mathbb{E}^{p^{indep}}\left(H_k\left(\theta_{-i}\right)\right) \neq \mathbb{E}^{p^{corr}}\left(H_k\left(\theta_{-i}\right)\right)$, which would imply the contradiction that K_i does not satisfy the expected value condition. And therefore, H_k must be separable.

Step 1b: (Each H_k gives 0 at identical profiles.) Fix a type θ_i. Consider beliefs of i which are identical point-distributions; distributions which are concentrated on the same type of all other agents. Formally, consider a belief b_{θ_i} such that, for some $\theta \in [\underline{\theta}, \overline{\theta}]$, the probability $b_{\theta_i}\left(\{\theta_j = \theta \text{ for all } j \neq i\}\right)$ is 1 for all $j \neq i$. Then, b_{θ_i} is included in $B^{id}_{\theta_i}$, moreover such point-beliefs exist for all θ. Fix this (independent) point belief b_{θ_i}. The expected value condition implies that for the polynomial format $0 \equiv \sum_{k=1}^{\infty} \mathbb{E}^{b_{\theta_i}}\left(H_k\left(\theta_{-i}\right)\right) \theta_i^k$ and thus for any k $\mathbb{E}^{b_{\theta_i}}\left(H_k\left(\theta_{-i}\right)\right) = 0$. At identical profiles as represented by b_{θ_i}, this latter means that $H_k(\theta, \theta \ldots, \theta) = 0$ for all $\theta \in [\underline{\theta}, \overline{\theta}]$, which proves that the H_k are 0 at identical profiles.

To prove the other direction of this Step 1, assume that K_i satisfies the two conditions above, that is H_ks are additively separable and H_ks give 0 at identical profiles. For a type θ_i and belief $b_{\theta_i} \in B^{id}_{\theta_i}$, by the separability of H_ks and by the boundedness of K_i, the conditional expectation

is such that

$$\mathbb{E}^{b_{\theta_i}}\left(K_i\left(\theta\right)\right) = \int_{\Theta_{-i}} \sum_{k=1}^{\infty} H_k\left(\theta_{-i}\right)\theta^k db_{\theta_i} = \int_{\Theta_{-i}} \sum_{k=1}^{\infty}\sum_{j\neq i} H_{kj}\left(\theta_j\right)\theta^k db_{\theta_i}$$
$$= \sum_{k=1}^{\infty}\sum_{j\neq i}\left[\int_{\Theta_j} H_{kj}\left(\theta_j\right) d\operatorname*{marg}_{\Theta_j} b_{\theta_i}\right]\theta^k \tag{18}$$

Let p denote the identical distribution over $[\underline{\theta},\overline{\theta}]$ such that $p := \operatorname{marg}_{\Theta_j} b_{\theta_i}$ for all $j \neq i$. With this notation, Equation (18) is

$$\mathbb{E}^{b_{\theta_i}}\left(K_i\left(\theta\right)\right) = \sum_{k=1}^{\infty}\sum_{j\neq i}\left[\int_{\underline{\theta}}^{\overline{\theta}} H_{kj}\left(\theta\right)dp\right]\theta^k = \int_{\underline{\theta}}^{\overline{\theta}}\sum_{k=1}^{\infty}\sum_{j\neq i} H_{kj}\left(\theta\right)\theta^k dp = \int_{\underline{\theta}}^{\overline{\theta}} K_i\left(\theta_i,\theta,\theta,\ldots,\theta\right)dp,$$

and the two conditions,

$$\mathbb{E}^{b_{\theta_i}}\left(K_i\left(\theta\right)\right) = \int_{\underline{\theta}}^{\overline{\theta}} K_i\left(\theta_i,\theta,\theta,\ldots,\theta\right)dp = \int_{\underline{\theta}}^{\overline{\theta}} 0dp = 0.$$

and thus K_i satisfies the expected value condition under $\mathcal{B}^{id}$. □

Step 2: (Properties of K_is for Full Implementation) If K_i satisfies the expected value condition in 7, then based on the characterization in Step 1, we have that

(1) $\partial K_i\left(m_i,m_{-i}\right)/\partial m_i = \sum_{k=0}^{\infty} k m_i^{k-1}\sum_{j\neq i} H_{ij}^k\left(m_j\right) = \sum_{k=0}^{\infty} k m_i^{k-1} 0 = 0$ for all m_i and m_{-i} such that $m_l = m_j$ for all $j,l \neq i$; and

(2) $\sum_{j\neq i}\left(\partial K_i\left(m_i,m_{-i}\right)/\partial m_j\right) = \sum_{j\neq i}\left(\sum_{k=0}^{\infty} m_i^k \sum_{s\neq i} H_{is}^k\left(m_s\right)\right) = 0$ for all m_i and m_{-i} such that $m_l = m_s$ for all $s,l \neq i$. □

Finally, if (d,t) is $\mathcal{B}^{id}$-IC, then by Lemma 2, there exist $K_i : M \to \mathbb{R}$ which satisfies the expected value condition in 7; and is such that $U_i^t\left(m;\theta\right) = U_i^*\left(m;\theta\right) + \int^{m_i} K_i\left(s,m_{-i}\right)ds$. This equivalence and the two properties above in Step 2 complete the proof of this Lemma. ■

Proof of Proposition 1. If (d,t) achieves full $\mathcal{B}^{id}$-implementation, then Lemmas 4 and 5, applied to SC-PC and linear moment conditions, imply points (1.) and (2.) of this Proposition. In the other direction, if (1.) and (2.) hold, then (2.) implies that t_i's belief-based term defines a moment condition which satisfies the three conditions of Lemma 3, and thus (d,t) is a $\mathcal{B}^{id}$-IC mechanism. Therefore, (1.) implies full $\mathcal{B}^{id}$-implementation which completes the proof of this Proposition. ■

Proof of Theorem 2. To prove part 1 of Theorem 2, it is useful to characterize the resulting set of strategies from the step by step eliminations of $\mathcal{B}^{id}$-rationalizability.

Step 1: In every round k, for all i and θ_i, the set of rationalizable messages $R_i^{id,k}\left(\theta_i|t^l\right)$ is a closed interval around θ_i.

