Review of Economic Studies (2023) **90**, 2517–2554 doi:10.1093/restud/rdac084

Advance access publication 24 January 2023

A Network Solution to Robust Implementation: The Case of Identical but Unknown Distributions

MARIANN OLLÁR
University of Edinburgh
and
ANTONIO PENTA
ICREA, UPF, BSE and TSE

First version received September 2021*; Editorial decision October* 2022*; Accepted December* 2022 (Eds.)

We study robust mechanism design in environments in which agents commonly believe that others' types are identically distributed, but we do not assume that the actual distribution is common knowledge, nor that it is known to the designer. First, we characterize all incentive compatible transfers under these assumptions. Second, we characterize the conditions under which full implementation is possible via direct mechanisms, that only elicit payoff relevant information, and the transfer schemes which achieve it whenever possible. The full implementation results obtain from showing that the problem can be transformed into one of designing a *network of strategic externalities*, subject to suitable constraints which are dictated by the incentive compatibility requirements.

Key words: Robust full implementation, Rationalizability, Interdependent values, Identical but unknown distributions

JEL Codes: D62, D82, D83, D85

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 their opponents as ex-ante symmetric from an informational viewpoint, or more broadly that they come from a common population of "types." Standard modelling 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*

The editor in charge of this paper was Andrea Galeotti.

is a natural way to capture a basic qualitative property of these environments, *common knowledge* of the distribution is a different kind of assumption: not only is it strong and unlikely to be satisfied; it is also well-known to heavily affect the results. Inspired by Wilson (1987)'s call for a "[...] repeated weakening of common knowledge assumptions [...]," and in the spirit of the robust mechanism design literature,[1] in this paper, we explore questions of *partial* and *full implementation* under the assumption that agents commonly believe that others' types are identically distributed, but *without* assuming that the distribution of types is commonly known, or that it is known to the designer. We will refer to this assumption as "common belief in identicality," and to the restrictions it entails on agents beliefs as the $\mathcal{B}^{id}$-restrictions.

We focus on general environments in which a finite number of agents have preferences over allocations and a private good, "money." As it is standard in the mechanism design literature, we assume that preferences are quasi-linear in the latter and that each agent has payoff-relevant private information. The designer wishes to choose an allocation, depending on agents' preferences over outcomes, and hence the desired allocation is a function of the realized vector of *payoff types*. The designer's problem is thus to make the agents willing to reveal their types, so as to implement the desired allocation. We allow for general interdependence in agents' valuations, and hence agents' preferences over outcomes may depend both on their own and on the others' types. The main restrictions we impose are that types are one dimensional, and that both the valuations and the allocation rules are twice differentiable. For example, the designer may wish to induce the efficient level of provision of a public good, that equalizes the marginal cost of production with the sum of the agents' marginal utilities. These marginal utilities may depend on the agents' own type, as well as on the others' types, and the more an agent's marginal utility for the public good depends on others' types, the stronger the level of preference interdependence among the agents.

We assume that, in order to implement the desired decision rule, the social planner may only use *transfer schemes* that only elicit agents' information about preferences: for each profile of reports by the agent, the planner chooses the allocation that corresponds to the desired decision, treating the reported types as if they were true, and the transfer scheme determines how much each agent should pay or receive, as a function of everyone's reports. The implementation question is thus whether such transfers can be designed so that agents find it in their interest to report their types truthfully. For *partial* $\mathcal{B}^{id}$*-implementation*, this means that truthful revelation must be a mutual best response for all beliefs that the agents may hold about others' types, given the $\mathcal{B}^{id}$-restrictions ($\mathcal{B}^{id}$*-incentive compatibility*). *Full* $\mathcal{B}^{id}$*-implementation* instead requires that truthful revelation be *the only* rationalizable solution under common belief in identicality.[2] For each notion of implementation, we identify the transfer schemes that achieve it whenever possible, and the conditions on the environment under which partial and full $\mathcal{B}^{id}$-implementation are possible.

We start our analysis with the introduction of the *canonical transfers* (cf. Ollár and Penta, 2017). These are the transfers that one is bound to use if truthful revelation is required to be an ex-post equilibrium (so called *ex-post incentive compatibility*), and hence they characterize the possibility of achieving partial implementation in belief-free settings (cf.

1. The robust mechanism design literature was spurred by the seminal works in belief-free settings by Bergemann and Morris (2005, 2009a, 2009b, 2011) for static mechanisms, followed by Müller (2016, 2020) and Penta (2015) for dynamic ones. Settings with partial belief restrictions have been studied, for instance, by Lopomo *et al.* (2011), Artemov *et al.* (2013), and Ollár and Penta (2017, 2022b).

2. The solution concept we adopt, $\mathcal{B}^{id}$-rationalizability, is a special case of Battigalli and Siniscalchi (2003)'s Δ-rationalizability. Besides the papers cited in the previous footnote, special versions of Δ-rationalizability have also been used in implementation theory by Oury and Tercieux (2012) and Kunimoto *et al.* (2021).

Bergemann and Morris, 2005). For instance, if the designer wishes to implement the efficient allocation rule, then the canonical transfers coincide with the generalized Vickrey–Clarke–Groves (VCG) mechanism.

Our first result shows that, when only common belief of identicality is maintained, *partial implementation* is possible if and only if it can be achieved by the canonical transfers. This, however, is not to say that partial $\mathcal{B}^{id}$-implementation is possible if and only if ex-post incentive compatibility is (for a counterexample, see Example 2 below), but only that in both cases it suffices to check if incentive compatibility holds for the same transfers (namely, the *canonical* ones). The intuition for this result is the following: under the $\mathcal{B}^{id}$-restrictions, from the designer's viewpoint, payoff types of the same agent do not differ in terms of their beliefs about others' types, and hence beliefs cannot be used to better screen types at the truthful profile (which is the focus of *partial* implementation). Compared with the belief-free case, however, the $\mathcal{B}^{id}$-restrictions do relax the incentive compatibility constraints and enlarge the set of incentive compatible transfers, which provides extra flexibility that the designer can leverage when trying to achieve *full implementation*, where the possibility of non-truthful profiles must also be taken into account.

Since *full $\mathcal{B}^{id}$-implementation* requires that truthful revelation be *the only* rationalizable solution under common belief in identicality, strict $\mathcal{B}^{id}$-incentive compatibility is necessary for full $\mathcal{B}^{id}$-implementation. Therefore, compared with partial $\mathcal{B}^{id}$-implementation, the extra desideratum is a uniqueness result. As we show, uniqueness crucially depends on the *strategic externalities* that are induced by a mechanism, that is on how agents' best responses are affected by the reports of the others. Specifically, first we define a measure of strategic externalities between each pair of agents and collect them in a *matrix of strategic externalities*. Then, we show that this matrix is key for uniqueness and, hence, for $\mathcal{B}^{id}$-full implementation. In particular, for environments that satisfy a standard single-crossing and a public-concavity condition, and for transfer schemes that are quadratic in the agents' reports, we show that a $\mathcal{B}^{id}$-incentive compatible transfer scheme also achieves full $\mathcal{B}^{id}$-implementation if and only if the spectral radius of the associated matrix of strategic externalities is less than one.[3]

With these results, in order to identify the transfer schemes that achieve full $\mathcal{B}^{id}$-implementation whenever possible, we aim to construct transfers that minimize the *spectral radius* of the corresponding strategic externality matrix, subject to preserving $\mathcal{B}^{id}$-incentive compatibility.[4] Crucially, it turns out that the latter is possible only if the associated matrix of strategic externalities features the same row-sums as those that are induced by the canonical transfers. The key to our design strategy is thus to *redistribute the strategic externalities* of the canonical transfers, subject to maintaining their row-sums and satisfy the relevant incentive compatibility constraints, in order to minimize the set of rationalizable reports. This is equivalent to a network design problem, in which nodes represent agents, and flows on the directed edges represent the strategic externalities, which must be arranged in order to minimize the spectral radius, up to an inflow constraint on each node. We find that the transfers that solve this problem induce a mechanism that features a stark hierarchical structure: besides preserving, for each

3. The spectral radius of a matrix is the largest absolute value of its eigenvalues. The case of SC-PC environments and quadratic transfers, which we discuss here, provides the easiest-to-read conditions for uniqueness. Outside of these settings, there is a gap between the necessity and sufficiency. The general conditions we provide in Lemma 1 are based on the spectral radii of an upper and lower-bound of the strategic externality matrix. In SC-PC settings with quadratic transfers, these two bounds coincide. In Section 4.1, we discuss the connections of our conditions with the global stability of linear dynamical systems.

4. A different characterization of economic concepts, based on the spectral radius of the matrix of payoff externalities, is provided by Elliott and Golub (2019), in the context of efficiency with public goods.

player, the total level of strategic externalities, these transfers *load* them all on the opponent who displays the lowest amount of preference interdependence. The strategic externalities associated with such *loading transfers* are thus described by a *star network* whose centre is the agent with the lowest level of total preference interdependence. In this star network, each node has one incoming edge; externalities flow to the peripheral nodes only from the centre, and to the centre only from the node with the second-lowest level of preference interdependence.

The structure of the *loading transfers* enables us to uncover a fairly surprising result: the possibility of full $\mathcal{B}^{id}$-implementation is characterized by the strength of the preference interdependence of the two agents for whom it is smallest, regardless of the number of the other agents, and of their preferences. At the extreme, whenever an environment includes at least *one* agent with private values, common belief in identicality ensures that full implementation is possible via a simple direct mechanism. Besides depicting a much more permissive picture for full implementation than Bergemann and Morris (2009a)'s belief-free benchmark (which may perhaps strike as surprising, given the weakness of the $\mathcal{B}^{id}$-restrictions), this characterization has potentially powerful implications from a broader market design perspective, which we will discuss in the conclusions.

2. FRAMEWORK

Preferences, types, and allocation rules. We consider environments with transferable utility with a finite set of agents $I = \{1, \ldots, n\}$, in which the space of allocations X is a compact and convex subset of a Euclidean space.

Agents privately observe their payoff types $\theta_i \in \Theta_i := [\underline{\theta}, \overline{\theta}] \subseteq \mathbb{R}$, drawn from a closed interval on the real line, which we assume is 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. 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.

Beliefs. We assume that agents commonly know that they each regard the types of the opponents to be identically distributed, 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 the idea that the opponents' types are identically distributed.[5] Formally, the designer's

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

assumptions about beliefs is represented by belief restrictions $\mathcal{B}^{id} = ((B^{id}_{\theta_i})_{\theta_i \in \Theta_i})_{i \in I}$, assumed common knowledge, such that:[6]

$$B^{id}_{\theta_i} = \left\{ b_{\theta_i} \in \Delta(\Theta_{-i}) : \underset{\Theta_j}{marg}\, b_{\theta_i} = \underset{\Theta_k}{marg}\, b_{\theta_i} \text{ for all } j, k \neq i \right\} \text{ for all } i \text{ and } \theta_i. \tag{2.1}$$

These belief restrictions entail weaker assumptions on agents' beliefs than many standard models in more applied theory and in empirical work. The belief restrictions in (2.1) 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. For instance, a type θ_i may subjectively believe that others' types are i.i.d. according to some distribution, whereas a different type θ_i' may believe that they are perfectly correlated (note that both such beliefs satisfy the marginality condition in (2.1)). Different types of agent i may thus entertain different beliefs about the opponents, which may or may not assume that types are i.i.d. across the opponents. However, the only thing that any of i's types *know* about others' beliefs (as well as the only thing that the designer knows about them), is that they must satisfy the marginality condition. Hence, our belief restrictions entail a very weak level of common knowledge in the environment.

Mechanisms. We consider *direct mechanisms*, in which agents report their payoff types and the allocation is chosen according to d. A direct mechanism is thus uniquely determined by a transfer scheme $t = (t_i)_{i \in I}$, where $t_i : M \to \mathbb{R}$ is twice differentiable and 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, and simply write $U_i(m; \theta)$. 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 $\mathbb{E}^{b_{\theta_i}}(U_i(m_i, \theta_{-i}; \theta_i, \theta_{-i}))) := \int_{\Theta_{-i}} U_i(m_i, \theta_{-i}; \theta_i, \theta_{-i})\, db_{\theta_i}$. 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, $j \neq i$ and $k \neq i, j$. Types $\theta_i \in [0, 1]$ are private

6. 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 $B_{\theta_i} = \{p(\cdot \mid \theta_i)\}$ for all i and θ_i; (iii) belief-free environments, in which $B_{\theta_i} = \Delta(\Theta_{-i})$ for all i and θ_i.

information to each agent i, and $\gamma = ((\gamma_{ij})_{j\neq i})_{i=1,2,3} \in \mathbb{R}^6$ are the parameters of preference interdependence. For instance, the public good in question could be quantity of public amenities in a city, and each agent's valuation for such a public good depends on both their own extroversion score, θ_i, as well as on others' (for references on the extroversion scale and other personality traits, and their use in economics, see *e.g.* Becker *et al.*, 2012). The effect of others' extroversion on one's own valuation, however, may vary—*e.g.* j's extroversion may have a larger effect than k's on the valuation of agent i (*i.e.* $\gamma_{ij} > \gamma_{ik}$), for instance because i is more likely to interact with j-types than with k-types (*e.g.* based on their geographic location within the city). In this context, the commonly known γ parameters represent how much the valuation of agents from different neighbourhoods depend on the extroversion of individuals' from other neighbourhoods. The $\mathcal{B}^{id}$-restrictions instead reflect the idea that extroversion (which is private information, and hence unobservable to others as well as to the designer) is commonly believed to be identically distributed across the (observable) neighbourhoods, although agents from different neighbourhoods or different types from each neighbourhood may differ in their beliefs about such a distribution.

