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

Mariann Ollár
University of Edinburgh

Antonio Penta
ICREA, UPF and Barcelona GSE

April 15, 2021

Abstract

We consider mechanism design environments in which agents commonly know that others' types are identically distributed, but without assuming that the actual distribution is common knowledge, nor that it is known to the designer *(common knowledge of identicality, CKI)*. Under these assumptions, we study problems of partial and full implementation, as well as robustness. First, we characterize the transfers which are incentive compatible under the CKI assumption, and provide necessary and sufficient conditions for partial implementation. Second, we characterize the conditions under which full implementation is possible via direct mechanisms, as well as the transfer schemes which achieve full implementation whenever it is possible. We do this by pursuing a *network approach*, which is based on the observation that the full implementation problem in our setting can be conveniently transformed into one of designing a network of strategic externalities, subject to suitable constraints which are dictated by the incentive compatibility requirements entailed by the CKI assumption. This approach enables us to uncover a fairly surprising result: the possibility of full implementation is characterized by the strength of the preference interdependence of the two agents with the least amount of preference interdependence, regardless of the total number of agents, and of their preferences. Finally, we study the robustness properties of the implementing transfers with respect to both misspecifications of agents' preferences and with respect to lower orders beliefs in rationality.

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

JEL: D62, D82, D83

1 Introduction

Many economic models assume that agents believe that the types of others are drawn from the same distribution. This is a natural way to represent situations in which agents regard each other

*Earlier versions of this paper circulated under the title "Implementation via Transfers with Identical but Unknown Distributions". We are grateful to Eddie Dekel, Philippe Jehiel, George Mailath, and Rakesh Vohra for their comments. We also thank seminar audiences at Bocconi, Caltech, MIT-Harvard, Michigan, Oxford, Carnegie-Mellon, Penn State, Univ. of Edinburgh, Bar-Ilan Univ, Tel-Aviv Univ., Hebrew Uni. of Jerusalem, ICEF (Moscow), and participants to the 2019 Warwick Economic Theory Workshop (Warwick Univ.) and to the Workshop on New Directions in Mechanism Design (Stony Brook, 2019), and the Canadian Economic Theory Conference. The GSE 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 also acknowledges the financial support of the European Research Council (ERC), ERC Starting Grant #759424.

as ex-ante symmetric from an informational viewpoint, or more broadly that they come from a common population. Standard modeling techniques, however, not only impose that the distribution of types is *identical* across agents, but also that it is common knowledge among them – and, in mechanism design, also known to the designer. But if *identicality* is a natural way to capture a basic qualitative property of these environments, *common knowledge* of the distribution is a different kind of assumption: not only is it strong and unlikely to be satisfied; it is also well-known to heavily affect the results.

A large and growing literature has taken up Wilson (1987) call for a "[...] repeated weakening of common knowledge assumptions [...]", and developed a robust approach to mechanism design (see, e.g., Bergemann and Morris (2005, 2009a)). In this paper we pursue the objectives of the *Wilson doctrine* in settings with informationally symmetric agents. More specifically, we maintain common knowledge that others' types are identically distributed, but *without* assuming a commonly known distribution. Under these assumptions, we study questions of both *partial* and *full implementation*: First, we characterize the transfers which are incentive compatible under the assumption of identicality, and provide necessary and sufficient conditions for partial implementation. Second, we characterize the conditions under which truthful revelation is *the only* solution under common knowledge of identicality, as well as the transfers which achieve it whenever possible (full implementation).[1] These results are enabled by a *network approach* we put forward (in many ways similar to the one pursued by Elliott and Golub (2019)), which is based on the observation that the full implementation problem in our setting can be conveniently mapped to a specific 'network design' problem, subject to suitable constraints which are dictated by the relevant incentive compatibility requirement. As we will discuss, this formulation also offers valuable insights for the literature on policy interventions on networks (e.g., Galeotti, Golub and Goyal (2020)), as well as on incomplete information in networks (e.g., Leister, Zenou and Zhou (2020)). Finally, we study the *robustness properties* of the implementing transfers with respect to the possibility that agents are 'slightly faulty' (e.g., Eliaz (2002) – or equivalently, that their preferences are slightly misspecified), and with respect to lower orders of rationality (which is also connected to recent work on level-k implementation by de Clippel et al. (2018, 2021)).

We start our analysis with the introduction of the *canonical transfers*. These transfers are pinned down by the necessary conditions for ex-post incentive compatibility, and hence they characterize the possibility of achieving partial implementation in belief-free settings (cf. Bergemann and Morris (2005)). Our first result shows that, when only common knowledge 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 implementation under common knowledge of identicality is as demanding as ex-post incentive compatibility: the latter is a strictly more demanding notion of implementation; nonetheless, in both settings the canonical transfers are all which needs to be considered to check if partial implementation is possible. Given these results, in analogy to standard methods on Bayesian incentive compatibility, necessary and sufficient conditions for partial implementation can be promptly obtained from studying the properties of the payoffs induced by the canonical transfers at the truthful profiles.