To show this, note that by construction $\theta_i \in R_i^{id,k}\left(\theta_i|t^l\right)$ and assume that $m_1, m_2 \in R_i^{id,k}\left(\theta_i|t^l\right)$. Then, there are conjectures for which these messages are best replies, that is, there exist μ_1 and μ_2 which are consistent with the $k-1$-st round and with identicality such that m_1 is best reply to μ_1 and m_2 is best reply to μ_2. Now, any convex combination $\lambda \in (0,1)$, $\lambda\mu_1 + (1-\lambda)\mu_2$ is also a conjecture which is consistent with the $k-1$-st round and with $\mathcal{B}^{id}$. Let m_λ denote the best reply

to this conjecture, which exists by the boundedness (which is implied by the differentiability) of v, d, t^l. Then, m_λ is continuous in λ, therefore the closed interval $[m_1, m_2] \subseteq R_i^{id,k}\left(\theta_i|t^l\right)$ and thus this set is closed and compact. $\square$

Recall that agents are ordered according to the absolute value of the ratio of the sum of their canonical externalities and own concavity, from the lowest to the highest, such that $\xi_{ij} := \partial^2 U_i^* / \left(\partial m_i \partial m_j\right) = -\left(\partial^2 v_i / \partial x \partial \theta_j\right) \cdot \left(\partial d / \partial \theta_i\right)$, $\xi_i := \sum_{j \neq i} \xi_{ij} / \xi_{ii}$ and $|\xi_1| \leq |\xi_2| \leq \ldots \leq |\xi_n|$. Recall that under SC-PC, these canonical externalities and the cross-derivatives in the resulting payoff functions in the loading mechanism $\left(d, t^l\right)$ are constants.

Step 2: In the loading mechanism, in every *two* rounds, the rate of shrinkage of the best reply sets in the iterative eliminations is $|\xi_1 \xi_2|$ for all agents.

To show this step, consider the loading direct mechanism $\left(d, t^l\right)$ and the iterative elimination process of $\mathcal{B}^{id}$-rationalizability.

In the first round of iterations, the size of the intervals which contain the strategies that survive the elimination derive from the loaded externality matrix such that:

$$SE^l = \begin{bmatrix} 0 & \xi_1 & 0 & \ldots & 0 \\ \xi_2 & 0 & 0 & \ldots & 0 \\ \xi_3 & 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \xi_n & 0 & 0 & \ldots & 0 \end{bmatrix} \text{ and } \left[R_i^{id,1}\left(\theta_i|t^l\right)\right]_{i \in I} = \begin{bmatrix} [\theta_1 \pm \xi_1] \cap [0,1] \\ [\theta_2 \pm \xi_2] \cap [0,1] \\ [\theta_3 \pm \xi_3] \cap [0,1] \\ \vdots \\ [\theta_n \pm \xi_n] \cap [0,1] \end{bmatrix}.$$

In the second round of iterations:

$$\left(SE^l\right)^2 = \begin{bmatrix} \xi_1\xi_2 & 0 & 0 & \ldots & 0 \\ 0 & \xi_1\xi_2 & 0 & \ldots & 0 \\ 0 & \xi_1\xi_3 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \xi_1\xi_n & 0 & \ldots & 0 \end{bmatrix} \text{ and } \left[R_i^{id,2}\left(\theta_i|t^l\right)\right]_{i \in I} = \begin{bmatrix} [\theta_1 \pm \xi_1\xi_2] \cap R_i^{id,1}\left(\theta_1|t^l\right) \\ [\theta_2 \pm \xi_1\xi_2] \cap R_i^{id,1}\left(\theta_2|t^l\right) \\ [\theta_3 \pm \xi_1\xi_3] \cap R_i^{id,1}\left(\theta_3|t^l\right) \\ \vdots \\ [\theta_n \pm \xi_1\xi_n] \cap R_i^{id,1}\left(\theta_n|t^l\right) \end{bmatrix}.$$

In the third round of iterations:

$$\left(SE^l\right)^3 = \begin{bmatrix} 0 & \xi_1^2\xi_2 & 0 & \ldots & 0 \\ \xi_1\xi_2^2 & 0 & 0 & \ldots & 0 \\ \xi_1\xi_2\xi_3 & 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \xi_1\xi_2\xi_n & 0 & 0 & \ldots & 0 \end{bmatrix} \text{ and } \left[R_i^{id,3}\left(\theta_i|t^l\right)\right]_{i \in I} = \begin{bmatrix} \left[\theta_1 \pm \xi_1^2\xi_2\right] \cap R_i^{id,2}\left(\theta_1|t^l\right) \\ \left[\theta_2 \pm \xi_1\xi_2^2\right] \cap R_i^{id,2}\left(\theta_2|t^l\right) \\ [\theta_3 \pm \xi_1\xi_2\xi_3] \cap R_i^{id,2}\left(\theta_3|t^l\right) \\ \vdots \\ [\theta_n \pm \xi_1\xi_2\xi_n] \cap R_i^{id,2}\left(\theta_n|t^l\right) \end{bmatrix}.$$

In the forth round of iterations:

$$\left(SE^l\right)^4 = \begin{bmatrix} \xi_1^2\xi_2^2 & 0 & 0 & \dots & 0 \\ 0 & \xi_1^2\xi_2^2 & 0 & \dots & 0 \\ 0 & \xi_1^2\xi_2\xi_3 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \xi_1^2\xi_2\xi_n & 0 & \dots & 0 \end{bmatrix} \text{ and } \left[R_i^{id,4}\left(\theta_i|t^l\right)\right]_{i\in I} = \begin{bmatrix} \left[\theta_1 \pm \xi_1^2\xi_2^2\right] \cap R_i^{id,3}\left(\theta_1|t^l\right) \\ \left[\theta_2 \pm \xi_1^2\xi_2^2\right] \cap R_i^{id,3}\left(\theta_2|t^l\right) \\ \left[\theta_3 \pm \xi_1^2\xi_2\xi_3\right] \cap R_i^{id,3}\left(\theta_3|t^l\right) \\ \vdots \\ \left[\theta_n \pm \xi_1^2\xi_2\xi_n\right] \cap R_i^{id,3}\left(\theta_n|t^l\right) \end{bmatrix}.$$

In the fifth round of iterations:

$$\left(SE^l\right)^5 = \begin{bmatrix} 0 & \xi_1^3\xi_2^2 & 0 & \dots & 0 \\ \xi_1^2\xi_2^3 & 0 & 0 & \dots & 0 \\ \xi_1^2\xi_2^2\xi_3 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \xi_1^2\xi_2^2\xi_n & 0 & 0 & \dots & 0 \end{bmatrix} \text{ and } \left[R_i^{id,5}\left(\theta_i|t^l\right)\right]_{i\in I} = \begin{bmatrix} \left[\theta_1 \pm \xi_1^3\xi_2^2\right] \cap R_i^{id,4}\left(\theta_1|t^l\right) \\ \left[\theta_2 \pm \xi_1^2\xi_2^3\right] \cap R_i^{id,4}\left(\theta_2|t^l\right) \\ \left[\theta_3 \pm \xi_1^2\xi_2^2\xi_3\right] \cap R_i^{id,4}\left(\theta_3|t^l\right) \\ \vdots \\ \left[\theta_n \pm \xi_1^2\xi_2^2\xi_n\right] \cap R_i^{id,4}\left(\theta_n|t^l\right) \end{bmatrix}.$$