Example 1. The designer wishes to implement the efficient allocation rule. With production costs $c(x) = x^2/2$, the efficient decision rule is $d(\theta) = \sum_{i=1}^{3} \kappa_i \theta_i$, where $\kappa_i \equiv 1 + \gamma_{ji} + \gamma_{ki}$ for all i, which 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. Our paper focuses on the latter environment, which will be discussed in part (iii) of the example. It is instructive, however, to first go over the examples about the belief-free and i.i.d. common prior benchmarks.

(i) 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 can be used (cf. Bergemann and Morris, 2009a).

In our example, the VCG transfers are the following:

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

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[7]

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

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 the equation above 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 interdependence 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 profiles are rationalizable, and hence belief-free full implementation fails.

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

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 ijth 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 mechanism, this amounts to

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

(ii) Identical and known distribution: reduction of strategic externalities. Strategic externalities and preference interdependence necessarily coincide in the VCG mechanism. But if the designer has some information about the agents' beliefs, then this coincidence is 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).

Implementation under 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):

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

Then, to a conjecture $\mu \in \Delta(\Theta_{-i} \times M_{-i})$, the best-reply function is

$$BR_{\theta_i}^{OP}(\mu) = \operatorname*{proj}_{[0,1]} \left(\theta_i + \sum_{l \neq i} \gamma_{il} [\mathbb{E}^{\mu}(\theta_l) - 0.5] \right).$$

Under the maintained assumptions, $\mathbb{E}(\theta_l \mid \theta_i) = 0.5$ 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* (what we refer to as *interim dominant strategy implementation*), and full implementation is achieved for any γ . Players' best-responses are not affected by other reports, and hence strategic externalities are 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 "$\mathbb{E}(\theta_l \mid \theta_i) = 0.5$ 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 part (ii) to weaken the strategic externalities of the baseline transfers from part (i), but in principle, others could be used too. 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.

(iii) Identical but unknown distribution: redistribution of strategic externalities. Now suppose that agents commonly know that they each regard the types of their opponents as being drawn from 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 "$\mathbb{E}(\theta_l \mid \theta_i) = 0.5$ for all θ_i and $l \neq i$," and hence incentive compatibility may fail. In fact, as we will show, Ollár and Penta (2017)'s 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.

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

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

In this case, to a conjecture $\mu \in \Delta(\Theta_{-i} \times M_{-i})$, the best-reply function becomes

$$\begin{aligned} BR^e_{\theta_i}(\mu) &= \operatorname*{proj}_{[0,1]} \left(\theta_i + \frac{1}{2}(\gamma_{ij} + \gamma_{ik}) \sum_{l \neq i} \mathbb{E}^\mu(\theta_l - m_l) + \frac{1}{2}(\gamma_{ij} - \gamma_{ik}) \mathbb{E}^\mu(\theta_j - \theta_k) \right) \\ &= \operatorname*{proj}_{[0,1]} \left(\theta_i + \frac{1}{2}(\gamma_{ij} + \gamma_{ik}) \sum_{l \neq i} \mathbb{E}^\mu(\theta_l - m_l) \right) \end{aligned}$$

The simplification in the last line follows from the fact that, under the $\mathcal{B}^{id}$ restrictions, it is common belief that $\mathbb{E}(\theta_j - \theta_k \mid \theta_i) = 0$ for all beliefs that any type θ_i may entertain. Thus, this mechanism is incentive compatible for all beliefs consistent with $\mathcal{B}^{id}$: if for all θ_l and $l \neq i$, $m_l = \theta_l$, then for all i, the best response is $m_i = \theta_i$. 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 in this setting.

Next, consider the following 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 $\gamma = \hat{\gamma}$, and compare them to the SE-matrix of the VCG transfers:

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

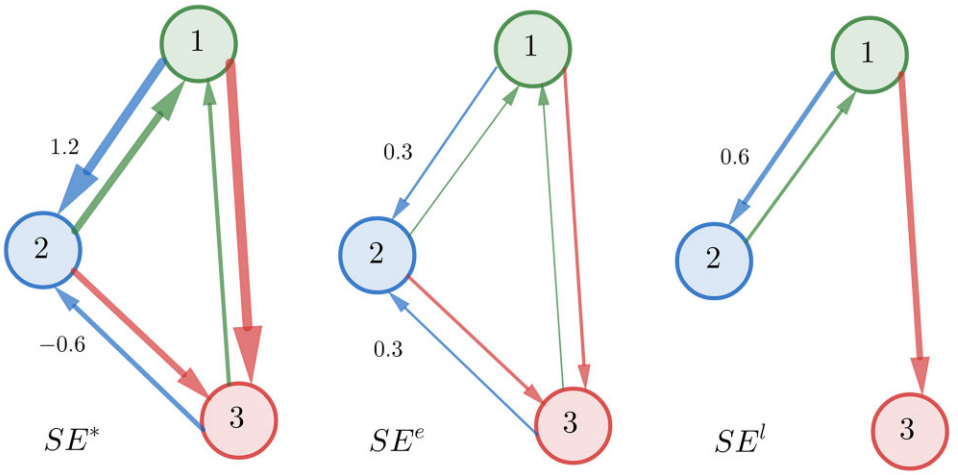

FIGURE 1

Strategic externalities and transfer schemes in part (iii) of Example 1. These network representations illustrate the strategic externalities induced, respectively, by the canonical, equal externality, and loading transfers. For example, the arrow pointing from agent 2 to 1 illustrates the absolute influence of 2's choice on 1's best reply.

First note that both $(t_i^e)_{i\in I}$ and $(t_i^l)_{i\in I}$ induce SE-matrices such that the sum of the strategic externalities 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 $(t_i^{OP})_{i\in I}$ from part (ii) of Example 1, 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 (Figure 1).

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 $(t_i^l)_{i\in I}$ 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:

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

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 $(t_i^l)_{i\in I}$ than under $(t_i^e)_{i\in I}$. In fact, it can be shown that among all matrices which preserve the row-sums of the SE^*-matrix, the strategic externality matrix associated with the loading transfers is the one with the smallest

spectral radius. This implies that, among all incentive compatible transfers, the loading transfers are those which induce the fastest contraction of the best-reply sets.

Our main results for full implementation show that, in a general class of environments, a suitable generalization of the *loading transfers* 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 the risk of mistakes in some agents' play (cf. Ollár and Penta, 2022b, also discussed in Section 4.3).

2.2. *Implementation concepts*

Next we formalize the notions of both partial and full implementation. We start from partial implementation, and first recall the standard notion of *ex-post incentive compatibility*, which requires truthful revelation to be an ex-post equilibrium of the game induced by a direct mechanism:

Definition 1 (ep-IC)**.** 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'.

As shown by Bergemann and Morris (2005), ex-post incentive compatibility characterizes the possibility of partial implementation when the designer has no information about agents' beliefs. In the present context, however, the designer knows that agents' beliefs are consistent with the $\mathcal{B}^{id}$-restrictions, and hence our analysis of partial implementation relies on the following less demanding notion of incentive compatibility:

Definition 2 ($\mathcal{B}^{id}$-IC)**.** 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 $b_{\theta_i} \in B^{id}_{\theta_i}$, $\mathbb{E}^{b_{\theta_i}}(U_i(\theta_i, \theta_{-i};\theta_i,\theta_{-i})) \geq \mathbb{E}^{b_{\theta_i}}(U_i(\theta_i', \theta_{-i};\theta_i,\theta_{-i}))$. (When d is clear from the context, we may say that t is $\mathcal{B}^{id}$-IC.) If the above inequalities hold strictly for all $\theta_i' \neq \theta_i$, then it is *strictly* $\mathcal{B}^{id}$-IC.

Definition 3 (Partial Implementation)**.** If (d,t) is $\mathcal{B}^{id}$-IC, then we say that the transfers t partially $\mathcal{B}^{id}$-implement the allocation function d. The allocation function d is *partially* $\mathcal{B}^{id}$-implementable if there exist transfers that partially $\mathcal{B}^{id}$-implement it.

First note that $\mathcal{B}^{id}$-IC is more demanding than standard Bayesian incentive compatibility, since it requires truthful revelation to be a mutual best-reply for all beliefs in the set $B^{id}_{\theta_i}$, as opposed to the single beliefs that each type would have in a standard Bayesian setting. However, since each $B^{id}_{\theta_i}$ is a strict subset of $\Delta(\Theta_{-i})$ (and, in particular, it does not contain all degenerate beliefs over each $\theta_{-i} \in \Theta_{-i}$), then $\mathcal{B}^{id}$-IC is less demanding than ex-post incentive compatibility.

Similar to Bergemann and Morris (2005), one could define Partial $\mathcal{B}^{id}$-Implementation as requiring truthful revelation to be *a* Bayes–Nash equilibrium for 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 Definition 2. Given this, the natural full implementation notion is to require truthful revelation to be *the only* Bayes–Nash equilibrium strategy for all type spaces consistent with the $\mathcal{B}^{id}$-restrictions. Once again, arguments similar to Bergemann and Morris (2009a) show that the set of all such Bayes–Nash equilibrium strategies is conveniently characterized by a suitable notion of rationalizability,

which will be introduced shortly, and which we refer to as $\mathcal{B}^{id}$-rationalizability.[8] Our notion of full implementation will thus require truthful revelation to be the only $\mathcal{B}^{id}$-rationalizable strategy. For the reasons, we explained, this notion can be seen as a shortcut to analyse standard questions of Bayesian implementation for all beliefs consistent with the $\mathcal{B}^{id}$ restrictions, and hence provides the natural counterpart to the notion of partial implementation that we introduced above.[9]

Formally, $\mathcal{B}^{id}$-rationalizability is defined by an iterated deletion procedure in which, for each type θ_i, a report survives the kth round of deletion if and only if it can be justified by conjectures (joint distributions over opponents' types and reports) that are consistent with the $\mathcal{B}^{id}$-restrictions, and with the previous rounds of deletion. For every i and θ_i, the set of conjectures that are consistent with the $\mathcal{B}^{id}$-restrictions is $C^{id}_{\theta_i} := \{\mu_i \in \Delta(M_{-i} \times \Theta_{-i}) : \text{marg}_{\Theta_{-i}} \mu_i \in B^{id}_{\theta_i}\}$.

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

$$R^{id,k}_i = \{(m_i, \theta_i) : m_i \in BR_{\theta_i}(\mu_i) \text{ for some } \mu_i \in C^{id}_{\theta_i} \cap \Delta(R^{id,k-1}_{-i})\}.$$

The set of $\mathcal{B}^{id}$-rationalizable messages for type θ_i is $R^{id}_i(\theta_i) := \{m_i : (m_i, \theta_i) \in \bigcap_{k \geq 0} R^{id,k}_i\}$.

Definition 5 (Full implementation)**.** The transfer scheme $t = (t_i)_{i \in I}$ fully implements d under common belief in identicality if $R^{id}_i(\theta_i) = \{\theta_i\}$ for all θ_i and all i. Allocation rule d is *fully* $\mathcal{B}^{id}$-implementable if there exist some transfers that fully $\mathcal{B}^{id}$-implement it.[10]

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 $\mathcal{B}^{id}$-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. Finally, note that 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 necessary condition for full $\mathcal{B}^{id}$-implementation. For this reason, while the main focus of the paper is on the analysis of full implementation, we first tackle the partial $\mathcal{B}^{id}$-implementation problem, and return to full $\mathcal{B}^{id}$–implementation in Section 4.

In the next two sections, we characterize the joint conditions on (v, d) under which partial and full $\mathcal{B}^{id}$-implementation is possible, as well as the transfer schemes that (partially or fully) implement d whenever possible.

8. $\mathcal{B}^{id}$-rationalizability is a special case of Battigalli and Siniscalchi (2003)'s Δ-rationalizability, which in general allows for general restrictions on players' first-order beliefs on others' types and strategies. Within robust mechanism design, special cases of Δ-rationalizability have been used by Bergemann and Morris (2009a), who impose no belief restrictions, and by Ollár and Penta (2017), who focused on belief restrictions that are only on others' types; Lipnowski and Sadler (2019) instead adopted restrictions on beliefs about others' behaviour for their concept of peer-confirming equilibrium, although not in an implementation setting.

9. 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).)

10. 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.

3. INCENTIVE COMPATIBILITY AND PARTIAL IMPLEMENTATION

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. We begin with introducing the *canonical transfers*, $t^* = (t_i^*(\cdot))_{i \in I}$, which are defined as follows: for each $i \in I$ and $m \in \Theta$,

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

In the following, we will refer to the pair (d, t^*) as the *canonical direct mechanism.*[11]

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 (Lemma 1). Hence, the canonical transfers characterize the mechanisms which may achieve partial implementation in the belief-free sense. As discussed, in the present context the designer knows that agents "commonly believe in identicality," and hence our analysis of partial implementation relies on the less demanding notion of incentive compatibility that we introduced in Definition 2. Nonetheless, as shown by the next result, the canonical transfers are still without loss of generality for partial $\mathcal{B}^{id}$-Implementation:

Theorem 1 (Partial Implementation)**.** *Under the maintained assumptions, d is partially $\mathcal{B}^{id}$-implementable if and only if (d, t^*) is $\mathcal{B}^{id}$-incentive compatible.*

Theorem 1 implies that, under the $\mathcal{B}^{id}$-restrictions, in order to decide whether partial implementation is possible, there is no reason to consider transfers other than the canonical ones. As we will see, this is not 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, $U_i^*(m; \theta)$, which only depend on the allocation function (d) and on the agents' preferences (v). First note that, under the maintained assumptions, the canonical direct mechanism induces twice differentiable payoff functions. 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.