[1] Our notion of full implementation is based on a special version of Battigalli and Siniscalchi (2003)'s Δ-rationalizability. Special version of Δ-rationalizability have also been used in implementation theory by Bergemann and Morris (2009a), Oury and Tercieux (2012), Ollár and Penta (2017), Kunimoto, Saran and Serrano (2021) and Lipnowski and Sadler (2019) in static mechanisms, and in dynamic ones by Müller (2016, 2020) and (albeit in a different sense) Catonini (2021).

We move next to the analysis of *full implementation*, which is the main focus of this paper. Here, it is useful to recall the result by Bergemann and Morris (2009a), who also study full implementation via direct mechanisms, but in belief-free settings. They show that full belief-free implementation is possible (in which case it is achieved by the canonical transfers) if and only if the interdependence in agents' valuations is not too strong. Since preference interdependence is often too strong, this characterization is typically regarded as a negative result. As shown by Ollár and Penta (2017), the reason for this result is that, when preferences interdependence is strong, then the canonical transfers induce too strong *strategic externalities*, which in turn induce multiplicity and hence a failure of full implementation. Ollár and Penta (2017)'s idea is then to use information about agents' beliefs, if available, to design incentive compatible transfers which induce small strategic externalities (and hence uniqueness and full implementation) even when preference interdependencies are strong. They thus provided sufficient conditions on agents' beliefs so that the designer could engineer such weakening of the strategic externalities.[2] It turns out, however, that if only common knowledge of identicality is maintained, without assuming knowledge of the actual distribution of types, then Ollár and Penta (2017)'s design strategy cannot be pursued: more precisely, we show that, under common knowledge of identicality, any incentive compatible mechanism must display the same total level of strategic externalities as the canonical direct mechanism. Hence, in these settings, the designer may only pursue a *redistribution* – not a reduction – of the strategic externalities, which in turn are pegged to the level of preference interdependence in the environment. This obviously limits the possibility of achieving full implementation, and requires developing a new design strategy.

Our analysis of full implementation develops such a novel design strategy. Intuitively, the key is to understand how to 'optimally' re-assign the strategic externalities induced by the canonical direct mechanism. For environments with single-crossing preferences and public concavity, we show that this idea is exactly formalized by the problem of minimizing the *spectral radius* of a matrix of externalities, subject to preserving the same row-sums as the matrix of strategic externalities associated with the canonical transfers.[3] As it turns out, the solution to the minimization problem identifies the transfers which achieve full implementation in our sense whenever possible, and generates a rather special hierarchical structure: besides preserving, for any player, the total level of strategic externalities he is subject to from his opponents – which, by the results above, is necessary to preserve incentive compatibility when only common belief in identicality is assumed – these transfers *load* all the strategic externalities on the opponent who displays the lowest amount of preference interdependence. The strategic externalities associated with the *loading transfers* are thus described by a directed graph, which takes the form of a *star network* whose center is the agent with the lowest preference interdependence. In this star network, each peripheral node has one and only one incoming edge, which comes from the center, and the center has only one incoming edge, which comes from the node with the next lowest level of preference interdependence.

The structure of the *loading transfers* enables us to uncover a fairly surprising result: the possibility of full implementation under common knowledge of identicality is characterized by the

[2]For instance, Ollár and Penta (2017) showed 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, in sharp contrast with the impossibility results in belief-free settings.

[3]The spectral radius of a matrix is the largest absolute value of its eigenvalues. 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.

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. Aside from depicting a much more permissive picture for full implementation than Bergemann and Morris (2009a)'s belief-free benchmark (which, in light of the weakness of the identicality assumption and of the results on partial implementation we discussed earlier, may perhaps strike as surprising), this characterization also has powerful 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. At the extreme, whenever an environment includes two agents with private values, common belief in identicality ensures that full implementation is possible via a simple direct mechanism, regardless of the number and interdependence of the other agents.

Besides the *loading transfers*, which as explained have a strongly asymmetric structure, we also study the properties of what we call *'equal-externality' transfers*, which are designed in order to evenly redistribute the strategic externalities across the opponents. The resulting network of externalities is thus a totally connected directed graph, in which each node has one outgoing edge to all other nodes. Such an alternative design strategy is not without loss of generality in our setting (i.e., there are environments in which full implementation is possible, but not with the equal-externality transfers). Nonetheless, we show that these transfers are still widely applicable, and that their symmetric structure grants them an important robustness property. In particular, while the loading transfers have several desirable robustness properties (for instance, they minimize the sensitivity of implementation with respect to lower-orders of rationality – cf. de Clippel et al. (2018)), we show that the equal-externality transfers minimize the impact on the implemented allocation with respect to the possibility of 'slightly faulty' agents or of misspecification of their preferences (cf., Eliaz (2002)).