And so on, in the k-th round of iteration, the size of the intervals which contain the strategies that survive the elimination derive from the loaded externality matrix to the power k and, if k is even, these intervals are given by

$$\left[R_i^{id,k}\left(\theta_i|t^l\right)\right]_{i\in I} = \begin{bmatrix} \left[\theta_1 \pm \xi_1^{k/2}\xi_2^{k/2}\right] \cap R_i^{id,k-1}\left(\theta_1|t^l\right) \\ \left[\theta_2 \pm \xi_1^{k/2}\xi_2^{k/2}\right] \cap R_i^{id,k-1}\left(\theta_2|t^l\right) \\ \left[\theta_3 \pm \xi_1^{k/2}\xi_2^{k/2-1}\xi_3\right] \cap R_i^{id,k-1}\left(\theta_3|t^l\right) \\ \vdots \\ \left[\theta_n \pm \xi_1^{k/2}\xi_2^{k/2-1}\xi_n\right] \cap R_i^{id,k-1}\left(\theta_n|t^l\right) \end{bmatrix},$$

and, if k is odd, these intervals are given by

$$\left[R_i^{id,k}\left(\theta_i|t^l\right)\right]_{i\in I} = \begin{bmatrix} \left[\theta_1 \pm \xi_1^{(k+1)/2}\xi_2^{(k-1)/2}\right] \cap R_i^{id,k-1}\left(\theta_1|t^l\right) \\ \left[\theta_2 \pm \xi_1^{(k-1)/2}\xi_2^{(k+1)/2}\right] \cap R_i^{id,k-1}\left(\theta_2|t^l\right) \\ \left[\theta_3 \pm \xi_1^{(k-1)/2}\xi_2^{(k-1)/2}\xi_3\right] \cap R_i^{id,k-1}\left(\theta_3|t^l\right) \\ \vdots \\ \left[\theta_n \pm \xi_1^{(k-1)/2}\xi_2^{(k-1)/2}\xi_n\right] \cap R_i^{id,k-1}\left(\theta_n|t^l\right) \end{bmatrix}.$$

In words, this means that in every *even round* of iteration, for each type of agent 1, the rationalizable set is either given by the previous rationalizable set or it is shrank to $|\xi_2|$ of this set and, for each type of agent $j \neq 1$, the rationalizable set is either the previous rationalizable set or it is shrank to $|\xi_1|$ of this set. The alternate holds for every *odd round* of iteration: for each type of agent 1, the rationalizable set is either the previous rationalizable set or it is shrank to $|\xi_1|$ of this set and, for each type of agent $j \neq 1$, the rationalizable set is either the previous rationalizable set or it is shrank to $|\xi_2|$ of this set. Combining the conclusions for odd and even rounds, we get that in every two rounds of iterations, for each type of each agent, the rationalizable set is either unchanged or it is shrank to $|\xi_1\xi_2|$ of this previous rationalizable set. □

And thus this step implies that if the sum of canonical externalities is such that $|\xi_1\xi_2| < 1$, then the size of the k-rationalizable sets converges to 0, and $R_i^{id}\left(\theta_i|t^l\right) = \{\theta_i\}$ for all i for all θ_i. On the other hand, if $|\xi_1\xi_2| \geq 1$, then $|\xi_2| \geq 1$ and in every round k, $R_i^{id,k}\left(\theta_2|t^l\right) = [\theta_2 \pm 1] \cap [0,1] = [0,1]$, in other words, all reports remain rationalizable for all types of agent 2 (and for all agents with an index larger than 2, too) and thus full implementation via t^l fails, which completes the proof of Part 1 of this Theorem.

To show the second part of this Theorem, we need to show that the allocation function d is $\mathcal{B}^{id}$-implementable if and only if it is $\mathcal{B}^{id}$-implementable via the loading transfers t^l in Equation 11. The if part is straightforward. The only if part, as we show next, relies on the fact that in any $\mathcal{B}^{id}$-implementing direct mechanism, the externalities can not be reduced beyond the sum of externalities in the canonical direct mechanism. The consequence of such irreducibility of externalities is reflected in each k-rationalizable set from the the step-by-step iterations, by Step 1 in the Proof of Lemma 4. This Step, in a more general setting without assuming SC-PC, shows that the canonical externalities are inherent in every iteration of $\mathcal{B}^{id}$-rationalizability.

Step 3: We use Step 1 in the Proof of Lemma 4 to show that in every round k, for all i and θ_i, the set of rationalizable messages of the loaded direct mechanism $R_i^{id,k}\left(\theta_i|t^l\right)$ are contained in $R_i^{id,k}\left(\theta_i|t\right)$, for any partially implementing direct mechanism (d,t).

To show this, fix a direct mechanism (d,t). Under SC-PC environments, Step 1 in the Proof of Lemma 4 implies that every k-rationalizable interval of θ_i of any implementing (d,t) direct mechanism contains the following set:

$$\underset{R_i^{id,k-1}(\theta_i|t)}{\text{proj}}\left[\theta_i - \xi_i \cdot l_{i,o}^{k-1,-}, \theta_i + \xi_i \cdot l_{i,o}^{k-1,+}\right] \subseteq \; R_i^{id,k}\left(\theta_i|t\right).$$

Recall that $l_{i,o}^{k-1,+}$ is the largest distance between positive misreport and the true type, which can arise for all opponents of i based on the previous round of iteration and $l_{i,o}^{k-1,-}$ is similarly this largest distance for negative misreport.

Next, we compare the k-rationalizable sets of (d,t) to the k-rationalizable sets of $\left(d,t^l\right)$, where the latter sets are already given in Step 2 of this proof. In particular, for the first round of iteration,

$$[\theta_i - \xi_i, \theta_i + \xi_i] \cap [0,1] \subseteq \; R_i^{id,1}\left(\theta_i|t\right).$$

For the second round of iteration,

$$[\theta_1 - \xi_1\xi_2, \theta_1 + \xi_1\xi_2] \cap [0,1] \subseteq \; R_i^{id,2}\left(\theta_i|t\right) \text{ if } i = 1 \text{ and}$$

$$[\theta_i - \xi_i\xi_1, \theta_i + \xi_i\xi_1] \cap [0,1] \subseteq \; R_i^{id,2}\left(\theta_i|t\right) \text{ if } i \neq 1.$$

For the third round of iteration,

$$[\theta_1 - \xi_1\left(\xi_1\xi_2\right), \theta_1 + \xi_1\left(\xi_1\xi_2\right)] \cap [0,1] \subseteq \; R_i^{id,3}\left(\theta_i|t\right) \text{ if } i = 1 \text{ and}$$

$$[\theta_i - \xi_i\left(\xi_1\xi_2\right), \theta_i + \xi_i\left(\xi_1\xi_2\right)] \cap [0,1] \subseteq \; R_i^{id,3}\left(\theta_i|t\right) \text{ if } i \neq 1.$$