Corollary 1 (Partial implementability and the canonical payoffs)**.**

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

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

11. The term "canonical mechanism" is traditionally used to refer to Maskin's mechanism for full implementation (Maskin, 1999). 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 mechanism are pervasive in the partial implementation literature. Examples arise, among others, in auctions (Myerson, 1981; Dasgupta and Maskin, 2000; Segal, 2003; Li, 2017), in pivot mechanisms (Milgrom, 2004; Jehiel and Lamy, 2018), in public goods problems (Green and Laffont, 1977; Laffont and Maskin, 1980), the one-dimensional results of Jehiel and Moldovanu (2001). Lemma 1 in Ollár and Penta (2017) generalizes the earlier results in the papers above. The term *canonical direct mechanism* was first used with this acceptation in Ollár and Penta (2017).

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: As the next example shows, there are environments in which (d, t^*) satisfies the second-order conditions in expectation, for all beliefs consistent with the $\mathcal{B}^{id}$ restrictions, but not in the ex-post sense:

Example 2. Consider an environment with three agents, $I = \{1, 2, 3\}$, with types $\theta_i \in [-1, 1]$ and valuations $v_i(x, \theta) = (\theta_i + \theta_i(\theta_j - \theta_k))x$ for all $i \in I$, where $x \in \mathbb{R}$, and consider the allocation rule $d(\theta) = \sum_{i=1}^{3} \theta_i$. In this environment, the second-order derivative of the payoff functions induced by the canonical transfers are the following:

$$\frac{\partial^2 U_i^*(m; \theta)}{\partial^2 m_i} = -2(1 + m_j - m_k) + (1 + m_j - m_k) = -(1 + m_j - m_k),$$

which, at the truth-telling profile $m = \theta$, is equal to:

$$\frac{\partial^2 U_i^*(\theta; \theta)}{\partial^2 m_i} = -(1 + \theta_j - \theta_k),$$

Since this term is positive at some $\theta \in \Theta$, truthful reporting is not optimal at all states. On the other hand, t^* ensures $\mathcal{B}^{id}$-incentive compatibility, since at the truthtelling profile,

$$\frac{\partial^2 \mathbb{E}^{b_{\theta_i}}(U_i^*(m_i, \theta_{-i}; \theta))}{\partial^2 m_i} = -1 < 0 \quad \text{for all } m_i \text{ and for all } b_{\theta_i} \in \mathcal{B}_{\theta_i}^{id}.$$

Hence, (d, t^*) is $\mathcal{B}^{id}$-IC, but not ep-IC. It follows that, with these preferences, this allocation rule is partially $\mathcal{B}^{id}$-implementable, but not belief-free implementable.

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, t^*. Similar to the way that *ex-post* monotonicity (of d) and single-crossing (of v) are sufficient for ep-IC, one can show that if interim monotonicity and single-crossing are satisfied *for all beliefs* consistent with the $\mathcal{B}^{id}$-restrictions, then the sufficient condition in part (ii) of Corollary 1 also holds, and hence they provide sufficient conditions for partial $\mathcal{B}^{id}$-implementation.[12]

The intuition for the result in Theorem 1 is the following: under the $\mathcal{B}^{id}$-restrictions, types do not differ in terms of their beliefs (*i.e.* $B_{\theta_i}^{id} = B_{\theta_i'}^{id}$ for all $\theta_i, \theta_i' \in \Theta_i$), and hence beliefs cannot be used to separate types, beyond what can be achieved without exploiting them. Thus, relative to the belief-free case, the role of the belief-restriction $\mathcal{B}^{id}$ is limited to relaxing the incentive compatibility constraint that the canonical transfers need to satisfy (from ex-post, to $\mathcal{B}^{id}$-IC), but it cannot be further leveraged to improve the design of transfers, to screen types.

The fact that $B_{\theta_i}^{id} = B_{\theta_i'}^{id} =: B_i^{id}$ for all $\theta_i, \theta_i' \in \Theta_i$ also has the following interesting implication, which in fact emerges from the proof of Theorem 1: For every $\mathcal{B}^{id}$-IC (d, t), and for any belief consistent with the $\mathcal{B}^{id}$-restrictions, the expected payment from every type of every agent at the truthtelling profile is the same as in the canonical mechanism (up to a constant). Formally:

Proposition 1 ("Payoff Equivalence" for $\mathcal{B}^{id}$-restrictions)**.** *If (d, t) is $\mathcal{B}^{id}$-IC, then for all $b \in B_{\theta_i}^{id}$, $\exists \kappa \in \mathbb{R}$ such that $\mathbb{E}^b(t_i(\theta_i, \theta_{-i})) = \mathbb{E}^b(t_i^*(\theta_i, \theta_{-i})) + \kappa$, for all $\theta_i \in \Theta_i$.*

12. Example 2 above is an instance of an environment with a (ex-post) monotonic allocation rule, in which the single crossing condition holds in expectation, for all $b_i \in B_{\theta_i}^{id}$, but not in the ex-post sense.

This result is an extension of the revenue-equivalence theorem, from the standard case of independent common prior, to the $\mathcal{B}^{id}$-restrictions. To understand this result, note that both the $\mathcal{B}^{id}$-restrictions and models of independent common prior share the feature that an agent's beliefs (a set, or a singleton) about others' types are the same for all his types. As further discussed in Ollár and Penta (2021), this property of generalized independence is key to revenue equivalence.

4. FULL IMPLEMENTATION

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 Environment)**.**

(i) Single-crossing environment: *for all* i and (x,θ), preferences are single crossing such that $\frac{\partial^2 v_i}{\partial x \partial \theta_i}(x,\theta) > 0$, and allocation is monotonic such that $\partial d/\partial \theta_i > 0$,

(ii) Public concavity environment: *for all* i, $\partial^2 v_i/\partial^2 x$ *and* $\partial d/\partial \theta_i$ *are constant in* θ, *and for all* i *and* j, $\partial^2 v_i/\partial x \partial \theta_j$ *is constant in* (x,θ).

We say that (d,v) is an environment with single crossing and public concavity (SC-PC) if it is both (i) and (ii).

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.[13] Compared with these applications, Definition 6 also accommodates more general dependence on x, as long as the concavity and the cross derivatives are public information.

These assumptions have two important consequences: Part (i) is a standard condition for ex-post incentive compatibility, which ensures in particular that partial $\mathcal{B}^{id}$-implementation is possible; Part (ii) ensures that, in the canonical direct mechanism, the 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 that $\partial^2 U_i^*/\partial^2 m_i \neq 0$. 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 $|\xi_1| \leq |\xi_2| \leq \cdots \leq |\xi_n|$. In SC-PC environments, these properties of the second-order derivatives of the payoff functions hold for all *transfers with constant curvature*, *i.e.* such that $\frac{\partial^3 t_i}{\partial m_i \partial m_j \partial m_k} = 0$ for all $i, j, k \in I$.

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. 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.

13. Quadratic-linear models are frequent in the literature of networks (*e.g.* Ballester *et al.*, 2006; Bramoullé and Kranton, 2007; Bramoullé *et al.*, 2014; Galeotti *et al.*, 2020), social interactions models (Blume *et al.*, 2015), markets with network externalities (*e.g.* Fainmasser and Galeotti, 2015), divisible good auctions (*e.g.* Wilson, 1979), and public goods (*e.g.* Duggan and Roberts, 2002).

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.) When the transfers in question are the canonical ones, t^*, then we write SE^* instead of SE^{t^*}. For example, in SC-PC settings, the canonical transfers t^* induce the following matrix of strategic externalities: for all (m,θ),

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

The next result shows that strategic externalities are key for full implementation. In particular, it shows that whether a strictly $\mathcal{B}^{id}$-IC transfer scheme t achieves full implementation, depends on the properties of two matrices which are closely related to $SE^t(m,\theta)$. Such matrices are obtained by focusing on the largest and smallest externalities across the domain, respectively, normalized by the smallest and largest concavity in the domain. Formally, let $|SE^t_{\max}|$ and $|SE^t_{\min}|$ be such that $|SE^t_{\max}|_{ii} = |SE^t_{\min}|_{ii} = 0$ for each i and, for each i and $j \neq i$, let $|SE^t_{\max}|_{ij} := \frac{\max_{(m,\theta)\in\Theta\times\Theta} |\partial^2 U_i^t(m,\theta)/\partial m_i \partial m_j|}{\min_{(m,\theta)\in\Theta\times\Theta} |\partial^2 U_i^t(m,\theta)/\partial^2 m_i|}$ and $|SE^t_{\min}|_{ij} := \frac{\min_{(m,\theta)\in\Theta\times\Theta} |\partial^2 U_i^t(m,\theta)/\partial m_i \partial m_j|}{\max_{(m,\theta)\in\Theta\times\Theta} |\partial^2 U_i^t(m,\theta)/\partial^2 m_i|}$. For any square matrix $A \in \mathbb{R}^{n\times n}$, we let $\rho(A)$ denote the *spectral radius* of A, *i.e.* the largest absolute value of its eigenvalues.[14] The next lemma formalizes the connection between the spectral radius of the $|SE^t_{\max}|$ and $|SE^t_{\min}|$-matrices and full $\mathcal{B}^{id}$-implementation:

Lemma 1 (Spectral radius and full $\mathcal{B}^{id}$-implementation)**.** *If t is $\mathcal{B}^{id}$-IC, then*

(i) $\rho(|SE^t_{\max}|) < 1$ *implies that t fully $\mathcal{B}^{id}$-implements d,*
(ii) $\rho(|SE^t_{\min}|) \geq 1$ *implies that t does not fully $\mathcal{B}^{id}$-implement d.*

First note that, if t is such that $SE^t(m,\theta)$ is constant in (m,θ) (as is the case, for instance, in SC-PC environments and transfers with constant curvature), then $|SE^t_{\max}| = |SE^t_{\min}| \equiv |SE^t|$, and then this Lemma implies that a transfer scheme t fully $\mathcal{B}^{id}$-implements d if and only if $\rho(|SE^t|) < 1$. Intuitively, the reason for this result is that eigenvalues in general describe the properties of iterated matrices. For strategic externality matrices, this amounts to describing the iterations of best replies which are implicit in the rationalizability operator. The condition that the spectral radius is smaller than one determines whether the transfers induce contractive best replies, and hence a unique rationalizable profile.[15] Incentive Compatibility—which is assumed in the Lemma—in turn ensures that such a unique profile is actually the truthful revelation profile. Since, in general, strategic externalities may vary over the domain, the necessary and sufficient conditions in the Lemma refer to the lower and upper bounds of such externalities, *i.e.* respectively to the $|SE^t_{\min}|$ and $|SE^t_{\max}|$-matrices.

As discussed, $\mathcal{B}^{id}$-IC is a necessary condition for full $\mathcal{B}^{id}$-implementation. Hence, we turn next to the implications of $\mathcal{B}^{id}$-IC for the mechanism's strategic externalities:

14. If A is such that $A_{ij} = \infty$ for some ij-entry, we let $\rho(A) := \lim_{K\to\infty} \rho(A_K)$, where A_K is s.t. $[A_K]_{ij} := K$ if $A_{ij} = \infty$ and $[A_K]_{ij} := A_{ij}$ otherwise.

15. Results analogous to Lemma 1 can be stated for other belief restrictions too, in that the spectral radius condition can be shown to characterize contractiveness of best replies in general games with payoff uncertainty. Other known conditions, such as diagonal dominance, are easier to check but only sufficient.

Lemma 2. *If 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 $j, k \neq i$,*

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

These conditions are also sufficient in SC-PC, when t has constant curvature.

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.

The intuition for this result, which is formalized by Lemmas 3 and 4 in Appendix A, is the following: by Lemma 3, the only way in which the designer can exploit the information on agents' beliefs to design $\mathcal{B}^{id}$-incentive compatible mechanisms, is to correct the baseline canonical transfers by adding a belief-dependent term which can be chosen for instance to minimize the spectral radius of the strategic externality matrix. In order to preserve incentive compatibility, however, the designer must know the expected value of this corrective term—formally, a function of the opponents' types—at the truthful strategy profile, for all beliefs that agents might have about others' types. Under the $\mathcal{B}^{id}$, essentially the only restriction which holds for all beliefs of all types is the idea that any player i regards the types of any two players as identically distributed. Hence, the only functions of the opponents' types whose expectation is known to the designer, regardless of which beliefs among those in $\mathcal{B}^{id}$ are entertained by the agents, are functions for which any "increase" on the effect of one opponent's type, must be offset by a commensurate "decrease" of some other opponent's type (cf. Lemma 4). The overall expectation of this corrective term must thus ensure a *rebalance* of the effects across the opponents, at least at profiles of identical types, which overall implies the constraint on the strategic externalities in the result above (cf. Appendix A).

The general design principle that emerges from combining Lemmas 1 and 2 is that the designer should seek to minimize the spectral radius of the $|SE^t_{\max}|$-matrix, subject to the constraints imposed by $\mathcal{B}^{id}$-IC (and, particularly, by Lemma 2). Such constraints imply that the designer may only *redistribute*, not reduce, the total strategic externalities induced by the canonical direct mechanism. In SC-PC environments and with quadratic transfers (which imply, in particular, that the SE^t-matrix is constant in (m, θ)), the conditions in Lemma 2 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 of the SE^*-matrix, $(\xi_i)_{i \in I}$, which in turn are uniquely pinned down by environment, (v, d). Our main result does not restrict the transfers to be quadratic, but it is nonetheless useful to consider that case. With such a restriction, in SC-PC settings, the design problem of identifying the transfers $\hat{t}$ that achieve full $\mathcal{B}^{id}$-implementation whenever some transfers in the same class do, is equivalent to a problem of minimizing the spectral radius subject to preserving the row-sums. This is formalized by the next result, which follows directly from Lemma 1:

Corollary 2. *A $\mathcal{B}^{id}$-IC transfer scheme $\hat{t}$ solves the "design problem" in the sense above if and only if the associated matrix of strategic externalities solves the following program:*

$$\begin{aligned} \min_{A} \quad & \rho(|A|) \\ \textit{subject to} \quad & A_{ii} = 0 \quad \textit{for every } i \in I, \\ & \sum_{j \neq i} A_{ij} = \xi_i \quad \textit{for every } i \in I. \end{aligned}$$

Moreover, the set of quadratic transfers that achieve full $\mathcal{B}^{id}$-implementation in a given environment (v, d) consists of all $\mathcal{B}^{id}$-IC transfers t whose matrix of strategic externalities, SE^t, satisfies the constraints in this program and is such that $\rho(|SE^t|) < 1$.