Finally, we illustrate the applicability of our main results considering a general class of *utilitarian public good problems with network effects*, in which the designer wishes to implement the optimal quantity of public good, for general weights of the utilitarian welfare functional.

We conclude this introduction with a few remarks on 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. Besides this observation, however, there are 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 favors the interpretability of the results and hence pushes a bit further Jackson's concern for economic 'relevance' of full implementation theory. It 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.

The rest of the paper is organized as follows: Section 2 introduces the model and presents some illustrating examples; Sections 3 and 4 provide the characterizations of partial and full implementation, respectively. Section 5 focuses on alternative design strategies for full implementation via transfers, and contains the sensitivity analysis. Section 6 provides the application to utilitarian public good problems with network effects. Section 7 concludes.

2 Framework

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

Agents privately observe their payoff types $\theta_i \in \Theta_i := [\underline{\theta}, \overline{\theta}] \subseteq \mathbb{R}$, 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.[4] Formally, the designer's 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:[5]

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

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

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 $t = (t_i)_{i \in I}$, $t_i : M \to \mathbb{R}$, which specifies the transfer to each agent i, for all profiles of reports $m \in \Theta$. (To distinguish the report from the state, we maintain the notation m_i even though the message spaces are $M_i = \Theta_i$.) Any transfer scheme induces a game with ex-post payoff functions $U^t_i(m;\theta) = v_i(d(m),\theta) + t_i(m)$. When the transfers are clear from the context, we don't emphasize the dependence of the payoff functions on t, 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 $E^{b_{\theta_i}}(U_i(m_i,\theta_{-i};\theta_i,\theta_{-i}))) := \int_{\Theta_{-i}} U_i(m_i,\theta_{-i};\theta_i,\theta_{-i})\, db_{\theta_i}$. 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

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

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

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

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. The social planner wishes to implement the efficient allocation rule. With production cost $c(x) = x^2/2$, the efficient decision rule is $d(\theta) = \sum_{i=1}^{3} \kappa_i \theta_i$, 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 Example 1.3. It is instructive, however, to first go over the examples about the belief-free and i.i.d. common prior benchmarks.

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)).

Example 1.1 (Belief-free Implementation.)**.** In our example, the VCG transfers are the following:

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

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

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

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

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

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

concavity of i's payoff function with respect to his own report. In the case of the canonical mechanism, this amounts to:

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

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).

Example 1.2 (Known i.i.d. Common Prior.)**.** Suppose that types are commonly known to be i.i.d. draws from a uniform distribution over $[0,1]$, and this is known to the designer. Consider the following transfers, which are a special case of Proposition 3 in Ollár and Penta (2017):

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

These transfers induce the following best response function to conjectures $\mu \in \Delta(\theta_{-i} \times M_{-i})$:

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

Under the maintained assumptions, $E(\theta_l|\theta_i) = 0.5$ for all θ_i and $l \neq i$. Hence the term in square brackets cancels out for all types. Truthful revelation therefore is *strictly dominant*, and full implementation is achieved for any γ. Players' best-responses are not affected by other reports, and hence strategic externalities 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 "$E(\theta_l|\theta_i) = 0.5$ for all θ_i and $l \neq i$". If this *moment condition* were not satisfied, these transfers would achieve neither full nor partial implementation. This moment condition was used in (3) to weaken the strategic externalities of the baseline transfers from Example 1.1, but in principle others could be used too.[8] Intuitively, the more information the designer has about agents' beliefs, the more freedom he has to choose a convenient moment condition. As shown by Ollár and Penta (2017), common prior models are maximal in the freedom they allow to the designer and, for a large class of environments, as in the example, strategic externalities can be completely eliminated when types are independent or affiliated.

Identical but Unknown Distribution: Redistribution of Strategic Externalities. Now suppose that agents commonly know that 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

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

moment condition "$E(\theta_l|\theta_i) = 0.5$ 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.