For the forth round of iteration,

$$\left[\theta_1 - \xi_1\left(\xi_1\xi_2^2\right), \theta_1 + \xi_1\left(\xi_1\xi_2^2\right)\right] \cap [0,1] \subseteq \ R_i^{id,4}\left(\theta_i|t\right) \text{ if } i = 1 \text{ and}$$

$$\left[\theta_i - \xi_i\left(\xi_1^2\xi_2\right), \theta_i + \xi_i\left(\xi_1^2\xi_2\right)\right] \cap [0,1] \subseteq \ R_i^{id,4}\left(\theta_i|t\right) \text{ if } i \neq 1.$$

Observe that in these expressions on the left hand side, the iterated sets derived based on Step 1 in the Proof of Lemma 4 for every k coincide with the iterated rationalizable sets of the loaded direct mechanism $\left(d, t^l\right)$, and thus by induction, for all k, $R_i^{id,k}\left(\theta_i|t^l\right) \subseteq \ R_i^{id,k}\left(\theta_i|t\right)$ for any partially implementing direct mechanism (d,t), which completes the proof of this step. □

Since, as we assumed, (d,t) achieves full $\mathcal{B}^{id}$-implementation, by the containments, we must have that as $k \to \infty$, $|R_i^{id,k}\left(\theta_i|t^l\right)| \to 0$, and thus $\left(d, t^l\right)$ achieves full $\mathcal{B}^{id}$-implementation too. ■

Proof of Lemma 6. By the Gershgorin circle theorem, both under condition (i) and (ii) the absolute value of all eigenvalues of $|SE^t|$ are smaller than 1, which by the eigenvalue lemma in Lemma 4 ensures full $\mathcal{B}^{id}$-implementation. ■

Proof of Proposition 2. The equal-externality transfer scheme t^e under the assumption of SC-PC is $\mathcal{B}^{id}$-IC. Moreover, t^e induces a strategic externality matrix which is such that for all $i, j \neq i$, $SE_{ij}^e = \left(\sum_{j\neq i} \frac{\partial^2 v_i}{\partial x \partial \theta_j} / \frac{\partial^2 v_i}{\partial x \partial \theta_i}\right) \frac{1}{n-1}$. For this externality matrix, notice that condition (i) of this Proposition implies condition (i) of Lemma 6; and condition (ii) of this Proposition implies condition (ii) of Lemma 6, and thus by Lemma 6, full $\mathcal{B}^{id}$-implementation follows. ■

Proof of Proposition 3. Under SC-PC, t^e ensures $\mathcal{B}^{id}$-incentive compatibility. Next, we show that t^e ensures full $\mathcal{B}^{id}$-implementation too.

First, assume that there exists a transfer scheme t which ensures full $\mathcal{B}^{id}$-implementation and limited strategic externalities as in (i) of Lemma 6.

By the characterization of belief-based terms for $\mathcal{B}^{id}$-IC in the Proof of Lemma 5, there exists (m,θ) for which $\sum_{j\neq i} SE_{ij}^t(m;\theta) = \sum_{j\neq i} SE_{ij}^*$.[28] Next we show that t^e induces an externality matrix which satisfies the conditions of the eigenvalue lemma in Lemma 4. By construction of t^e, $\sum_{j\neq i}|SE_{ij}^e| = \sum_{j\neq i}|\sum_{j\neq i} SE_{ij}^*/(n-1)| = |\sum_{j\neq i} SE_{ij}^*|$. And thus, there exists (m,θ) such that $\sum_{j\neq i}|SE_{ij}^e| = |\sum_{j\neq i} SE_{ij}^t(m;\theta)| \leq \sum_{j\neq i}|SE_{ij}^t(m;\theta)| < 1$. The latter strict inequality holds by (i) of Lemma 6 and thus by the Gershgorin circle theorem, $\rho|SE^e| < 1$ and thus by the eigenvalue lemma, Lemma 4 t^e too ensures full $\mathcal{B}^{id}$-implementation.

Second, assume that there exists a transfer scheme t which ensures full $\mathcal{B}^{id}$-implementation and limited strategic impacts as in (ii) of Lemma 6.

By the characterization of belief-based terms for $\mathcal{B}^{id}$-IC in the Proof of Lemma 5, there exists (m,θ) for which $|\sum_{i\in I}\sum_{j\neq i} SE_{ij}^*| = |\sum_{i\in I}\sum_{j\neq i} SE_{ij}^t(m;\theta)| \leq \sum_{i\in I}\sum_{j\neq i}|SE_{ij}^t(m;\theta)|$ and thus, by t satisfying (ii) of Lemma 6, $|\sum_{i\in I}\sum_{j\neq i} SE_{ij}^*| < n$ and writing this with the total externality notation, $\sum_{i\in I}\xi_i < n$. Now, consider the absolute externality matrix induced by the equal-externality transfers t^e. In what follows, using the Perron-Frobenius theorem, we show that this matrix has a spectral radius which is less than 1. $|SE^e|$ is a non-negative matrix, with zeros in its diagonal and by its construction, for all i and $j \neq i$, $|SE^e|_{ij} = |\sum_{j\neq i} SE_{ij}^*|/(n-1)$, in other notation, $|SE^e|_{ij} = |\xi_i|/(n-1)$. Let ρ denote the largest eigenvalue of this matrix. (Assume that ξ_is, the absolute total canonical externalities, are ordered as before, based on their absolute values, from the smallest to the largest.) By the Perron-Frobenius theorem, there is a positive 1-norm vector $v \in \mathbb{R}^n$ such that $\rho v = |SE^e|v$. The componentwise consequence of this is that,

[28] Closely related to the externality matrix, in this proof, we let $\mathcal{S}_{ij}^{(t)}(m;\theta) := \partial_{ij}^2 U_i^{(t)}(m;\theta)/\partial_{ii}^2 U_i^{(t)}(m;\theta)$.

for all i, $\rho\frac{v_i}{|\xi_i|} = \frac{\sum_{j\neq i} v_j}{n-1}$, which also implies that if $|\xi_i| \leq |\xi_j|$, then $v_i \geq v_j$. Adding up these n equations and expressing ρ, gives that $\rho = \frac{\sum_{j\in I} v_i}{\sum_{j\in I} \frac{v_i}{|\xi_i|}}$, which is a weighted harmonic mean of ξ_is with weights v_is. And thus from a weighted harmonic mean – arithmetic mean inequality, $\rho = \frac{\sum_{j\in I} v_i}{\sum_{j\in I} \frac{v_i}{|\xi_i|}} \leq \sum_{i\in I} \frac{v_i}{\sum_{j\in I} v_j} |\xi_i|$. Since larger $|\xi_i|$s have smaller weights, this latter expression is bounded by the average of $|\xi_i|$s, and thus $\rho \leq \sum_{i\in I} \frac{|\xi_i|}{n} < 1$. Recall from above that, the latter strict inequality is a consequence of t satisfying (ii) of Lemma 6. Therefore, by the eigenvalue lemma, Lemma 4, t^e ensures full $\mathcal{B}^{id}$-implementation. ■