This result formalizes the connection between the full implementation and the network design problems that we discussed in the introduction and in Section 2. The second part of the result, that characterizes the set of transfers that achieve full implementation, is useful to think about other desiderata that one can impose, beyond full implementation. In Ollár and Penta (2022b), for instance, we consider the problem of a designer who wishes to fulfil other robustness criteria, besides full $\mathcal{B}^{id}$-implementation. In that case, the program can be adapted by replacing the minimization of ρ (which corresponds to identifying the "most contractive" transfers, that achieve full implementation whenever possible—what is needed to identify general conditions on (v, d) for implementability in the next subsection) with some other objective, tailored to the specific desiderata, and adding the requirement $\rho(|SE^t|) < 1$ to the constraints of the program. In this sense, the connection between the approaches can prove fruitful for further questions of implementation. We discuss this point further in Section 4.3.[16]

The uniqueness results above are also related to the literature on rationalizability and on global stability in dynamical systems. As we explained, the matrix of strategic externalities is key to uniqueness. The literature on dominance solvability provides some insights in this sense, but mainly for complete information games (Moulin, 1984). One may intuit that uniqueness of rationalizability is related to global stability, meaning that it is guaranteed if the dynamical system which describes the iterations of best replies is globally stable. We give broad conditions under which this intuition is valid and extends to incomplete information environments with belief-restrictions.[17] For instance, with quadratic transfers and under the SC-PC restriction, the matrix of strategic externalities determines a linear dynamical system which describes the relevant best-reply sets given the belief-restrictions. Unique rationalizability is equivalent to the global stability of this system which, in turn, is characterized by the largest absolute eigenvalue (the spectral radius). Given that a $\mathcal{B}^{id}$-incentive compatible t already ensures that truthful revelation has the best-reply property, uniqueness (and, hence, full $\mathcal{B}^{id}$-implementation) is achieved if and only if the spectral radius of the associated matrix of strategic externalities is less than one. As mentioned, outside of this special case, the resulting dynamical system is not necessarily linear, and there may be a gap between necessary and sufficient conditions. Our general conditions therefore involve upper and lower bound matrices of strategic externalities (Lemma 1).

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 which can be used to identify whether or not full $\mathcal{B}^{id}$-Implementation is possible. To this end, we introduce the *loading transfers*. As illustrated in part (iii) of Example 1, 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

16. In Ollár and Penta (2022a), we show that the program in Corollary 2 can be rewritten as an *optimal transport problem* with a nonlinear cost function, thereby providing a connection and also a novel result from the viewpoint of that literature (see, for instance, Ekeland, 2010; Daskalakis *et al.*, 2017; Kattwinkel *et al.*, 2022, and chapters in Galichon, 2018).

17. Ollár and Penta (2017) already noted the relevance of strategic externalities for uniqueness of rationalizability in incomplete information games and for general belief restrictions, but they only focused on sufficient conditions. The spectral radius results that we provide in this paper (also applied in Ollár and Penta, 2022b) are novel.

with the smallest canonical externalities (given the relabelling above, these are agents 1 and 2). Formally, the *loading transfers* $(t_i^l)_{i\in I}$ are defined as follows: for each $i \in I$ and $m \in M_i \times M_{-i}$,

$$t_i^l(m) = \underbrace{t_i^*(m)}_{\text{canonical transfers}} + \underbrace{L_i^l(m_{-i})m_i}_{\substack{\text{redistribution of}\\ \text{canonical externalities}}}, \tag{4.2}$$

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

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

First, it can be checked that these transfers ensure $\mathcal{B}^{id}$-IC (cf. Lemma 3 in Appendix A). 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)**.** *Fix an SC-PC* (v, d).

(i) d is fully $\mathcal{B}^{id}$-implementable if and only if it is fully $\mathcal{B}^{id}$-implemented by t^l.
(ii) d is fully $\mathcal{B}^{id}$-implementable if and only if $|\xi_1\xi_2| < 1$.

Before discussing the logic of the proof, first note that the condition in Part (ii) is a constraint on the canonical externalities and it is equivalent to requiring that the preference interdependence of agents 1 and 2 be sufficiently small. Formally: $|\xi_1\xi_2| < 1$ if and only if $|\sum_{j\neq 1} \frac{\partial^2 v_1}{\partial x \partial \theta_j} \cdot \sum_{j\neq 2} \frac{\partial^2 v_2}{\partial x \partial \theta_j}| < \frac{\partial^2 v_1}{\partial x \partial \theta_1} \cdot \frac{\partial^2 v_2}{\partial x \partial \theta_2}$.

Part (i) of the theorem derives from the following observations. First, in SC-PC environments, the loading transfers are strictly $\mathcal{B}^{id}$-IC and induce constant strategic externalities. Hence (by Lemma 1) they achieve full implementation if and only if $\rho(|SE^l|) < 1$. Second, by examining the iterative rounds of rationalizability, we show that all $\mathcal{B}^{id}$-IC transfers induce sets of rationalizable profiles that contain the ones induced by the loading transfers. The steps in the proof also imply that $\rho(|SE^l|) \leq \rho(|SE^t_{\max}|)$ for all $\mathcal{B}^{id}$-IC transfers t. The reason why t^l achieves the minimal spectral radius among the possible strategic externality matrices is perhaps best illustrated in Example part (iii) of Example 1. To concentrate all of i's strategic externalities on the opponent with the smallest $|\xi_j|$ (that is, 1's externalities on agent 2, and all other agents' externalities on agent 1) decreases the impact of the flow of externalities on best replies. This impact is key to

rationalizability and is linear algebraically represented by the iterated matrix of strategic externalities. The intuition, as our proof shows, indeed carries over to all rounds of iterations and is optimal in the here considered broader space of transfers. Hence, if $\rho(|SE^l|) \geq 1$, then full implementation necessarily fails for all other transfers too. On the other hand, if $\rho(|SE^l|) < 1$, then full implementation is possible, and it is achieved, for example, by t^l.

Part (ii) follows from the fact that $\rho(|SE^l|) < 1$ if and only if $|\xi_1\xi_2| < 1$. As we explained, this implies that the possibility of achieving full $\mathcal{B}^{id}$-implementation depends on the canonical externalities of the two agents with the smallest canonical externalities (equivalently, of the two smallest levels of preference interdependence). Thus, full implementation is possible if and only if the combined effect of these two agents' canonical externalities is not too large, regardless of the strength of the preference interdependence of the other agents and their number. At the extreme, if an environment involves just *one* agent with private values, then full implementation is possible.

4.3. *Discussion*

It may be useful at this point to discuss how the results above and our approach more broadly compare with the typical approach in the literature on full implementation.

On the restriction to direct mechanisms. The first, main point of departure, is our restriction to direct mechanisms. As it is well known, this restriction is without loss of generality for the purpose of partial implementation, but it may make the task of achieving full implementation harder. Note, however, that if this means that the necessity part of our characterization may be stronger than what could be identified with unrestricted mechanisms, the opposite is true for the sufficiency direction: the fact that we provide remarkably permissive results, *despite* the restriction to the class of mechanisms, strengthens those results. There are, however, other reasons for restricting the class of mechanisms.

First, classical results on full implementation typically involve unrealistically complicated mechanisms, which have been criticized for providing limited economic insight (*e.g.* Jackson, 1992). The artificial nature of those mechanisms, and the related emphasis in the literature on necessity results, in our view explain why the full implementation approach has overall been less successful than the partial implementation one, in terms of delivering clear qualitative insights on the design of real world mechanisms. 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 favours the interpretability of the results and hence pushes a bit further Jackson's concern for economic 'relevance' of full implementation theory.

This restriction also enables us to uncover what features of an incentive compatible transfer scheme—namely, 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, while maintaining the same fundamental structure as the transfer schemes for partial implementation, which have a clear economic interpretation and may thus be more portable to the real world. One by-product of this is the possibility of recasting the implementation problem in terms of a *weighted network design* problem, thereby connecting full implementation with more familiar concepts of mainstream economics, such as networks and externalities. As we further discuss in the conclusions, we think that this connection may benefit both the implementation and the network literature.

Constructive insights for transfer design. The classical approach in the full implementation literature considers preferences which are not necessarily quasi-linear, and focuses on

social choice functions (SCFs) $f : \Theta \to Y$, where Y denotes the space of outcomes (see, *e.g.* Bergemann and Morris, 2009a). With quasi-linear preferences, $Y = X \times \mathbb{R}^n$, and hence such characterizations can be used to check whether a given $f(\cdot) = (d(\cdot), t(\cdot))$ is implementable by a direct mechanism (and hence, similar to Lemma 1, whether a given t implements d), but they do not provide insights on *how to design* transfers for full implementation. Since we are interested in this kind of constructive insights, we adopted here the standard setup of the partial implementation literature, only taking $d : \Theta \to X$ as given, and letting the designer choose $t : \Theta \to \mathbb{R}^n$. Second, as we already discussed, the restriction to *direct mechanisms* also entails some loss of generality for full implementation, but in these environments it enables an easier comparison with the partial implementation literature, and to focus on the structural properties of the transfer schemes. The emphasis on the ability to generate insights for the design of transfers represents an important point of departure from the full implementation literature, and is also reflected in the kind of conditions we provide (cf. Lemma 1).[18] By referring to the eigenvalues of the strategic externality matrices, these conditions also enabled us to draw a bridge between full implementation and networks (*e.g.* Elliott and Golub, 2019; Galeotti *et al.*, 2020), which may prove fertile for both strands of the literatures (these points are further discussed in the Conclusions).

Alternative robustness criteria. In many settings, it may be desirable to ensure that the implementing mechanism does not rely too heavily on agents' behaviour exactly coinciding with that entailed by the maintained assumptions on their preferences and rationality. In Ollár and Penta (2022b), 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 agents' reports. Such "mistakes" can be interpreted as stemming from agents' *slightly faulty* behaviour (similar to Eliaz, 2002), or as a shorthand to account for possible misspecification of their preferences in the model.[19] Intuitively, the extreme hierarchical structure of the strategic externalities induced by the loading transfers may entail an unnecessarily high fragility of the system in this context, and hence if full $\mathcal{B}^{id}$-implementation is possible, other transfer schemes may provide a better compromise between the various desiderata.

The robustness notion in Ollár and Penta (2022b) reflects the idea 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 a solution concept, $R_i^{F_\varepsilon}$, which characterizes the behavioural implication of assuming common knowledge that a subset F of players may

18. As a comparison, Bergemann and Morris (2009a) characterize belief-free rationalizable implementation via direct mechanisms in environments with monotone aggregators (*i.e.* such that $\forall i,\ v_i(x,\theta) = w_i(x, h_i(\theta))$ for some $w_i : X \times \mathbb{R} \to \mathbb{R}$ and $h_i : \Theta \to \mathbb{R}$ strictly increasing in θ_i) in terms of strict ep-IC and the following "contraction property" (Def.5, p.1183, ibid.): $\forall \beta : \Theta \to 2^\Theta$ s.t. $\theta \in \beta(\theta)$ for all θ, but $\beta(\theta') \neq \{\theta'\}$ for some θ', there exists i, θ_i and $\theta_i'' \in \beta_i(\theta_i)$ with $\theta_i'' \neq \theta_i$ such that, for all θ_{-i} and $\theta'_{-i} \in \beta_{-i}(\theta_{-i})$, $sign(\theta_i - \theta_i'') = sign(h_i(\theta_i, \theta_{-i}) - h_i(\theta_i'', \theta'_{-i}))$. With more general preferences and with unrestricted mechanisms, the analogous condition for belief-free rationalizability is *robust monotonicity* (Bergemann and Morris, 2011): $\forall \beta : \Theta \to 2^\Theta$ s.t. $\exists \theta, \theta' : \theta' \in \beta(\theta)$ and $f(\theta) \neq f(\theta')$, $\exists i, \theta_i, \theta_i'' \in \beta_i(\theta_i)$ s.t. $\forall \theta_{-i}$ and $\psi \in \Delta(\beta_{-i}^{-1}(\theta'_{-i}))$, $\exists y \in Y$: (i) $\sum_{\theta'_{-i} \in \beta_{-i}^{-1}(\theta_{-i})} \psi(\theta_{-i}) u_i(y, (\theta_i, \theta_{-i})) > \sum_{\theta'_{-i} \in \beta_{-i}^{-1}(\theta_{-i})} \psi(\theta_{-i}) u_i(f(\theta_i', \theta'_{-i}), (\theta_i, \theta_{-i}))$; and (ii) $\forall \theta_i'', u_i(f(\theta_i'', \theta'_{-i}), (\theta_i'', \theta'_{-i})) > u_i(y, (\theta_i'', \theta'_{-i}))$. Similar characterizations, alternative to Lemma 1, could be provided for full $\mathcal{B}^{id}$-implementation.