Example 1.3 ($\mathcal{B}^{id}$-Implementation.)**.** Suppose that $\gamma = \hat{\gamma}$ as at the end of Example 1.1, and hence belief-free implementation is not possible. Now consider the following transfers:

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

In this case, the best replies become

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

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

Next consider the following, more complex, transfers:

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

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

To understand the logic behind these transfers, it is useful to look at the induced SE-matrices when $\gamma = gam\hat{m}a$, and compare them to the SE-matrix of the VCG transfers:

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

First notice that both $(t_i^e)_{i\in I}$ and $\left(t_i^l\right)_{i\in I}$ induce SE-matrices such that the sum of the strategic 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

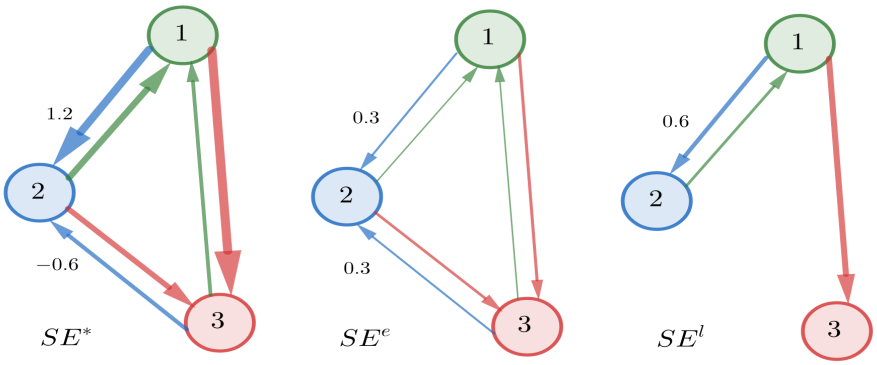

Figure 1: Strategic Externalities and Transfer Schemes in Example 1.3. These network representations illustrate the strategic externalities induced, respectively, by the canonical, equal externality, and loading transfers. For example, the green arrow from agent 2 to 1 illustrates the absolute influence of 2's choice on 1's best reply.

down by the total level of preference interdependence. (So, for instance, transfers such as $\left(t_i^{OP}\right)_{i\in I}$ from Example 1.2, whose SE-matrix consists of all zeros, will not be incentive compatible under the $\mathcal{B}^{id}$-restrictions.) In this sense, strategic externalities can only be *redistributed*, not reduced.

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

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

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

Recursively, this also extends to all powers $k \geq 2$, which implies that, from the second round of elimination on, the set of rationalizable reports shrinks more under $\left(t_i^l\right)_{i\in I}$ than under $(t_i^e)_{i\in I}$. 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* in the example characterizes the mechanisms which achieve full B^{id}-implementation: under these belief restrictions, full implementation is possible if and only if it is achieved by the loading transfers. This in turn enables us to characterize the environments in which full implementation is possible. We also show that the loading transfers induce the fastest contraction among all implementing mechanisms, and that they are the 'most robust' with respect to lower order beliefs in rationality. The *equal-externality* transfers, instead, are 'most robust' if one considers the possibility of mispecifications of agents' preferences.

2.2 Implementation Concepts

We next 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. *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. *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* $b_{\theta_i} \in B^{id}_{\theta_i}$*,* $E^{b_{\theta_i}}\left(U_i\left(\theta_i, \theta_{-i};\theta_i,\theta_{-i}\right)\right) \geq E^{b_{\theta_i}}\left(U_i\left(\theta_i', \theta_{-i};\theta_i,\theta_{-i}\right)\right)$*. If the inequality holds strictly for all* i*,* θ_i*,* $b_{\theta_i} \in B^{id}_{\theta_i}$ *and* $\theta_i' \neq \theta_i$*, then we say that it is* strictly $\mathcal{B}^{id}$*-IC.*

Definition 3. *If* (d,t) *is* $\mathcal{B}^{id}$*-IC, then we say that the transfers* t *partially* $\mathcal{B}^{id}$*-implement the allocation function* d*. Allocation rule* d *is* partially $\mathcal{B}^{id}$*-implementable if there exist some 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 $\Delta(\Theta)$ (and, in particular, it does not contain all degenerate beliefs of each $(\theta_1, ..., \theta_n) \in [\underline{\theta}, \overline{\theta}]^n$), 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 Def. 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.[9] 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 analyze 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 notion introduced above.[10]

[9] $\mathcal{B}^{id}$-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' behavior for their concept of peer-confirming equilibrium, although not in an implementation setting.

[10] 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).)

Formally, $\mathcal{B}^{id}$-rationalizability is defined by an iterated deletion procedure in which, for each type θ_i, a report survives the k-th round of deletion if and only if it can be justified by conjectures (joint distributions over opponents' types and strategies) which are consistent with the belief restrictions $\mathcal{B}^{id}$, and with the previous rounds of the deletion procedure. For every i and θ_i, the set of conjectures that are consistent with common belief in identicality is defined as $C_{\theta_i}^{id} := \left\{\mu_i \in \Delta\left(M_{-i} \times \Theta_{-i}\right) : marg_{\Theta_{-i}}\mu_i \in B_{\theta_i}^{id}\right\}$.