Proof of Proposition 4. Observe that, under symmetric aggregators in valuations, $\sum_{k\neq i} \frac{\partial^2 v_i}{\partial x \partial \theta_k}$ is the same for all agents:

$$\begin{aligned}\sum_{k\neq i} \frac{\partial^2 v_i}{\partial x \partial \theta_k} &= \sum_{k\neq i} \partial^2_{x,h} w \cdot \frac{\partial h_i}{\partial \theta_k} = \partial^2_{x,h} w \cdot \sum_{k\neq i} \frac{\partial h_i}{\partial \theta_k} = \partial^2_{x,h} w \sum_{k\neq j} \frac{\partial h_j}{\partial \theta_k} \text{ for all } j \\ &= \sum_{k\neq j} \frac{\partial^2 v_j}{\partial x \partial \theta_k} \text{ for all } j.\end{aligned}$$

Observe also that, under symmetric aggregators in valuations, $\frac{\partial^2 v_i}{\partial x \partial \theta_i}$ is the same for all agents:

$$\begin{aligned}\frac{\partial^2 v_i}{\partial x \partial \theta_i} &= \partial^2_{x,h} w \cdot \frac{\partial h_i}{\partial \theta_i} = \partial^2_{x,h} w \cdot \frac{\partial h_j}{\partial \theta_j} \text{ for all } j \\ &= \frac{\partial^2 v_j}{\partial x \partial \theta_j} \text{ for all } j.\end{aligned}$$

To prove part 2 of this proposition, recall the characterization of Theorem 2, which says that an increasing allocation function is full $\mathcal{B}^{id}$-implementable if and only if $|\xi_1 \xi_2| < 1$. This latter condition is equivalent to $|\sum_{k\neq 1} \frac{\partial^2 v_1}{\partial x \partial \theta_k} \cdot \sum_{k\neq 2} \frac{\partial^2 v_2}{\partial x \partial \theta_k}| < \frac{\partial^2 v_1}{\partial x \partial \theta_1} \cdot \frac{\partial^2 v_2}{\partial x \partial \theta_k}$. Under symmetric aggregators, this latter inequality is equivalent to $|\sum_{k\neq i} \frac{\partial^2 v_i}{\partial x \partial \theta_k}| < \frac{\partial^2 v_i}{\partial x \partial \theta_i}$ for all i, which completes the proof of this part.

To prove part 1 of this proposition, note that from Theorem 2, if the equal-externality mechanism (d, t^e) achieves full $\mathcal{B}^{id}$-implementation, then the loaded direct mechanism $\left(d, t^l\right)$ achieves this too. To prove the other direction, note that if the loaded direct mechanism $\left(d, t^l\right)$ achieves full $\mathcal{B}^{id}$-implementation, then by the previous part of this proof, we have $\left|\sum_{k\neq i} \frac{\partial^2 v_i}{\partial x \partial \theta_k}\right| < \frac{\partial^2 v_i}{\partial x \partial \theta_i}$ for every i, and (i) of Proposition 2 implies that the equal-externality transfers ensure full $\mathcal{B}^{id}$-implementation too, which completes the proof of this part. ■

Proof of Proposition 3. Fix α and fix an SC-PC environment $\left(d^{PC}, v^{PC}\right)$. Consider an environment (d^α, v^α), which is α-close to $\left(d^{PC}, v^{PC}\right)$. In this environment, consider the direct mechanism given by the transfer scheme in Equation 13.

$$t_i^\alpha(m) = t_i^{*,\alpha}(m) + L_i^{(l,PC)}(m_{-i})\, m_i.$$

These transfers feature an additive modification of this environment's canonical transfers, and the additive term is defined by what the loading modification would be in the α-close SC-PC environment $\left(d^{PC}, v^{PC}\right)$.

First, notice that this direct mechanism is $\mathcal{B}^{id}$-incentive compatible. This is so because t^*

is incentive compatible and because under the belief restriction of identicality (identicality and independence), in the first order condition, $\mathbb{E}^{b_{\theta_i}}\left(L_i^{(l,PC)}\left(\theta_{-i}\right)\right)=0$ for all θ_i and for all i, moreover the second order condition is the same as the second order condition of the canonical direct mechanism, which by assumption is negative, and thus ensures optimality.

Second, we show that there exists $\alpha_0>0$, such that for all smaller $\alpha>0$, the direct mechanism t^α guarantees externalities which satisfy the eigenvalue condition of the eigenvalue lemma, Lemma 4. To this end, fix α such that $0<\alpha<\min_i|\xi_{ii}|$. To study the externality matrix of the direct mechanism (d,t^α), notice that, because of α-closeness, for all $(i,j)\in\{(1,2),(2,1),(3,1),\ldots,(n,1)\}$, $|SE^\alpha|_{ij}=\frac{\max_{m,\theta}|\partial^2_{ij}U_i^\alpha(m;\theta)|}{\min_{m,\theta}|\partial^2_{ii}U_i^\alpha(m;\theta)|}\leq\frac{|\sum_{j\neq i}\xi_{ij}|+\alpha}{|\xi_{ii}|-\alpha}=:|\mathcal{E}^\vee_\alpha|_{ij}$, and for all other (i,j) such that $i\neq j$, $|SE^\alpha|_{ij}=\frac{\max_{m,\theta}|\partial^2_{ij}U_i^\alpha(m;\theta)|}{\min_{m,\theta}|\partial^2_{ii}U_i^\alpha(m;\theta)|}\leq\frac{\alpha}{|\xi_{ii}|-\alpha}=:|\mathcal{E}^\vee_\alpha|_{ij}$. Let $|\mathcal{E}^\vee_\alpha|_{ii}:=0$ for all $i\in I$. Based on these definitions, if $\alpha\to 0$, then element-wise, $|SE^\alpha|<|\mathcal{E}^\vee_\alpha|\to|SE^{l,PC}|$, where the latter matrix is the loaded externality matrix in $\left(d^{PC},v^{PC}\right)$. Since $\left(d^{PC},v^{PC}\right)$ admits full $\mathcal{B}^{id}$-implementation, by Theorem 2, $\rho\left(|SE^{l,PC}|\right)=\sqrt{\xi_1\xi_2}<1$. And thus, by the continuity of the spectral radius, there exists $\alpha_0>0$ such that for all $\alpha<\alpha_0$, $\rho\left(|\mathcal{E}^\vee_\alpha|\right)<1$, and thus, $\rho\left(|SE^\alpha|\right)\leq\rho\left(|\mathcal{E}^\vee_\alpha|\right)$[29], therefore, $\rho\left(|SE^\alpha|\right)<1$ too. And thus the eigenvalue condition of Lemma 4 holds and it implies that for all $\alpha<\alpha_0$, for an α-close environment (d^α,v^α), the transfers t^α, as given in Equation 13 ensure $\mathcal{B}^{id}$-Implementation. ■