19. Ollár and Penta (2022b) only focus on the implementation of efficient allocation rules, and for settings in which the distributions of other players' types have identical means (but not necessarily identical distributions). Their results, however, can be extended to more general allocation rules and the $\mathcal{B}^{id}$-restrictions of this paper.

be ε-faulty, across all agents and all configurations of the set of faulty agents. As shown in Ollár and Penta (2022b), in SC-PC environments with symmetric aggregators, among the set of transfers that achieve full implementation, the transfers that are the most robust in this sense are characterized by an *equal* redistribution of the given total strategic externalities, among the opponents of every player.

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 of these agents being faulty. In that case, the loading transfers would perform rather 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 is especially critical for the mechanism.

Beyond SC-PC environments. Ollár and Penta (2022a) also consider environments that do not satisfy the SC-PC restriction. In those settings, the key difficulty is that the canonical strategic externality matrix may not be constant over the domain of types and reports, and hence operationalizing the general principle of redistributing the strategic externalities subject to the incentive compatibility constraints requires tracing how they vary over the entire domain. One way to approach this problem is to construct the modification of the baseline transfers, based on a midpoint between the lowest and highest strategic externalities generated by the environment. Theorem 3 in Ollár and Penta (2022a) shows that such a design strategy ensures full $\mathcal{B}^{id}$-implementation, if the strategic externalities at such a midpoint are not too large for at least two agents, and as long as the strategic externalities do not vary too much across the entire domain. So, in that sense, the main qualitative insight obtained under the SC-PC restriction carry over to general settings, provided that the design of the loading transfer is suitably generalized.

5. CONCLUSIONS

This paper continues a long tradition of works on implementation, that have taken up Wilson (1987) and Jackson (1992)'s call for a greater "relevance" of full implementation theory, through a repeated weakening of common knowledge assumptions on the environment, and the exploration of restricted classes of mechanisms.[20] In this paper, we focused specifically on implementation via transfers that only elicit agents' payoff-relevant information, under weak common knowledge assumptions that reflect a natural economic idea: namely, that agents' types are drawn from an identical distribution. Our main results characterize the transfer schemes that achieve, respectively, partial and full implementation whenever possible, under such a "common belief in identicality" restriction, as well as the conditions on the agents' preferences and on the allocation rule under which these notions of implementation are possible. Despite the restriction to the class of mechanisms, which ensures a clear economic interpretation of the results, we uncovered surprisingly permissive results. For instance, we showed that the possibility of

20. For instance, under standard common knowledge assumptions, Jackson (1992) studied implementation via bounded mechanisms, and Bergemann and Morris (2009a), Oury and Tercieux (2012) studied implementation via direct mechanisms; With unrestricted mechanisms, Bergemann and Morris (2011), Müller (2020) studied implementation in belief-free settings; papers that included both non-standard (weak) common knowledge restrictions and restricted mechanisms, include Bergemann and Morris (2009a, 2009b) and Ollár and Penta (2017); etc.

full implementation is determined 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.

Our analysis also revealed that the joint restrictions on the mechanisms and on the common knowledge assumptions impose a peculiar mathematical structure on the implementation problem, which enabled us to recast the mechanism design problem as one of "optimally" designing a network of strategic externalities, subject to suitable constraints. The objective of this design exercise (dictated by the aim of identifying the transfer schemes that achieve full implementation whenever possible) is to minimize the spectral radius of the matrix of strategic externalities; the constraints (which are dictated by incentive compatibility under "common belief in identicality") require preserving the total level of such externalities. Aside from the implementation results in a strict sense, this formulation of the problem generates further insights, which may prove valuable for other strands of the literature.

For instance, Galeotti *et al.* (2020) recently studied the problem of optimally intervening on the nodes of a game with networked externalities. The interventions considered in that paper concern the idiosyncratic (non-strategic) components of players' preferences, taking as given a network of externalities which is assumed to induce contractive best replies and uniqueness of equilibrium. In contrast, our analysis concerns the design of the very network of strategic externalities (subjects to certain constraints, as we discussed in the previous paragraph). The objective of minimizing the spectral radius, within a class of networks of strategic externalities, may prove useful in itself, as several properties of a networked economy may be related to the spectral radius of its matrix of strategic externalities: for instance, when the spectral radius is less than one, it is closely related to Cournot stability of the associated Nash equilibrium (cf. Moulin, 1984). Our solution to the spectral radius-minimization problem is thus also informative about structural properties of networks, well beyond the full implementation problem from which it stemmed in this paper. In fact, the solution we identified (namely, the *star network* that describes the strategic externalities induced by the loaded transfers in Theorem 2) has interesting structural features, which we think are quite revealing from a pure network perspective.

Our characterization of full implementation in terms of a spectral radius condition on a suitable matrix of strategic externalities is also related to Elliott and Golub (2019)'s characterization of efficient allocations in economies with networked externalities, which is also based on a spectral radius condition of a matrix of externalities. The main difference is that their spectral radius condition refers to a matrix of *payoff* externalities, which are captured by the first-order derivatives of agents' payoff functions. In contrast, our condition refers to a matrix of *strategic* externalities, which describes how players' best responses are affected by others' actions, and hence are described by the second-order derivatives of agents' payoff function. Nonetheless, both papers provide clear cases in point on how a network approach may shed a new light on classical problems, and enable novel results. For the problem, we consider, specifically, this connection favours a more clear integration of full implementation theory with more familiar concepts of mainstream economics, such as transfers schemes, networks and externalities. The other important difference is that Elliott and Golub (2019) consider complete information settings, whereas we allow for incomplete information with both private and interdepenndent values. From this viewpoint, our results also contribute to the growing literature on network games with incomplete information (*e.g.*Calvó-Armengol and De Martí, 2007; Galeotti *et al.*, 2009; Calvó-Armengol *et al.*, 2015; De Martí and Zenou, 2015; Golub and Morris, 2017; Myatt and Wallace, 2019; Leister, 2020; Leister *et al.*, 2020). With respect to this literature, our results

on the spectral radius of the strategic externality matrix provide sufficient conditions for equilibrium uniqueness (as well as a characterization of uniqueness of rationalizable solutions) for incomplete information games, with both private and interdependent values.

With respect to robust mechanism design, this paper contributes to the literature which has explored environments with limited information about agents' beliefs, intermediate between standard Bayesian settings (*e.g.* Postlewaite and Schmeidler, 1986; Jackson, 1991), and the belief-free benchmark (*e.g.*Bergemann and Morris, 2005, 2009a). Compared with Ollár and Penta (2017), which introduced general belief-restrictions and studied sufficient conditions under which full implementation may be achieved via a *reduction* of strategic externalities,[21] this paper represents an example of a specific belief restriction based on an interesting class of economic environments (namely, the common belief assumption only about identicality). As discussed, these restrictions turn out to induce a tractable mathematical structure, that translates into a different design principle—namely, a *redistribution* of the strategic externality—that also enables strong implementation results. Interesting directions for future research include exploring other belief restrictions, similarly motivated to capture primitive qualitative properties of beliefs, without imposing the standard common prior assumption. For instance, it would be interesting to study implementation under qualitative restrictions such as independence, affiliation, positive correlation, etc., without the extra common knowledge assumptions of standard models.

In a similar spirit, it would also be important to explore different restrictions to the class of mechanisms, especially tailored to specific environments, or by imposing specific properties on the mechanism.[22] This is important because, if direct mechanisms are ideal to provide economic insights on incentive compatibility, they are not always the simplest to implement in practice. In some settings, indirect yet simpler mechanisms may also achieve implementation (auctions are a classical example). While our results are silent on such specific indirect mechanisms, the general idea of focusing on the matrix of strategic externality, and to pursue contractive best replies via the addition of belief-dependent terms (cf. Appendix 2), is based on general game theoretic principles which are applicable to other baseline mechanisms as well.

Finally, we note that the characterization of full implementation under common belief in identicality may have potentially interesting implications from a broader market design perspective: for instance, if full implementation cannot be achieved for a given set of agents, then adding two more agents whose preferences do not depend much on others' information would suffice to make full implementation possible. In practical problems of market design, however, these possibility results ought to be weighted against other considerations, which may entail a different structure for the implementing mechanism. One such example, which we discussed in Section 4.3, is robustness with respect to "mistakes in play" (Ollár and Penta, 2022b), which suggests a more even redistribution of the strategic externalities. Exploring further desiderata and robustness criteria is another interesting direction for future research.

21. For instance, Ollár and Penta (2017) show that (under certain preference restrictions) strategic externalities can always be eliminated in common prior models with independent or affiliated types and hence full implementation be achieved in (interim) dominant strategies. When strategic externalities cannot completely eliminated, they provide sufficient conditions for contractive best replies, so as to obtain uniqueness of the rationalizable strategy profiles.

22. In recent years, many papers have re-visited standard implementation problems imposing extra desiderata on the mechanisms. Deb and Pai (2017), for instance, pursue symmetry of the mechanism; Mathevet (2010) and Mathevet and Taneva (2013) pursue supermodularity; Healy and Mathevet (2012) and Ollár and Penta (2017) pursue contractiveness. In the classical literature, the broader idea of modifying ex-post incentive compatible transfers using information about beliefs has been pursued, among others, by d'Aspremont *et al.* (1979), Arrow (1979), Cremer and McLean (1988).

APPENDIX A. ON PARTIAL $\mathcal{B}^{ID}$-IMPLEMENTATION

A.1. *On the Proof of Theorem 1: main ideas*

The key for the proof of Theorem 1 is provided by the following Lemma:

Lemma 3 ($\mathcal{B}^{id}$-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 $m \in M \equiv \Theta$,*

$$t_i(m) = \underbrace{t_i^*(m) + \tau_i(m_{-i})}_{\substack{\text{belief-free transfers} \\ \text{(ep-IC characterization)}}} + \underbrace{\int_{\underline{\theta}}^{m_i} K_i(s_i, m_{-i})\, ds_i}_{\text{belief-based component}} \tag{A.1}$$

where $\tau_i : M_{-i} \to \mathbb{R}$ and $K_i : M \to \mathbb{R}$ are differentiable functions and K_i is such that:[23]

$$\mathbb{E}^{b_{\theta_i}}(K_i(\theta_i, \theta_{-i})) = 0 \quad \text{for all } \theta_i \text{ and for all } b_{\theta_i} \in B^{id}_{\theta_i}. \tag{A.2}$$

[Sufficiency:] *If (d,t) is twice differentiable, t satisfies (A.1) and (A.2), and the resulting payoffs are such that $\mathbb{E}^{b_{\theta_i}}(\partial^2 U_i(m_i, \theta_{-i}; \theta)/\partial^2 m_i) < 0$ for all m_i and $b_{\theta_i} \in \mathcal{B}^{id}_{\theta_i}$, then (d,t) is $\mathcal{B}^{id}$-IC.*

Equation (A.1) implies that, as far as $\mathcal{B}^{id}$-IC is concerned, it is without loss of generality to design transfers starting from the canonical transfers, and then adding a *belief-based* term $K_i : M \to \mathbb{R}$. The sense in which the extra component is "belief-dependent" is clarified by the condition in Equation (A.2), which has to be satisfied for all beliefs consistent with $\mathcal{B}^{id}$. Note that any twice continuously differentiable mechanism is $\mathcal{B}^{id}$-IC 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 the $\mathcal{B}^{id}$ restrictions. Moreover, 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) $\mathbb{E}^{b_{\theta_i}}(\partial U_i(\theta_i, \theta_{-i}; \theta_i, \theta_{-i})/\partial m_i) = 0$ and (ii) $E^{b_{\theta_i}}(\partial^2 U_i(\theta_i, \theta_{-i}; \theta_i, \theta_{-i})/\partial^2 m_i) \leq 0$. But if t partially implements d, then by Lemma 3 it can be written as in (A.1), 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

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

Condition (A.2) in Lemma 3 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 (A.2) also ensures that the second term of right-hand side of the second equation is zero, for all beliefs $b_{\theta_i} \in B^{id}_{\theta_i}$. Hence, the first- and second-order conditions are met in (d,t) if and only if they are met in the canonical direct mechanism. Theorem 1 expands on this observation.

A.2. *Incentive compatibility and moment conditions*

Further intuition on the belief-based components in condition (A.2) of Lemma 3 can be gathered by looking at the special case in which the K_i function can be written as $K_i(m) = L_i(m_{-i}) -$

23. For any $f : \Theta \to \mathbb{R}$, $\theta_i \in \Theta_i$ and $b_{\theta_i} \in B^{id}_{\theta_i}$, we let $\mathbb{E}^{b\theta_i}(f(\theta_i, \theta_{-i})) := \int_{\Theta_{-i}} f(\theta_i, \theta_{-i})\, db_{\theta_i}$.

$f_i(m_i)$, for some $L_i : \Theta_{-i} \to \mathbb{R}$ and $f_i : \Theta_i \to \mathbb{R}$. Then, the expected value condition (A.2) can be written as

$$\mathbb{E}^{b_{\theta_i}}(L_i(\theta_{-i})) = f_i(\theta_i) \quad \text{for all } \theta_i \text{ and for all } b_{\theta_i} \in B^{id}_{\theta_i}. \tag{A.3}$$

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 (A.3) 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 7. 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* the $\mathcal{B}^{id}$-restrictions if it satisfies (A.3) 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 3, Eq. (A.1) specializes to

$$t_i(m) = \underbrace{t_i^*(m) + \tau_i(m_{-i})}_{\substack{\text{characterization}\\ \text{of ep-IC transfers}}} + \underbrace{L_i(m_{-i})m_i - \int^{m_i} f_i(s_i)\,ds_i}_{\text{moment condition-based term}}. \tag{A.4}$$

This is precisely the class of transfers for which Ollár and Penta (2017) provide sufficient conditions for full implementation.[24] By Lemma 3, there may exist incentive compatible transfers which cannot be written as in Equation (A.4), since not all functions $K_i : \Theta \to \mathbb{R}$ in that Lemma are equivalent to moment conditions in the sense of Definition 7. 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 4 (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*

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 $\theta_{-i} \in \Theta_{-i}$.*

Proof of Lemma 4. 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.