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

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

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

Definition 5 (Full Implementation)**.** *The transfer scheme $t = (t_i)_{i \in I}$ fully implements d under common knowledge of identicality if $R_i^{id}\left(\theta_i\right) = \{\theta_i\}$ 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.*[11]

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 $b_{\theta_i} \in \Delta\left(\Theta_{-i}\right)$. Strict $\mathcal{B}^{id}$-IC therefore is a 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.

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*, $t^* = \left(t_i^*\left(\cdot\right)\right)_{i \in I}$, which are defined as follows: for each $i \in I$ and $m \in \Theta$,

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

In the following, we will refer to the pair (d, t^*) as the *canonical direct mechanism.*[12]

[11] 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.

[12] The term 'canonical mechanism' is traditionally used to refer to Maskin's mechanism for full implementation. That mechanism is not 'direct' and it induces an integer game to eliminate undesirable equilibria. We call (d, t^*)

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 (*ibid.*, 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: Characterization)**.** *Under the maintained assumptions: d is partially $\mathcal{B}^{id}$-implementable if and only if it is partially $\mathcal{B}^{id}$-implemented by t^*.*

Theorem 1 implies that, under the $\mathcal{B}^{id}$-restrictions, there is no reason to consider transfers other than the canonical ones. As we will see, this will not be the case for full implementation: full implementation may fail under the canonical transfers, but be achieved by other transfers. Besides its intrinsic interest, this result also simplifies the task of identifying which conditions on the environment are necessary or sufficient for partial implementation: it suffices to study properties of the payoff functions induced by the canonical mechanism, $U_i^*(m;\theta)$, which only depend on the allocation function and on the agents' preferences. First note that, under the maintained assumptions, the canonical direct mechanism induces payoff functions which are twice continuously differentiable. 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. That is:

Corollary 1. *Under the maintained assumptions:*

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

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 (as in part 2 of Theorem 1), 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:*

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

the canonical *direct* mechanism, since special cases of this mechanism are pervasive in the partial implementation literature. For example, 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), etc. Lemma 1 in Ollár and Penta (2017) generalized the earlier results in the papers above. The term *canonical direct mechanism* was first used with this acceptation in Ollár and Penta (2017).

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,

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

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.[13]

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., $B^{id}_{\theta_i} = B^{id}_{\theta'_i}$ 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 $B^{id}_{\theta_i} = B^{id}_{\theta'_i}$ 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:

Proposition 1 ('Payoff Equivalence' for $\mathcal{B}^{id}$-restrictions)**.** *For any $\mathcal{B}^{id}$-incentive compatible direct mechanism, for any type $\theta_i \in \Theta_i$ and belief $b_{\theta_i} \in B^{id}_{\theta_i}$, the expected payments are the same as in the canonical mechanism. Formally: if (d, t) is $\mathcal{B}^{id}$-IC, then for each $i \in I$, $\theta_i \in \Theta_i$ and $b_{\theta_i} \in B^{id}_{\theta_i}$, $E^{b_{\theta_i}}(t_i(\theta_i, \theta_{-i})) = E^{b_{\theta_i}}(t_i^*(\theta_i, \theta_{-i}))$.*

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

[13] 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 $b_i \in B^{id}_{\theta_i}$, but not in the ex-post sense.

1. [Single-Crossing] For all i and (x,θ), $\frac{\partial^2 v_i}{\partial x \partial \theta_i}(x,\theta) > 0$ and $\partial d/\partial \theta_i > 0$

2. [Public Concavity] For all i and j, $\partial^2 v_i/\partial^2 x$ and $\partial^2 v_i/\partial x \partial \theta_j$ are constant in θ, and $\partial d/\partial \theta_i$ is constant in θ_j for all i and j .

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.[14] Compared to these applications, Definition 6 also accommodates more general dependence on x, as long as the concavity and the cross derivatives are public information. The application to utilitarian public good problems with network externalities in Section 6 provides an example of a general class of economically important environments that satisfy the SC-PC assumptions. (Nonetheless, we stress that most of the results below do not rely on these assumptions.)