Proof of Lemma 7. Recall that it is the consequence of SC-PC and linear moment conditions that in the utility functions of the direct mechanism (d,t), the second order derivatives are constants. Formally, let G_i be t's moment condition-based part and since it is linear in m_i for all i it can be written as $G_i(m)=L_i(m_{-i})\,m_i$; and $\partial^2_{ij}U_i$ is constant in θ and m, for all i and j. Then, for any $\mu\in C^{\mathcal{B}^{id}}_{\theta_i}$, adding and subtracting $L_i(\theta_{-i})$, applying Leibniz's rule and the triangle inequality, we have that

$$\left|\frac{\partial EU^\mu_{\theta_i}}{\partial m_i}(\theta_i)\right|=\left|\int_{\Theta_{-i}\times M_{-i}}\left(\frac{\partial v_i}{\partial x}\left(d\left(\theta_i,m_{-i}\right),\theta\right)-\frac{\partial v_i}{\partial x}\left(d\left(\theta_i,m_{-i}\right),\theta_i,m_{-i}\right)\right)\frac{\partial d}{\partial\theta_i}\left(\theta_i,m_{-i}\right)\right.$$

$$\left.+L_i(m_{-i})-L_i(\theta_{-i})+L_i(\theta_{-i})-f_i(\theta_i)\,d\mu\right|$$

$$\leq\int_{\Theta_{-i}\times M_{-i}}\sum_{j\neq i}\left|\frac{\partial^2U_i}{\partial m_i\partial m_j}\right||\theta_j-m_j|\,d\mu+\varepsilon\leq\sum_{j\neq i}\left|\frac{\partial^2U_i}{\partial m_i\partial m_j}\right|+\varepsilon. \tag{19}$$

These inequalities hold with equality when the RHS is projected to the $[0,1]$ interval for a conjecture μ, which is concentrated on profiles such that θ_j-m_j has the opposite sign as the partial derivative and is of distance one. [check such μ conjecture is valid].

For any $m_i\in R_i^{\mathcal{B},1}(\theta_i)$, there exists $\mu\in C^{\mathcal{B}}_{\theta_i}$ such that $m_i\in BR_{\theta_i}(\mu)$, fix this μ. Since m_i is best reply, it minimizes the first-order partial derivative. Using (19), by the concavity of the expected utility function, for this $\mu\in C^{\mathcal{B}}_{\theta_i}$, we have that $\left|\frac{\partial EU^\mu_{\theta_i}}{\partial m_i}(m_i)-\frac{\partial EU^\mu_{\theta_i}}{\partial m_i}(\theta_i)\right|\leq\sum_{j\neq i}\left|\frac{\partial^2U_i}{\partial m_i\partial m_j}\right|+\varepsilon$. By the mean value theorem, there exists $s_i\in M_i$ such that $\left|\frac{\partial^2EU^\mu_{\theta_i}}{\partial^2m_i}(s_i)\right||m_i-\theta_i|\leq\sum_{j\neq i}\left|\frac{\partial^2U_i}{\partial m_i\partial m_j}\right|+$

[29] To see this inequality, recall Gelfand's formula, which characterizes the spectral radius of a matrix A such that $\rho(A)=\lim_{k\to\infty}||A^k||^{1/k}$. One can check using Gelfand's formula that if a non-negative matrix A is element-wise dominated by a non-negative matrix B, that is if $A_{ij}\leq B_{ij}$ for all i,j, then $\rho(A)\leq\rho(B)$, which implies the inequality above.

ε. Recall that the second order derivative is constant and the same as in the canonical direct mechanism, therefore, for all θ_i and $m_i \in R_i^{\mathcal{B},1}(\theta_i)$,

$$|m_i - \theta_i| \leq \frac{\sum_{j \neq i} \left|\partial_{ij}^2 U_i\right| + \varepsilon}{|\partial_{ii}^2 U_i|}.$$

Thus in this first round of iterative eliminations, with notation $A_1 := |SE^t| + [\varepsilon, 0]\, C$ for all θ_i the rationalizable messages are

$$R_i^{\mathcal{B}^{id},\varepsilon,1}(\theta_i) = [\theta_i \pm [A_1 \mathbf{1}]_i] \cap [0,1]$$

In the second round of iterative eliminations, for any $m_i \in R_i^{\mathcal{B},2}(\theta_i)$, there exists $\mu \in C_{\theta_i}^{\mathcal{B}} \cap R_i^{\mathcal{B},1}(\theta_i)$ such that $m_i \in BR_{\theta_i}(\mu)$. For the Taylor-expansion of $\partial EU_{\theta_i}^{\mu}/\partial m_i$ at θ_i around m_i there exists $s_i \in M_i$ such that $\frac{\partial EU_{\theta_i}^{\mu}}{\partial m_i}(\theta_i) = \frac{\partial EU_{\theta_i}^{\mu}}{\partial m_i}(m_i) + \frac{\partial^2 EU_{\theta_i}^{\mu}}{\partial^2 m_i}(s_i)(\theta_i - m_i)$. Since m_i is best reply to μ and $EU_{\theta_i}^{\mu}(m_i)$ is strictly concave, we have that $\left|\frac{\partial^2 EU_{\theta_i}^{\mu}}{\partial^2 m_i}(s_i)\right| |\theta_i - m_i| \leq \left|\frac{\partial EU_{\theta_i}^{\mu}}{\partial m_i}(\theta_i)\right|$. Further bounding this by adding and subtracting $L_i(\theta_{-i})$, applying Leibniz's rule and the triangle inequality, we get that $\left|\frac{\partial EU_{\theta_i}^{\mu}}{\partial m_i}(\theta_i)\right| \leq \int_{\Theta_{-i} \times M_{-i}} \sum_{j \neq i} \left|\partial_{ij}^2 U_i\right| |\theta_j - m_j|\, d\mu + \varepsilon$. Therefore, for all θ_i and $m_i \in R_i^{\mathcal{B},2}(\theta_i)$

$$|\theta_i - m_i| \leq \sum_{j \neq i} \frac{|\partial_{ij}^2 U_i|}{|\partial_{ii}^2 U_i|} \left[\sum_{l \neq j} \frac{|\partial_{jl}^2 U_j| + \varepsilon}{|\partial_{jj}^2 U_j|}\right] + \frac{\varepsilon}{|\partial_{ii}^2 U_i|}$$

Thus in this second round of iterative eliminations, with notation $A_2 := |SE^t|^2 + [\varepsilon, 0]\, |SE^t| C + [\varepsilon, 0]\, C$ for all θ_i the rationalizable messages are

$$R_i^{\mathcal{B}^{id},\varepsilon,2}(\theta_i) = [\theta_i \pm [A_2 \mathbf{1}]_i] \cap [0,1]$$