24. 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 of this paper, the conditional averages of types are neither common knowledge nor known to the designer.)

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 (2.1) to be the independent product of an identical distribution over $[\underline{\theta}, \overline{\theta}]$. It can be shown that results analogous to Lemma 3 obtain for $\mathcal{B}^{iid}$-restrictions, as well as a characterization analogous to Lemma 4, with the only difference that part 3 of Lemma 4 is not required. Intuitively, the stronger information that the designer has about agents beliefs in $\mathcal{B}^{iid}$, compared with $\mathcal{B}^{id}$, 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.

A.3. *Proofs*

Proof of Lemma 3. Assume that t ensures $\mathcal{B}^{id}$-incentive compatibility which, by t's differentiability and the applicability of Leibniz's rule, means that for all i and θ_i

$$\mathbb{E}^{b_{\theta_i}}(\partial(v_i(d(m_i, \theta_{-i}), \theta) + t_i(m_i, \theta_{-i}))/\partial m_i)|_{m_i=\theta_i} = 0 \quad \text{for all } b_{\theta_i} \in B^{id}_{\theta_i}.$$

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

$$\mathbb{E}^{b_{\theta_i}}(\partial(t_i(m_i, \theta_{-i}) - t_i^*(m_i, \theta_{-i}))/\partial m_i)|_{m_i=\theta_i} = 0 \quad \text{for all } b_{\theta_i} \in B^{id}_{\theta_i}.$$

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_i and let this part be $\tau_i(m_{-i}) := D_i(m) - \int_{\underline{\theta}}^{m_i} \frac{\partial D_i}{\partial m_i}(s_i, m_{-i})\, ds_i$, and further let $K_i(m) := \partial D_i(m)/\partial m_i$ for all m. Then, the transfer t_i takes the form $t_i(m) = t_i^*(m) + \tau_i(m_{-i}) + \int_{\underline{\theta}}^{m_i} K_i(s_i, m_{-i})\, ds_i$ for all m and K_i satisfies the expected value condition in (A.2). Moreover, if (d, t) is twice differentiable, then by the definition of canonical transfers t^* is twice differentiable, and thus K_i is differentiable. Since K_i is differentiable in all its arguments, τ_i is twice differentiable, which completes the proof of the necessity part of this Lemma.

If (d, t) is twice differentiable and t satisfies the characterization in (A.1) and the expected value condition in (A.2), then

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

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: If $K_i : M \to \mathbb{R}$ satisfies condition (A.2), then for all θ_i $\mathbb{E}^{b_{\theta_i}}(K_i(m_i, \theta_{-i})) = 0$ for all m_i and for all $b_{\theta_i} \in B^{id}_{\theta_i}$.

To show this step, recall the expected value condition in (A.2), $\mathbb{E}^{b_{\theta_i}}(K_i(\theta_i, \theta_{-i})) = 0$ 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

$p \in B^{id}_{m_i}$ for all $m_i \in [\underline{\theta}, \overline{\theta}]$, that is $\mathbb{E}^p(K_i(m_i, \theta_{-i})) \equiv 0$ as a function of m_i, and this holds for any $p \in B^{id}_{\theta_i}$, which proves this Step.[25]

To show the Theorem, if (d, t) partially implements d, then by Lemma 3, t can be written as in (A.1), and hence—letting U^* denote the payoff function of the canonical direct mechanism—for any θ_i and $b_{\theta_i} \in B^{id}_{\theta_i}$:

$$\begin{aligned}\mathbb{E}^{b_{\theta_i}}(\partial U_i(m_i, \theta_{-i}; \theta_i, \theta_{-i})/\partial m_i) &= \mathbb{E}^{b_{\theta_i}}(\partial U_i^*(m_i, \theta_{-i}; \theta_i, \theta_{-i})/\partial m_i) + \mathbb{E}^{b_{\theta_i}}(K_i(m_i, \theta_{-i})) \\ &= \mathbb{E}^{b_{\theta_i}}(\partial U_i^*(m_i, \theta_{-i}; \theta_i, \theta_{-i})/\partial m_i),\end{aligned}$$

where the latter is a well-defined function of m_i. Hence, for all types, the set of optimal reports for all beliefs in $\mathcal{B}^{id}$ are equivalent in (d, t) and (d, t^*), which proves this Theorem.

APPENDIX B. PROOFS OF RESULTS FROM SECTION 4

Proof of Lemma 1. [26] (i) (Sufficiency: eigenvalue condition for full implementation.)[27] Fix θ_i $in(\underline{\theta}, \overline{\theta})$ and examine the kth round of eliminations: fix $m_i \in R_i^k(\theta_i)$. Thus for m_i, there exists a conjecture which supports m_i as a best reply and is concentrated on R_{-i}^{k-1}. (Recall that a conjecture of agent i is a probability distribution over $M_{-i} \times \Theta_{-i}$.) Let this conjecture be μ_L. At the same time, 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; and pick μ_T such that $\text{marg}_{\Theta_{-i}}\mu_T = \text{marg}_{\Theta_{-i}}\mu_L$.

We use the notation $EU_i^{\mu}(m_i; \theta_i)$ to 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

$$\begin{aligned}0 &= \partial_i EU_i^{\mu_L}(m_i; \theta_i) - \partial_i EU_i^{\mu_T}(\theta_i; \theta_i) \\ &= \underbrace{\partial_i EU_i^{\mu_L}(m_i; \theta_i) - \partial_i EU_i^{\mu_L}(\theta_i; \theta_i)}_{\text{difference due to own action}} + \underbrace{\partial_i EU_i^{\mu_L}(\theta_i; \theta_i) - \partial_i EU_i^{\mu_T}(\theta_i; \theta_i)}_{\text{difference due to external (others') actions}}.\end{aligned}$$

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

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

Second, let $b_l \leq b_u$ be the boundary points of the set of $k-1$-rationalizable messages of θ_i. If m_i is such that $m_i = b_l$, then, because m_i is best reply,

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

25. Note that K_i need not be the 0 function. For example, $(\theta_j - \theta_k)\theta_i$ satisfies the expected value condition for all identical distributions. Moreover, if K_i^1 and K_i^2 satisfy the condition, then any linear combination $\alpha K_i^1 + \beta K_i^2$ satisfies the condition as well.

26. The sufficiency of the eigenvalue condition for full implementation and the points in this lemma are stated for identical distributions but, as it is clear from the proofs, they generalize beyond $\mathcal{B}^{id}$ to arbitrary belief restrictions.

27. Recall that to extend the spectral radius operator to the affinely extended reals, given a non-negative matrix A, we let A_K be such that $[A_K]_{ij} := K$ if $A_{ij} = \infty$ and $[A_K]_{ij} := A_{ij}$ otherwise. We let $\rho(A) := \lim_{K \to \infty} \rho(A_K)$. Beyond the standard extensions of operators, we adopt the understanding that $0/0 = \infty$ and $\infty/\infty = \infty$.

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

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

After examining the signs of $\partial^2_{ii} EU_i^{\mu_L}(s_i;\theta_i)$ and the respective signs of $(m_i-\theta_i)$ in the latter two cases, we can summarize that for all, 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

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

From this, for each agent j and round k, letting $l_j^k := \max_{\theta_j, m_j \in R_j^k(\theta_j)} |\theta_j - m_j|$, and letting $l_j^0 = l = \overline{\theta} - \underline{\theta}$, we have

$$|m_i-\theta_i| \leq \frac{\sum_{j\neq i} |\partial^2_{ij} U_i(\theta_i, s_{-i};\theta)| l_j^{k-1}}{|\partial^2_{ii} EU_i^{\mu_L}(s_i;\theta_i)|} \leq [|SE^t_{\max}| l^{k-1}]_i.$$

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

$$|m_i-\theta_i| \leq [|SE^t_{\max}| l^{k-1}]_i \leq [|SE^t_{\max}||SE^t_{\max}| l^{k-2}]_i \leq \cdots \leq [|SE^t_{\max}|^k \mathbf{1} l]_i.$$

Since $\rho(|SE^t_{\max}|) < 1$, we have $|SE^t_{\max}|^k \to \mathbf{0}$, and thus full $\mathcal{B}^{id}$-implementation follows.

(ii) (Necessity: eigenvalue condition for failure of full implementation.) The key step for this part is to show that for all rounds k there is an agent i such that for all types θ_i, there is a kth round $\mathcal{B}$-rationalizable message—a message in $R_i^k(\theta_i)$—which falls outside a positive measure open set around θ_i. In particular, consider the largest subset of agents whose interaction matrix in $|SE^t_{\min}|$ is irreducible and features no 0 eigenvalues. (Such subset $I_E \subseteq I$ of the agents exists and, since $\rho(|SE^t_{\min}|) > 1$ and the diagonal contains 0s, it has at least two agents.) We maintain the ordering of the agents and use notation E for this irreducible block of $|SE^t_{\min}|$. We will show next, that for each round k for some $i \in I_E$, there is a best reply outside the open set $(\theta_i \pm [E \cdot \mathbf{l}^{k-1}_{\min,E}]_i) \cap \text{int cl } R_i^{k-1}(\theta_i)$. The notation $\mathbf{l}^k_{\min,E}$ is such that: for each agent $j \in I_E$ and round k, let $l^k_{j,\min,E} := \inf_{\theta_j} \min\{\sup_{m_j \in R_j^k(\theta_j); m_j \leq \theta_j} (\theta_j - m_j); \sup_{m_j \in R_j^k(\theta_j); m_j > \theta_j} (m_j - \theta_j)\}$, and let $l^0_{j,\min} := \overline{\theta} - \underline{\theta}$.[28]

To show this, consider an internal type θ_i for some agent $i \in I_E$. First notice that the previous statement is true for $k = 1$. Moreover, since the truthtelling profile is never eliminated, $R_i^k(\theta_i)$ is always non-empty. Next, consider round k and let m_i be a message that is best reply to a conjecture $\mu_L^E \in \Delta(M_{-i} \times \Theta_{-i})$ that is consistent with $\mathcal{B}$, with round $k-1$ rationalizability; and is such that (i) for all $j \in I_E$, μ_L^E it places probability one on positive misreports that are $l^k_{j,\min,E}$ apart from θ_j if the absolute smallest $\partial^2_{ij} U_i$ is positive and places probability one on negative misreports if it is negative; and (ii) for all $j \notin I_E$, μ_L^E it places probability one on the true type θ_j being reported. (Here, we write U_i for the payoffs resulting from the given t.) Now, if the considered m_i is an extremal point of cl $R_i^{k-1}(\theta_i)$, then we are done. However,

28. The intuition for $\mathbf{l}^k_{\min,E}$ is that it is a vector that keeps track of the minimum distance of worst-case positive or negative misreports; resulting from interactions based on the irreducible E, among agents in I_E.

if it is an internal point, then $\partial^2_{ii} EU_i^{\mu_L^E}(m_i;\theta_i) \leq 0$ and there is a small ε such that the modified function $EU_i^{\mu_L^E,\varepsilon} := EU_i^{\mu_L^E}(s_i;\theta_i) - \varepsilon(s_i - m_i)^2$ admits m_i as a strict optimizer. For the difference between the derivative of this function and the expected utility at the corresponding truthtelling conjecture; using mean value theorems, we can establish that for m_i there exist messages s_i, s_{-i}, m_{-i} such that m_j reflects the distances in μ_L^E and

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

Taking absolute values and lower bounding by the relevant minimum partial derivatives, we get that for all small $\varepsilon > 0$

$$(-\partial^2_{ii} EU_i^{\mu_L^E}(s_i;\theta_i) + \varepsilon)|(m_i - \theta_i)| \geq \sum_{j\neq i} \min_{m,\theta} |\partial^2_{ij} U_i(m;\theta)| l^{k-1}_{j,\min},$$

which further implies for such m_i that

$$|m_i - \theta_i| \geq \frac{\sum_{j\neq i} \min_{m,\theta} |\partial^2_{ij} U_i(m;\theta)| l^{k-1}_{j,\min}}{|\partial^2_{ii} EU_i^{\mu_L^E}(s_i;\theta_i)|} \geq [E\mathbf{l}^{k-1}_{\min}]_i.$$

Thus, summarizing this, for each k, there is a kth round rationalizable message that is outside the set $(\theta_i \pm [E\cdot \mathbf{l}^{k-1}_{\min,E}]_i) \cap \operatorname{int}\operatorname{cl} R_i^{k-1}(\theta_i)$, which when iterated gives that it is outside the set $(\theta_i \pm [E^k \cdot \mathbf{l}^0_{\min,E}]_i) \cap (\underline{\theta}_i, \overline{\theta}_i)$. Iteratively, one can see that $\mathbf{l}^0_{\min,E}, \mathbf{l}^1_{\min,E}$ are strictly positive. Assuming that $\mathbf{l}^{k-1}_{\min,E}$ is strictly positive, and by the irreducibility of the non-negative E, we have that $\mathbf{l}^k_{\min,E}$ is strictly positive. From this, we can see that if the spectral radius $\rho(|SE^t_{\min}|) \geq 1$, then the sequence $\{E^k\}_{k=1}^\infty$ of non-negative matrices is bounded away from $\mathbf{0}$ and thus there are rationalizable messages for agents in I_E which are distinct from their true types; and thus full $\mathcal{B}$-implementation fails.

Proof of Lemma 2. First, we give a characterization of belief-based terms under $\mathcal{B}^{id}$. (The following step is again used in Theorem 2 below.)