The important consequences of these assumptions are two: Part 1 ensures that partial $\mathcal{B}^{id}$-implementation is possible;[15] Part 2 also ensures that, in the canonical direct mechanisms, all second order derivatives $\frac{\partial^2 U_i^*}{\partial m_i \partial m_j} = -\frac{\partial^2 v_i}{\partial x \partial \theta_j} \cdot \frac{\partial d}{\partial \theta_i}$ are constant in (θ, m) and s.t. $\partial^2 U_i^*/\partial^2 m_i \neq 0$. We can thus define the (normalized) *canonical externalities* as real numbers $\xi_{ij} := \frac{\partial^2 U_i^*/\partial m_i \partial m_j}{\partial^2 U_i^*/\partial^2 m_i}$. For each i, let $\xi_i := \sum_{j \neq i} \xi_{ij}$, and relabel agents if necessary so that $|\xi_1| \leq |\xi_2| \leq \ldots \leq |\xi_n|$. In SC-PC environment, the property that the second-order derivatives of the payoff function are constant in (θ, m) actually holds for all mechanisms based on *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.

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^*

[14] 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, Golub and Goyal (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)).

[15] This can be checked by noting that the environment satisfies the conditions of Corollary 1. In fact, these conditions are also sufficient for ex-post incentive compatibility.

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, respecitvely 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 $|SE^t_{max}|_{ij} := \frac{\max_{(m,\theta)\in\Theta\times\Theta} |\partial^2 U^t_i(m,\theta)/\partial m_i \partial m_j|}{\min_{(m,\theta)\in\Theta\times\Theta} |\partial^2 U^t_i(m,\theta)/\partial^2 m_i|}$ and $|SE^t_{min}|_{ij} := \frac{\min_{(m,\theta)\in\Theta\times\Theta} |\partial^2 U^t_i(m,\theta)/\partial m_i \partial m_j|}{\max_{(m,\theta)\in\Theta\times\Theta} |\partial^2 U^t_i(m,\theta)/\partial^2 m_i|}$. Given allocation rule d, and preferences $v = (v_i)_{i\in I}$, let $IC^{id}(v,d)$ denote the set of twice continuously differentiable transfers $t = (t_i)_{i\in I}$ which are strictly $\mathcal{B}^{id}$-incentive compatible. For any square matrix $X \in \mathbb{R}^{n\times n}$, we let $\rho(X)$ denote the *spectral radius* of X, i.e. the largest absolute value of its eigenvalues.[16]

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)**.** *Let $t \in IC^{id}(v,d)$. Then:*

(i) If $\rho(|SE^t_{max}|) < 1$, then t fully $\mathcal{B}^{id}$-implements d.

(ii) If $\rho(|SE^t_{min}|) \geq 1$, then t does not fully $\mathcal{B}^{id}$-implement d

First note that, if $t \in IC^{id}(v,d)$ 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.[17] 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:[18]

[16] 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.

[17] 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).

[18] We note that, for the analysis of partial implementation (cf. Appendix A), it suffices to check conditions on the agents' optimization problem at the truthful profile (which is all that matters for partial implementation). In

Lemma 2. *Let $t \in IC^{id}(v,d)$. Then, for all θ and $(m_i, \bar{m}_{-i})$ s.t. $\bar{m}_j = \bar{m}_k$ for all $j,k \neq i$:*

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

These conditions are also sufficient in SC-PC, when t 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. App. A).

The overall design strategy that emerges from combining Lemma 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, in which the SE^*-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 as in the SE^*-matrix. Hence, in SC-PC settings, the design strategy boils down to a problem of minimizing the spectral radius of a matrix, subject to preserving the row-sums of the SE^*-matrix. With this in mind, we let $\underline{\rho}^{max}(v,d)$ and $\underline{\rho}^{min}(v,d)$ denote the lowest spectral radii induced by transfers in $IC^{id}(v,d)$, for the $|SE^t_{max}|$ and $|SE^t_{min}|$-matrices, respectively:

$$\underline{\rho}^{\max}(v,d) := \min_t \rho\left(|SE^t_{\max}|\right) \quad \text{s.t.: } t \in IC^{id}(v,d) \qquad \text{and} \qquad \underline{\rho}^{\min}(v,d) := \min_t \rho\left(|SE^t_{\min}|\right). \quad \text{s.t.: } t \in IC^{id}(v,d)$$

contrast, the next result refers to restrictions that $\mathcal{B}^{id}$-IC imposes on the strategic externalities at the non-truthful profiles. This is needed because the result in Lemma 1 refers to properties of the $|SE^t|$-matrix, which in turn depend on the properties of the strategic externalities at all profiles (m,θ), not only at the truthful ones.