By induction, at the k^{th} round, with notation $A_k := |SE^t|^k + [\varepsilon, 0]\, |SE^t|^{k-1} C + \ldots + [\varepsilon, 0]\, |SE^t| C + [\varepsilon, 0]\, C = |SE^t|^k + [\varepsilon, 0]\left(I - |SE^t|^k\right)\left(I - |SE^t|\right)^{-1} C$ (the latter equality assuming that $\rho(|SE^t|) < 1$.) for all θ_i the rationalizable messages are

$$R_i^{\mathcal{B}^{id},\varepsilon,k}(\theta_i) = [\theta_i \pm [A_k \mathbf{1}]_i] \cap [0,1]$$

Taking limits as $k \to \infty$, we have that for all i and θ_i, the rationalizable messages for all θ_i are

$$R_i^{\mathcal{B}^{id},\varepsilon}(\theta_i) = \left[\theta_i \pm \left[\left([\varepsilon, 0]\left(I - |SE^t|\right)^{-1} C\right) \mathbf{1}\right]_i\right] \cap [0,1]$$

■

Proof of Proposition 5.

For the loading transfers t^l, the inverse of $I - |SE^l|$ is as follows:

$$\left(I - |SE^l|\right)^{-1} = \begin{bmatrix} 1 & -|\xi_1| & 0 & \dots & 0 \\ -|\xi_2| & 1 & 0 & \dots & 0 \\ -|\xi_3| & 0 & 1 & \vdots & \vdots \\ \vdots & \vdots & 0 & \ddots & 0 \\ -|\xi_m| & 0 & \dots & 0 & 1 \end{bmatrix}^{-1}$$

$$= \frac{1}{1 - |\xi_1\xi_2|} \begin{bmatrix} 1 & |\xi_1| & 0 & \dots & 0 \\ |\xi_2| & 1 & 0 & \dots & 0 \\ |\xi_3| & |\xi_1\xi_3| & 1 - |\xi_1\xi_2| & \vdots & \vdots \\ \vdots & \vdots & 0 & \ddots & 0 \\ |\xi_m| & |\xi_1\xi_m| & \vdots & 0 & 1 - |\xi_1\xi_2| \end{bmatrix}.$$

For the equal-externality transfers t^e, the inverse of $I - |SE^e|$, under symmetric aggregators, is

$$(I - |SE^e|)^{-1} = \begin{bmatrix} 1 & -\frac{|\xi|}{(n-1)} & -\frac{|\xi|}{(n-1)} & \dots & -\frac{|\xi|}{(n-1)} \\ -\frac{|\xi|}{(n-1)} & 1 & -\frac{|\xi|}{(n-1)} & \dots & -\frac{|\xi|}{(n-1)} \\ -\frac{|\xi|}{(n-1)} & -\frac{|\xi|}{(n-1)} & 1 & \vdots & \vdots \\ \vdots & \vdots & -\frac{|\xi|}{(n-1)} & \ddots & -\frac{|\xi|}{(n-1)} \\ -\frac{|\xi|}{(n-1)} & -\frac{|\xi|}{(n-1)} & \dots & -\frac{|\xi|}{(n-1)} & 1 \end{bmatrix}^{-1}$$

$$= \frac{1}{\left(1 + \frac{|\xi|}{n-1}\right)(1 - |\xi|)} \begin{bmatrix} 1 - \frac{(n-2)|\xi|}{n-1} & \frac{|\xi|}{n-1} & \dots & \frac{|\xi|}{n-1} \\ \frac{|\xi|}{n-1} & 1 - \frac{(n-2)|\xi|}{n-1} & \dots & \frac{|\xi|}{n-1} \\ \vdots & \vdots & \ddots & \frac{|\xi|}{n-1} \\ \frac{|\xi|}{n-1} & \dots & \frac{|\xi|}{n-1} & 1 - \frac{(n-2)|\xi|}{n-1} \end{bmatrix}.$$

Applying Lemma 7 to these inverses, one can notice that for the loading transfers, for all $k > 1$,

$$\eta^{t^l}(\varepsilon, k) = \frac{1 + |\xi|}{1 - \xi^2}\varepsilon = \frac{1}{1 - |\xi|}\varepsilon,$$

and for the equal-exgternality transfers,

$$\eta^{t^e}(\varepsilon, k) = \frac{1 - \frac{n-2}{n-1}|\xi| + \frac{k-1}{n-1}|\xi|}{\left(1 + \frac{|\xi|}{n-1}\right)(1 - |\xi|)}\varepsilon = \frac{1 - \frac{n-k-1}{n-1}|\xi|}{\left(1 + \frac{|\xi|}{n-1}\right)(1 - |\xi|)}\varepsilon.$$

Comparing

$$\frac{1}{1 - |\xi|} \text{ to } \frac{1 - \frac{n-k-1}{n-1}|\xi|}{\left(1 + \frac{|\xi|}{n-1}\right)(1 - |\xi|)}$$

is equivalent to comparing

$$1 + \frac{|\xi|}{n-1} \text{ to } 1 - \frac{n-k-1}{n-1}|\xi|,$$

from which we get that for all $1 < k < n$ and for all $\varepsilon > 0$, $\eta^{t^e}(\varepsilon, k) < \eta^{t^l}(\varepsilon, k)$, in other words, the equal-externality transfers are less sensitive to mistakes in play. ■

Proof of Theorem 4. See the proof of Step 3 in the Proof of Theorem 2. ■

References

Aghion, P., D. Fudenberg, R. Holden, T. Kunimoto, O. Tercieux, *Subgame-perfect Implementation under Information Perturbations*, The Quarterly Journal of Economics, 127(4), 1843–1881, 2012.

Arrow, K. J., *The Property Rights Doctrine and Demand Revelation under Incomplete Information*, in Economics and Human Welfare, 23–39. Academic Press, 1979

Artemov, G., T. Kunimoto and R. Serrano, *Robust Virtual Implementation with Incomplete Information: Towards a Reinterpretation of the Wilson Doctrine*, Journal of Economic Theory, 148(2): 424–447, 2013.

Athey, S. and P. Haile, *Nonparametric Approaches to Auctions*, Chapter 60 in Handbook of Econometrics, vol. 6A, Elsevier, 2007.

Battigalli, P. and M. Siniscalchi, *Rationalization and Incomplete Information*, Advances in Theoretical Economics, 3(1), 2003.

Bergemann, D. and S. Morris, *Robust Mechanism Design*, Econometrica, 73(6), 1771–1813, 2005.

Bergemann, D. and S. Morris, *Robust Implementation in Direct Mechanisms*, Review of Economic Studies, 76, 1175–1204, 2009.