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 (A.2) if and only if it can be written as

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

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

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

To show this step, assume, that K_i satisfies the expected value condition in (A.2) 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 (A.2) 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.[29]

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

$$H_k(\theta^1, \theta^2, \ldots, \theta^2) - H_k(\theta^2, \theta^2, \ldots, \theta^2) \neq H_k(\theta^1, \theta^1, \ldots, \theta^1) - H_k(\theta^2, \theta^1, \ldots, \theta^1). \quad \text{(B.5)}$$

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 $k \neq i$, $\text{marg}_{\Theta_k} p^{\text{corr}} = \text{marg}_{\Theta_k} p^{\text{indep}}$, and on θ^1: $\text{marg}_{\Theta_k} p^{\text{corr}}(\{\theta_k = \theta^1\}) = \text{marg}_{\Theta_k} p^{\text{indep}}(\{\theta_k = \theta^1\}) = 0.5$, and on θ^2: $\text{marg}_{\Theta_k} p^{\text{corr}}(\{\theta_k = \theta^2\}) = \text{marg}_{\Theta_k} p^{\text{indep}}(\{\theta_k = \theta^2\}) = 0.5$. Observe that both p^{corr} and p^{indep} are available under the belief restrictions $\mathcal{B}^{id}$, formally, $p^{\text{corr}} \in B^{id}_{\theta_i}$ and $p^{\text{indep}} \in B^{id}_{\theta_i}$. For ease of notations, let p be a probability measure over $[\underline{\theta}, \overline{\theta}]$ such that $p(\{\theta_k = \theta^1\}) = p(\{\theta_k = \theta^2\}) = 0.5$ and let f_p be p's distribution function.

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

$$\begin{aligned}\mathbb{E}^{p^{\text{corr}}}(H_k(\theta_{-i})) &= \int_{\Theta_{-i}} H_k(\theta_{-i})\, dp^{\text{corr}} = \int_{\underline{\theta}}^{\overline{\theta}} H_k(\theta, \theta, \ldots, \theta) f_p\, d\theta \\ &= 0.5 H_k(\theta^1, \theta^1, \ldots, \theta^1) + 0.5 H_k(\theta^2, \theta^2, \ldots, \theta^2).\end{aligned}$$

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

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

The last negation follows from Equation (B.5), which recall was the consequence of non-separability, and this negation implies that $\mathbb{E}^{p^{\text{indep}}}(H_k(\theta_{-i})) \neq \mathbb{E}^{p^{\text{corr}}}(H_k(\theta_{-i}))$, 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

29. This proof is a proof by coupling, a proof technique here applied to distributions over continuous support.

agents. Formally, consider a belief b_{θ_i} such that, for some $\theta \in [\underline{\theta}, \overline{\theta}]$, the probability $b_{\theta_i}(\{\theta_j = \theta \text{ for all } j \neq i\})$ 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 $0 \equiv \sum_{k=1}^{\infty} \mathbb{E}^{b_{\theta_i}}(H_k(\theta_{-i}))\theta_i^k$ and thus for any k $\mathbb{E}^{b_{\theta_i}}(H_k(\theta_{-i})) = 0$. 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

$$\begin{aligned}\mathbb{E}^{b_{\theta_i}}(K_i(\theta)) &= \int_{\Theta_{-i}} \sum_{k=1}^{\infty} H_k(\theta_{-i})\theta^k \, db_{\theta_i} = \int_{\Theta_{-i}} \sum_{k=1}^{\infty} \sum_{j\neq i} H_{kj}(\theta_j)\theta^k \, db_{\theta_i} \\ &= \sum_{k=1}^{\infty} \sum_{j\neq i} \left[\int_{\Theta_j} H_{kj}(\theta_j) \, d \operatorname{marg}_{\Theta_j} b_{\theta_i} \right] \theta^k \end{aligned} \tag{B.6}$$

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 (B.6) is

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

and the two conditions,

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

and thus K_i satisfies the expected value condition under $\mathcal{B}^{id}$ and thus proves the characterization result in this Step.

If K_i satisfies the expected value condition in (A.1), then based on the characterization in Step 1 of Proof of Lemma 1, we have

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

If (d, t) is $\mathcal{B}^{id}$-IC, then by Lemma 3, there exist $K_i : M \to \mathbb{R}$ which satisfies the expected value condition in (A.1); and is such that $\partial U_i^t(m; \theta)/\partial m_i = \partial U_i^*(m; \theta)/\partial m_i + K_i(m_i, m_{-i})$. This equation and the two properties above imply the points of the lemma. Finally, the characterization's application to SC-PC environments and constant curvature in t proves this Lemma.

Proof of Theorem 2. Consider the loading transfers t^l. It is useful to characterize the resulting sets of rationalizable 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 $R_i^{id,k}(\theta_i \mid t^l)$ is a closed interval around θ_i.[30]

To show this, note that by construction $\theta_i \in R_i^{id,k}(\theta_i \mid t^l)$. By the boundedness (which is implied by the differentiability) of v, d, t^l and by the SC-PC conditions, the best reply map is single valued and continuous. Using this, one can show in a proof by induction that for every k, the set of conjectures which are consistent with the $k-1$-st round and with identicality is closed in the sup-norm. By continuity of the best reply function, the set of best replies is closed, and thus $R_i^{id,k}(\theta_i \mid t^l)$ is a closed interval which contains θ_i.

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 $\xi_{ij} := \partial^2 U_i^*/(\partial m_i \partial m_j) = -(\partial^2 v_i/\partial x \partial \theta_j) \cdot (\partial d/\partial \theta_i)$, $\xi_i := \sum_{j \neq i} \xi_{ij}/\xi_{ii}$ and $|\xi_1| \leq |\xi_2| \leq \cdots \leq |\xi_n|$. Recall that under SC-PC, these canonical externalities and the cross-derivatives in the resulting payoff functions in the loading mechanism (d, t^l) 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 (d, t^l) 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:

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

In the second round of iterations:

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

30. Note that this property is stated for t^l but it extends in SC-PC to every bounded and smooth $\mathcal{B}^{id}$-IC t.

In the third round of iterations:

$$\left(SE^l\right)^3 = \begin{bmatrix} 0 & \xi_1^2\xi_2 & 0 & \dots & 0 \\ \xi_1\xi_2^2 & 0 & 0 & \dots & 0 \\ \xi_1\xi_2\xi_3 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \xi_1\xi_2\xi_n & 0 & 0 & \dots & 0 \end{bmatrix} \quad \text{and} \quad [R_i^{id,3}(\theta_i \mid t^l)]_{i\in I} = \begin{bmatrix} [\theta_1 \pm \xi_1^2\xi_2] \cap R_1^{id,2}(\theta_1 \mid t^l) \\ [\theta_2 \pm \xi_1\xi_2^2] \cap R_2^{id,2}(\theta_2 \mid t^l) \\ [\theta_3 \pm \xi_1\xi_2\xi_3] \cap R_3^{id,2}(\theta_3 \mid t^l) \\ \vdots \\ [\theta_n \pm \xi_1\xi_2\xi_n] \cap R_n^{id,2}(\theta_n \mid t^l) \end{bmatrix}.$$

And so on, in the kth 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

$$[R_i^{id,k}(\theta_i \mid t^l)]_{i\in I} = \begin{bmatrix} [\theta_1 \pm \xi_1^{k/2}\xi_2^{k/2}] \cap R_1^{id,k-1}(\theta_1 \mid t^l) \\ [\theta_2 \pm \xi_1^{k/2}\xi_2^{k/2}] \cap R_2^{id,k-1}(\theta_2 \mid t^l) \\ [\theta_3 \pm \xi_1^{k/2}\xi_2^{k/2-1}\xi_3] \cap R_3^{id,k-1}(\theta_3 \mid t^l) \\ \vdots \\ [\theta_n \pm \xi_1^{k/2}\xi_2^{k/2-1}\xi_n] \cap R_n^{id,k-1}(\theta_n \mid t^l) \end{bmatrix},$$

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

$$[R_i^{id,k}(\theta_i \mid t^l)]_{i\in I} = \begin{bmatrix} [\theta_1 \pm \xi_1^{(k+1)/2}\xi_2^{(k-1)/2}] \cap R_1^{id,k-1}(\theta_1 \mid t^l) \\ [\theta_2 \pm \xi_1^{(k-1)/2}\xi_2^{(k+1)/2}] \cap R_2^{id,k-1}(\theta_2 \mid t^l) \\ [\theta_3 \pm \xi_1^{(k-1)/2}\xi_2^{(k-1)/2}\xi_3] \cap R_3^{id,k-1}(\theta_3 \mid t^l) \\ \vdots \\ [\theta_n \pm \xi_1^{(k-1)/2}\xi_2^{(k-1)/2}\xi_n] \cap R_n^{id,k-1}(\theta_n \mid t^l) \end{bmatrix}.$$

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. Similarly, it holds for every *odd round* of iteration that 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 $R_i^{id}(\theta_i \mid t^l) = \{\theta_i\}$ 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, $R_2^{id,k}(\theta_2 \mid t^l) = [\theta_2 \pm (\overline{\theta} - \underline{\theta})] \cap [\underline{\theta}, \overline{\theta}] = [\underline{\theta}, \overline{\theta}]$, 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 will lead to the characterizing inequalities in part 2 of this Theorem).

Recall that in this proof for Part 1, 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 (4.2). The if part is straightforward. The only if part, relies on the following step, which shows that a $\mathcal{B}^{id}$-IC transfer scheme ensures that the step-by-step iterative eliminations result in sets of k-rationalizable strategies whose sizes reflect the canonical externalitites.

Step 3: (Iterations and canonical externalities, given $\mathcal{B}^{id}$.) 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

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

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

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

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

To show this step, fix $\theta_i\in(\underline{\theta},\overline{\theta})$ and fix some type $\theta_o\in(\underline{\theta},\overline{\theta})$ and some message $m_o\in(\underline{\theta},\overline{\theta})$ 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, assigns probability 1 to the event that all opponents types are $\theta_j=\theta_o$ and report their true types. Let this—concentrated truth-reporting—conjecture be $\mu_T\in\Delta(M_{-i}\times\Theta_{-i})$. There exists also a message of i which is best-reply to the conjecture that assigns probability 1 to the event that opponents are $\theta_j=\theta_o$ and report m_o regardless of their types. Denote this undominated strategy by m_i and let this—concentrated m_o-reporting—conjecture be $\mu_L\in\Delta(M_{-i}\times\Theta_{-i})$. 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

$$\begin{aligned}0&=\partial_i EU_i^{\mu_L}(m_i;\theta_i)-\partial_i EU_i^{\mu_T}(\theta_i;\theta_i)=\partial_i EU_i^{*\mu_L}(m_i;\theta_i)-\partial_i EU_i^{*\mu_T}(\theta_i;\theta_i)\\&=\underbrace{\partial_i EU_i^{*\mu_L}(m_i;\theta_i)-\partial_i EU_i^{*\mu_L}(\theta_i;\theta_i)}_{\text{difference due to own action}}+\underbrace{\partial_i EU_i^{*\mu_L}(\theta_i;\theta_i)-\partial_i EU_i^{*\mu_T}(\theta_i;\theta_i)}_{\text{difference due to external (others') actions}},\end{aligned}$$

where the first equality holds because of the canonical representation of (d,t) in Lemma 3, the of belief-based terms in step 1 of Theorem 1 and because of the conjectures μ_T and μ_L are constructed such that they satisfy identicality on the margins of the messages too.

In this step, 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 (θ_i,s_{-i}^o).

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

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

and there exists s_o such that

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

Note that any kth-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,

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

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

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

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

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

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

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

which gives that, for some profile,

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

For this step, let $l_{i,o}^{0,+} = l_{i,o}^{0,-} := \overline{\theta} - \underline{\theta}$. To measure the size of higher-than-true misreports, let $l_{i,o}^{k,+} := \min_{j\neq i} \max_{\theta_j} \max_{m_j \in Rat_j^{\mathcal{B}^{id},k}(\theta_j)} (m_j - \theta_j)$ and similarly, for lower-than-true misreports, let $l_{i,o}^{k,-} := \min_{j\neq i} \max_{\theta_j} \max_{m_j \in Rat_j^{\mathcal{B}^{id},k}(\theta_j)} (\theta_j - m_j)$.

We summarize the above 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

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

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 kth 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

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

which, completes the proof of this Step.

Step 3 as established above is the key step to the if and only if result. In words, it implies that in any $\mathcal{B}^{id}$-implementing direct mechanism, the externalities cannot 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 of the step-by-step iterations; for all $\mathcal{B}^{id}$-IC t. Next, the final step below formalizes the observation that it is the loading transfer scheme that minimizes the size of rationalizable sets, given the constraint on necessary externalitites and therefore leads to full implementation whenever that is possible.

Step 4: We use Step 3 of this proof to show that in every round k, for all i and θ_i, the set of rationalizable messages of the loading direct mechanism $R_i^{id,k}(\theta_i \mid t^l)$, which we characterized in Step 1, are contained in $R_i^{id,k}(\theta_i \mid t)$ for any partially implementing direct mechanism (d, t).