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 Example 1.3, the logic of the construction is to redistribute the strategic externalities so that, in the resulting mechanism, they are all concentrated on the two agents with the smallest canonical externalities (given the relabeling above, these are agents 1 and 2). Formally, the *loading transfers* $\left(t_i^l\right)_{i\in I}$ are defined as follows: for each $i \in I$ and $m \in M_i \times M_{-i}$,

$$t_i^l(m) = \underbrace{t_i^*(m)}_{\text{canonical transfers}} + \underbrace{L_i^l(m_{-i})\, m_i}_{\substack{\text{redistribution of} \\ \text{canonical externalities}}}, \tag{6}$$

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

$$L_i^l(m_{-i}) = \begin{cases} \left[-\sum_{\substack{k\neq 1 \\ k\neq 2}} \frac{\partial^2 v_1}{\partial x \partial \theta_k} m_2 + \sum_{\substack{k\neq 1 \\ k\neq 2}} \frac{\partial^2 v_1}{\partial x \partial \theta_k} m_k\right] \frac{\partial d}{\partial \theta_1} \text{ if } i = 1 \\ \left[-\sum_{\substack{k\neq 1 \\ k\neq j}} \frac{\partial^2 v_j}{\partial x \partial \theta_k} m_1 + \sum_{\substack{k\neq 1 \\ k\neq j}} \frac{\partial^2 v_j}{\partial x \partial \theta_k} m_k\right] \frac{\partial d}{\partial \theta_j} \text{ if } i \neq 1 \end{cases} \tag{7}$$

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 $\partial^2_{i1} U_i^l = \sum_{j\neq i} \partial^2_{ij} U_i^*$ for all $i \neq 1$; $\partial^2_{12} U_1^l = \sum_{j\neq 1} \partial^2_{1j} U_1^*$ and otherwise $\partial^2_{ij} U_i^l = 0$. 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)**.** *For any environment (v,d) that satisfies the SC-PC conditions, the following holds:*

1. *d is fully $\mathcal{B}^{id}$-implementable if and only if it is fully $\mathcal{B}^{id}$-implemented by t^l.*
2. *d is fully $\mathcal{B}^{id}$-implementable if and only if the canonical externalities are such that $|\xi_1\xi_2| < 1$ (or, equivalently, if and only if the preferences interdependence of agents 1 and 2 are sufficiently small: that is, 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 1 of the theorem follows from the following facts, which are shown in the proof: first, the loading transfers in SC-PC environments are strictly $\mathcal{B}^{id}$-IC and induce constant strategic externalities, and hence (by Lemma 1) they achieve full implementation if and only if $\rho(|SE^{t^l}|) < 1$; second, among the class of $\mathcal{B}^{id}$-IC transfers, the loading transfers are those that minimize the spectral radius of both the $|SE^{t^l}_{min}|$ and of the $|SE^{t^l}_{max}|$ matrices. Hence, if $\rho(|SE^{t^l}|) \geq 1$, then

full $\mathcal{B}^{id}$-implementation is impossible, because any $\mathcal{B}^{id}$-IC transfer $t \in IC^{id}(v,d)$ would be such that $\rho(|SE^t_{min}|) \geq 1$ (cf. Lemma 1, part (ii)); on the other hand, if $\rho(|SE^{t^l}|) < 1$, then full implementation is possible, and it is achieved by t^l. The reason why t^l achieves both $\underline{\rho}^{\max}(v,d)$ and $\underline{\rho}^{\min}(v,d)$ is that, as it turns out, the best way of minimizing the spectral radius of the strategic externality matrix, subject to the constraint (imposed by $\mathcal{B}^{id}$-IC) of preserving the same row-sums as in the SE^*-matrix (cf. Lemma 2), is to concentrate all the strategic externalities of any agent i on the opponent with the smallest $|\xi_j|$: that is on agent 2 for $i = 1$, and on agent 1 for all $i \neq 1$. This is precisely what is achieved by the SE^{t^l}-matrix, and hence by the t^l transfers.

Part 2 follows from the fact that $\rho(|SE^{t^l}|) < 1$ if and only if $|\xi_1 \xi_2| < 1$. This in turn implies that the possibility of achieving full $\mathcal{B}^{id}$-implementation is pinned down by the canonical externalities of the two agents with the smallest canonical externalities (or, equivalently, those with the smallest level 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, and that is regardless of the strength of the preference interdependence of other agents, or of their number.

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

Before moving on to non-SCPC environments, it may be useful to discuss how the results above compare with the typical characterizations in the full implementation literature. First, that literature typically considers preferences which are not necessarily quasi-linear, and focuses on social choice function (SCFs) $f : \Theta \to Y$, where Y denotes the space of outcomes (see, e.g., Bergemann and Morris (2009a)). With quasilinear 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 allows 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).[19] By referring to the eigenvalues of the strategic externality matrices,

[19] 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.

these conditions also enabled us to draw a bridge between full implementation and networks (e.g., Elliott and Golub (2019), Galeotti, Golub and Goyal (2020)), which may prove fertile for both strands of the literatures (these points are further discussed in the Conclusions).