Bergemann, D. and S. Morris, *Robust Virtual Implementation*, Theoretical Economics, 4(1), 2009

Bergemann, D. and S. Morris, *Robust Implementation in General Mechanisms*, Games and Economic Behavior, 71(2): 261-281, 2011

Borgers, T. and D. Smith, *Robust Mechanism Design and Dominant Strategy Voting Rules*, Theoretical Economics 9, 339-360, 2014.

Carlsson, H. and E. Van Damme, *Global Games and Equilibrium Selection*, Econometrica, 989–1018, 1993.

Carroll, G., *Robustness and Linear Contracts*, American Economic Review 105, 536–563, 2015.

Dasgupta, P. and E. Maskin, *Efficient Auctions*, The Quarterly Journal of Economics 115(2), 341–388, 2000.

d'Aspremont, C., J. Cremer and L-A. Gerard-Varet, *Incentives and Incomplete Information*, Journal of Public Economics 11:25–45, 1979.

Deb, R. and M. M. Pai, *Discrimination via Symmetric Auctions*, Journal of American Economic Journal: Microeconomics,275–314, 2017.

de Clippel, G., R. Saran and R. Serrano, *Level-k Mechanism Design*, Review of Economic Studies 86(3) , 1207–1227, 2018.

Cremer, J. and R.P. McLean, *Full extraction of the surplus in Bayesian and dominant strategy auctions*, Econometrica, 1247–1257, 1988.

Eliaz, K., *Fault Tolerant Implementation*, The Review of Economic Studies 69(3), 589–610, 2002.

Green, J., and JJ. Laffont, *Characterization of Satisfactory Mechanisms for the Revelation of Preferences for Public Goods*, Econometrica, 427–438, 1977.

Healy, P. J., and L. Mathevet, *Designing Stable Mechanisms for Economic Environments*, Theoretical Economics 7(3), 609–661, 2012.

Hendricks, K., J. Pinkse and R. Porter, *Empirical Implications of Equilibrium Bidding in First-Price, Symmetric, Common Value Auctions*, Review of Economic Studies, 70, 115–145, 2003.

Jackson, M. O., *Implementation in Undominated Strategies: A Look at Bounded Mechanisms*, Review of Economic Studies, 59(4), 757–75, 1992.

Jehiel, P., L. Lamy, *A Mechanism Design Approach to the Tiebout Hypothesis*, Journal of Political Economy, Forthcoming, 2017.

Kajii, A., S. Morris, *The Robustness of Equilibria to Incomplete Information*, Econometrica, 283–1309, 1997.

Kneeland, T., *Mechanism Design with Level-k Types: Theory and an Application to Bilateral Trade*, Working Paper, 2018.

Laffont, J-J. and E. Maskin, *A Differential Approach to Dominant Strategy Mechanisms*, Econometrica, 1507–1520, 1980.

Levy, G. and R. Razin, *Correlation Neglect, Voting Behavior, and Information Aggregation*, American Economic Review, 105, 4, 1634–45, 2015.

Li, Y., *Approximation in Mechanism Design with Interdependent Values*, Games and Economic Behavior, 103, 225–253, 2017.

Lipnowski, E. and E. Sadler, *Peer–Confirming Equilibrium*, Econometrica 87(2), 567–591, 2019.

Guo, H. and N. C. Yannelis, *Robust Strong Nash Implementation*, mimeo, University of Iowa, 2017.

Lopomo, G., L. Rigotti, and C. Shannon, *Uncertainty in mechanism design*, Working Paper, University of Pittsburgh, 2011.

Maskin, E., *Nash Equilibrium and Welfare Optimality*, The Review of Economic Studies, 66(1), 23–38, 1999.

Mathevet, L., *Supermodular Mechanism Design*, Theoretical Economics 5(3), 403–443, 2010.

Mathevet, L. and I. Taneva, *Finite Supermodular Design with Interdependent Valuations*, Games and Economic Behavior 82, 327–349, 2013.

Milgrom, P.R., *Putting auction theory to work*, Cambridge University Press. Vancouver, 2004.

Moore, J. and R. Repullo, *Subgame Perfect Implementation*, Econometrica, 1191–1220, 1988.

Morris, S. and H. S. Shin, *Global Games: Theory and Applications*, Advances in Economics and Econometrics (56), 2003.

Morris, S. and H. S. Shin, *Unique Equilibrium in a Model of Self-fulfilling Currency Attacks*, American Economic Review, 587–597, 1998.

Müller, C., *Robust Implementation in Weakly Perfect Bayesian Strategies*, mimeo, 2018.

Müller, C., *Robust Virtual Implementation under Common Strong Belief in Rationality*, Journal of Economic Theory (162), 407–450, 2016.

Myerson, R.B., *Optimal auction design*, Mathematics of operations research, 6(1), 58–73, 1981.

Ollár, M., *Shared Information Sources in Exchanges*, Working Paper, 2017.

Ollár, M. and A. Penta, *Full Implementation and Belief Restrictions*, American Economic Review, August, 2017.

Penta, A., *Higher Order Uncertainty and Information: Static and Dynamic Games*, Econometrica 80(2), 631–660, 2012.

Penta, A., *On the Structure of Rationalizability for Arbitrary Spaces of Uncertainty*, Theoretical Economics 8(2), 405–430, 2013.

Penta, A., *Robust Dynamic Mechanism Design*, Journal of Economic Theory 160, 280–316, 2015.

Penta, A., P. Zuazo-Garin, *Rationalizability and Observability*, Working Paper, 2017.

Rubinstein, A., *The Electronic Mail Game: Strategic Behavior Under "Almost Common Knowledge"*, The American Economic Review, 385–391, 1989.

Segal, I., *Optimal Pricing Mechanisms with Unknown Demand*, The American Economic Review 93 (3), 509–529, 2003.

Weinstein, J., M. Yildiz, *A Structure Theorem for Rationalizability with Application to Robust Predictions of Refinements*, Econometrica 75(2), 365–400, 2007.

Weinstein, J., M. Yildiz, *Impact of Higher-order Uncertainty*, Games and Economic Behavior 60 (1), 200–212, 2007.

Weinstein, J., M. Yildiz, *Sensitivity of Equilibrium Behavior to Higher-order Beliefs in Nice Games*, Games and Economic Behavior 72(1), 288–300, 2011.

Weinstein, J., M. Yildiz, *Reputation without Commitment in Finitely Repeated Games*, Theoretical Economics 11(1), 157–185, 2016.

Wilson, R., *Game-Theoretic Analysis of Trading Processes*, Advances in Economic Theory, ed. by Bewley, Cambridge University Press, 1987.

Wolitzky, A., *Mechanism Design with Maxmin Agents: Theory and an Application to Bilateral Trade*, Theoretical Economics 11(3), 971–1004, 2016.

Yamashita, T., *Implementation in Weakly Undominated Strategies: Optimality of Second-Price Auction and Posted-Price Mechanism*, Review of Economic Studies 82, 1223–1246, 2015.