To show this, fix a direct mechanism (d, t). Under SC-PC environments, Step 3 implies that every k-rationalizable interval of θ_i of any implementing (d, t) direct mechanism contains the following set:

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

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 (d, t^l), where the latter sets are already given in Step 2 of this proof. In particular, for the first round of iteration,

$$[\theta_i - \xi_i, \theta_i + \xi_i] \cap [\underline{\theta}, \overline{\theta}] \subseteq R_i^{id,1}(\theta_i \mid t).$$

For the second round of iteration,

$$\begin{aligned} &[\theta_1 - \xi_1\xi_2, \theta_1 + \xi_1\xi_2] \cap [\underline{\theta}, \overline{\theta}] \subseteq R_i^{id,2}(\theta_i \mid t) \quad \text{if } i = 1 \\ \text{and} \quad &[\theta_i - \xi_i\xi_1, \theta_i + \xi_i\xi_1] \cap [\underline{\theta}, \overline{\theta}] \subseteq R_i^{id,2}(\theta_i \mid t) \quad \text{if } i \neq 1. \end{aligned}$$

For the third round of iteration,

$$\begin{aligned} &[\theta_1 - \xi_1(\xi_1\xi_2), \theta_1 + \xi_1(\xi_1\xi_2)] \cap [\underline{\theta}, \overline{\theta}] \subseteq R_i^{id,3}(\theta_i \mid t) \quad \text{if } i = 1 \\ \text{and} \quad &[\theta_i - \xi_i(\xi_1\xi_2), \theta_i + \xi_i(\xi_1\xi_2)] \cap [\underline{\theta}, \overline{\theta}] \subseteq R_i^{id,3}(\theta_i \mid t) \quad \text{if } i \neq 1. \end{aligned}$$

For the forth round of iteration,

$$\begin{aligned} &[\theta_1 - \xi_1(\xi_1\xi_2^2), \theta_1 + \xi_1(\xi_1\xi_2^2)] \cap [\underline{\theta}, \overline{\theta}] \subseteq R_i^{id,4}(\theta_i \mid t) \quad \text{if } i = 1 \\ \text{and} \quad &[\theta_i - \xi_i(\xi_1^2\xi_2), \theta_i + \xi_i(\xi_1^2\xi_2)] \cap [\underline{\theta}, \overline{\theta}] \subseteq R_i^{id,4}(\theta_i \mid t) \quad \text{if } i \neq 1. \end{aligned}$$

Observe that in these expressions on the left-hand side, the iterated sets derived in Step 3, for every k, coincide with the iterated rationalizable sets of the loaded direct mechanism (d, t^l), and thus by induction, for all k, $R_i^{id,k}(\theta_i|t^l) \subseteq R_i^{id,k}(\theta_i|t)$.[31] This latter holds for any partially implementing direct mechanism (d, t), which completes the proof of this Step.

Turning to Part 1, if t^l ensures full $\mathcal{B}^{id}$-implementation, then, clearly, d is fully $\mathcal{B}^{id}$-implementable. If the direct mechanism (d, t) achieves full $\mathcal{B}^{id}$-implementation, by the containment above, we must have that as $k \to \infty$, $|R_i^{id,k}(\theta_i|t^l)| \to 0$, and thus (d, t^l) achieves full $\mathcal{B}^{id}$-implementation too, which completes the proof of Part 1 of this Theorem. Applying Lemma 1 to the loaded externality matrix, completes Part 2 of this Theorem.

31. Notice that the matrix algebraic content of this latter line is that $|SE^l|^k\mathbf{1} \leq |SE^t_{\max}|^k\mathbf{1}$ for all k. By Gelfand's formula (that is by $\rho(A) = \lim_{k\to\infty} ||A^k||^{1/k}$) and by the definition of the norm, we have that $\rho(|SE^l|) \leq \rho(|SE^t|)$ for every t that is $\mathcal{B}$-IC.

Acknowledgments. Earlier versions of this paper circulated under the title "Implementation via Transfers with Identical but Unknown Distributions". We are grateful to Pierpaolo Battigalli, Eddie Dekel, Philippe Jehiel, George Mailath, and Rakesh Vohra for their comments. We also thank seminar audiences at Yale, LSE, Bocconi, Caltech, MIT-Harvard, Michigan, Oxford, Cambridge, Carnegie-Mellon, Penn State, Univ. of Edinburgh, Bar-Ilan Univ, Tel-Aviv Univ., Hebrew Uni. of Jerusalem, ICEF (Moscow), and participants to the 2020.1 World Congress of the Game Theory Society (Budapest), the 2019 Warwick Economic Theory Workshop (Warwick Univ.) the Workshop on New Directions in Mechanism Design (Stony Brook, 2019), and the Canadian Economic Theory Conference. The BSE benefited from the financial support of the Spanish Ministry of Economy and Competitiveness, through the Severo Ochoa Programme for Centres of Excellence in R&D (CEX2019-000915-S). Antonio Penta acknowledges the financial support of the European Research Council, Starting Grant 759424.

REFERENCES

ARROW, K. (1979), "The Property Rights Doctrine and Demand Revelation under Incomplete Information", in *Economics and Human Welfare* (New York: Academic Press), 23–39.
ARTEMOV, G., KUNIMOTO, T. and SERRANO, R. (2013), "Robust Virtual Implementation with Incomplete Information: Towards a Reinterpretation of the Wilson Doctrine", *Journal of Economic Theory*, **148**, 424–447.
BALLESTER, C., CALVÓ-ARMENGOL, A. and ZENOU, Y. (2006), "Who's Who in Networks. Wanted: The Key Player", *Econometrica*, **74**, 1403–1417.
BATTIGALLI, P. and SINISCALCHI, M. (2003), "Rationalization and Incomplete Information", *Advances in Theoretical Economics*, **3**, 1073.
BECKER, A., DECKERS, T., DOHMEN, T., FALK, A. and KOSSE, F. (2012), "The Relationship Between Economic Preferences and Psychological Personality Measures", *Annual Review of Economics*, **4**, 453–478.
BERGEMANN, D. and MORRIS, S. (2005), "Robust Mechanism Design", *Econometrica*, **73**, 1771–1813.
BERGEMANN, D. and MORRIS, S. (2009a), "Robust Implementation in Direct Mechanisms", *Review of Economic Studies*, **76**, 1175–1204.
BERGEMANN, D. and MORRIS, S. (2009b), "Robust Virtual Implementation", *Theoretical Economics*, **4**, 45–88.
BERGEMANN, D. and MORRIS, S. (2011), "Robust Implementation in General Mechanisms", *Games and Economic Behavior*, **71**, 261–281.
BLUME, L. E., BROCK, W. A., DURLAUF, S. N. and JAYARAMAN, R. (2015), "Linear Social Interactions Models", *Journal of Political Economy*, **123**, 444–496.
BRAMOULLÉ, Y. and KRANTON, R. (2007), "Public Goods in Networks", *Journal of Economic Theory*, **135**, 478–494.
BRAMOULLÉ, Y., KRANTON, R. and D'AMOURS, M. (2014), "Strategic Interaction and Networks", *American Economic Review*, **104**, 898–930.
CALVÓ-ARMENGOL, A. and DE MARTÍ, J. (2007), "Communication Networks: Knowledge and Decisions", *American Economic Review*, **97**, 86–91.
CALVÓ-ARMENGOL, A., DE MARTÍ, J. and PRAT, A. (2015), "Communication and Influence", *Theoretical Economics*, **10**, 649–690.
CREMER, J. and MCLEAN, R. P. (1988), "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions", *Econometrica*, **56**, 1247–1257.
DASGUPTA, P. and MASKIN, E. (2000), "Efficient Auctions", *The Quarterly Journal of Economics*, **115**, 341–388.
DASKALAKIS, C., DECKELBAUM, A. and TZAMOS, C. (2017), "Strong Duality for a Multiple-Good Monopolist", *Econometrica*, **85**, 735–767.
D'ASPREMONT, C., CREMER, J. and GERARD-VARET, L-A. (1979), "Incentives and Incomplete Information", *Journal of Public Economics*, **11**, 25–45.
DEB, R. and PAI, M. M. (2017), "Discrimination via Symmetric Auctions", *American Economic Journal: Microeconomics*, **9**, 275–314.
DE MARTÍ, J. and ZENOU, Y. (2015), "Network Games with Incomplete Information", *Journal of Mathematical Economics*, **61**, 221–240.
DUGGAN, J. and ROBERTS, J. (2002), "Implementing the Efficient Allocation of Pollution", *American Economic Review*, **92**, 1070–1078.
EKELAND, I. (2010), "Notes on Optimal Transportation", *Economic Theory*, **42**, 437–459.
ELIAZ, K. (2002), "Fault Tolerant Implementation", *The Review of Economic Studies*, **69**, 589–610.
ELLIOTT, M. and GOLUB, B. (2019), "A Network Approach to Public Goods", *Journal of Political Economy*, **127**, 730–776.
FAINMASSER, I. P. and GALEOTTI, A. (2015), "Pricing Network Effects", *Review of Economic Studies*, **83**, 1–36.
GALEOTTI, A., GOLUB, B. and GOYAL, S. (2020), "Targeting Interventions in Networks", *Econometrica*, **88**, 2445–2471.
GALEOTTI, A., GOYAL, S., JACKSON, M. O., VEGA-REDONDO, F. and YARIV, L. (2009), "Network Games", *Review of Economic Studies*, **77**, 1175–1204.
GALICHON, A. (2018), *Optimal Transport Methods in Economics* (Princeton: Princeton University Press).
GOLUB, B. and MORRIS, S. (2017), "Expectations, Networks, and Conventions", arxiv.

GREEN, J. and LAFFONT, J. J.. (1977), "Characterization of Satisfactory Mechanisms for the Revelation of Preferences for Public Goods", *Econometrica*, **45**, 427–438.
HEALY, P. J. and MATHEVET, L. (2012), "Designing Stable Mechanisms for Economic Environments", *Theoretical Economics*, **7**, 609–661.
JACKSON, M. O. (1991), "Bayesian Implementation", *Econometrica*, **59**, 461–477.
JACKSON, M. O. (1992), "Implementation in Undominated Strategies: A Look at Bounded Mechanisms", *Review of Economic Studies*, **59**, 757–75.
JEHIEL, P. and LAMY, L. (2018), "A Mechanism Design Approach to the Tiebout Hypothesis", *Journal of Political Economy*, **126**, 735–760.
JEHIEL, P. and MOLDOVANU, B. (2001), "Efficient Design with Interdependent Valuations", *Econometrica*, **69**, 1237–1259.
KATTWINKEL, D., NIEMEYER, A., PREUSSER, J. and WINTER, A. (2022), "Mechanisms without Transfers for Fully Biased Agents" (Mimeo).
KUNIMOTO, T., SARAN, R. and SERRANO, R. (2021), "Interim Rationalizable Implementation of Functions" (Mimeo).
LAFFONT, J-J. and MASKIN, E. (1980), "A Differential Approach to Dominant Strategy Mechanisms", *Econometrica*, **48**, 1507–1520.
LEISTER, C. M. (2020), "Information Acquisition and Welfare in Network Games", *Games and Economic Behavior*, **122**, 453–475.
LEISTER, C. M., ZENOU, Y. and ZHUNG, J. (2020), "Social Connectedness and Contagion" (Mimeo).
LI, Y. (2017), "Approximation in Mechanism Design with Interdependent Values", *Games and Economic Behavior*, **103**, 225–253.
LIPNOWSKI, E. and SADLER, E. (2019), "Peer–Confirming Equilibrium", *Econometrica*, **87**, 567–591.
LOPOMO, G., RIGOTTI, L. and SHANNON, C. (2011), "Uncertainty in Mechanism Design" (Working Paper, University of Pittsburgh).
MASKIN, E. (1999), "Nash Equilibrium and Welfare Optimality", *The Review of Economic Studies*, **66**, 23–38.
MATHEVET, L. (2010), "Supermodular Mechanism Design", *Theoretical Economics*, **5**, 403–443.
MATHEVET, L. and TANEVA, I. (2013), "Finite Supermodular Design with Interdependent Valuations", *Games and Economic Behavior*, **82**, 327–349.
MILGROM, P. R. (2004), *Putting Auction Theory to Work* (Vancouver: Cambridge University Press).
MOULIN, H. (1984), "Dominance Solvability and Cournot Stability", *Mathematical Social Sciences*, **7**, 83–102.
MÜLLER, C. (2016), "Robust Virtual Implementation under Common Strong Belief in Rationality", *Journal of Economic Theory*, **162**, 407–450.
MÜLLER, C. (2020), "Robust Implementation in Weakly Perfect Bayesian Strategies", *Journal of Economic Theory*, **189**, 105038.
MYATT, D. P. and WALLACE, C. (2019), "Information Acquisition and Use by Networked Players", *Journal of Economic Theory*, **182**, 360–401.
MYERSON, R. B. (1981), "Optimal Auction Design", *Mathematics of Operations Research*, **6**, 58–73.
OLLÁR, M. and PENTA, A. (2017), "Full Implementation and Belief Restrictions", *American Economic Review*, **107**, 2243–2277.
OLLÁR, M. and PENTA, A. (2021), "Incentive Compatibility and Belief Restrictions' (Mimeo).
OLLÁR, M. and PENTA, A. (2022a), "A Network Solution to Robust Implementation: the Case of Identical but Unknown Distribution" (BSE Working Paper No. 1248).
OLLÁR, M. and PENTA, A. (2022b), "Efficient Implementation via Transfers: Uniqueness and Sensitivity in Symmetric Environments" (AEA: papers and proceeding).
OURY, M. and TERCIEUX, O. (2012), "Continuous Implementation", *Econometrica*, **80**, 1605–1637.
PENTA, A. (2015), "Robust Dynamic Mechanism Design", *Journal of Economic Theory*, **160**, 280–316.
POSTLEWAITE, A. and SCHMEIDLER, D. (1986), "Implementation in Differential Information Economies", *Journal of Economic Theory*, **39**, 14–33.
SEGAL, I. (2003), "Optimal Pricing Mechanisms with Unknown Demand", *The American Economic Review*, **93**, 509–529.
WILSON, R. (1979), "Auctions of Shares", *The Quarterly Journal of Economics*, **93**, 675–689.
WILSON, R. (1987), "Game-Theoretic Analysis of Trading Processes", in Bewley (ed.), *Advances in Economic Theory* (Cambridge University Press).