4.3 Full $\mathcal{B}^{id}$-Implementation in Non SC-PC Environments

We turn next to environments which do not satisfy the SC-PC restriction. First of all, we recall that the design principle we developed in Section 4.1 – namely, the idea of *redistributing* the strategic externalities in order to minimize the spectral radius of the strategic externality matrix – also holds outside of SC-PC environments. More specifically, the necessity part of Lemma 2 (which shows that $\mathcal{B}^{id}$-IC requires maintaining the same total strategic externalities as the canonical direct mechanism at profiles of identical reports) and Lemma 1 (which shows, in particular, that full $\mathcal{B}^{id}$-implementation is achieved by any transfers which induces a strategic externality matrix with spectral radius smaller than one) hold for general environments.

In this Section we exploit those general insights to study full implementation in non SC-PC environments. The key difficulty in these settings is that the strategic externalities of the canonical direct mechanism 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 t^* based on a midpoint between the lowest and highest stategic externalities generated by the environment. The next result 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 if they are not too far from the strategic externalities attained over the full domain.

Formally, given the entries of the strategic externalities matrix $\left(SE^{*}\left(m;\theta\right)_{ij}\right)_{(i,j)\in I\times I}$, we define the 'midpoint' strategic externalities $\bar{SE}^{*}_{ij}$ as follows:[20]

$$\bar{SE}^{*}_{ij} := \frac{\max_{(m,\theta)} SE^{*}_{ij}\left(m;\theta\right) + \min_{(m,\theta)} SE^{*}_{ij}\left(m;\theta\right)}{2}.$$

Similar to Section 4.1, for each $i \in I$ we let $\bar{\xi}^{l}_{i} := \sum_{j\neq i} \bar{SE}^{*}_{ij}$ denote i's total strategic externalities in the $\bar{SE}^{*}_{ij}$ matrix, and relabel agents if needed so that $|\bar{\xi}_1| \leq |\bar{\xi}_2| \leq \ldots \leq |\bar{\xi}_n|$. Assume that $|\bar{\xi}_n| < \infty$. The *Generalized Loading Tranfers*, $(\bar{t}^{l}_{i})_{i\in I}$, are defined as follows:

$$\bar{t}^{l}_{i}\left(m\right) = \underbrace{t^{*}_{i}\left(m\right)}_{\text{canonical transfers}} + \underbrace{\bar{L}^{l}_{i}\left(m_{-i}\right) m_{i}}_{\substack{\text{redistribution of} \\ \text{externalities in } \bar{SE}^{*}}}, \tag{8}$$

[20] Any convex combination of $SE^{*}_{ij}\left(m;\theta\right)$ points, i.e. $\int SE^{*}_{ij}\left(m;\theta\right) d\mu$ where μ is a probability measure over $\Theta\times\Theta$ is a viable definition for $\bar{SE}^{*}_{ij}$. Note that the relevant values of distances in A are impacted by the choice of μ. Here we do not expand on the implications of this, but one may further reduce the resulting spectral radius ρ via the optimal choice of μ.

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

$$\bar{L}_i^l(m_{-i}) = \begin{cases} \left[-\sum_{\substack{k\neq 1\\ k\neq 2}} \bar{SE}_{1k}^* m_2 + \sum_{\substack{k\neq 1\\ k\neq 2}} \bar{SE}_{1k}^* m_k \right] \frac{\partial d}{\partial \theta_1} \text{ if } i = 1 \\ \left[-\sum_{\substack{k\neq 1\\ k\neq j}} \bar{SE}_{jk}^* m_1 + \sum_{\substack{k\neq 1\\ k\neq j}} \bar{SE}_{jk}^* m_k \right] \frac{\partial d}{\partial \theta_j} \text{ if } i \neq 1 \end{cases}$$

To measure the distances between the actual strategic externalities $SE^*(m,\theta)$ and the midpoints above, we let

$$\alpha_{ij} := \max_{(m,\theta)} SE_{ij}^*(m;\theta) - \bar{SE}_{ij}^*.$$

and define the *matrix of approximation errors*, A, s.t. $[A]_{ij} := \alpha_{ij}$ for all i,j. Finally, we also define $\alpha_1 := \max_j \alpha_{1j}$, $\alpha_2 := \max_j \alpha_{2j}$, and $\alpha := \max_{ij:i\neq 1,2} \alpha_{ij}$.

The next result provides necessary and sufficient conditions for the *generalized loading transfers* to achieve full $\mathcal{B}^{id}$-Implementation in general environments.[21] The key idea is to study conditions that ensure $\rho(|\bar{SE}^{\vec{t}^l}| - A) < 1$ for necessity, and $\rho(|\bar{SE}^{\vec{t}^l}| + \mathcal{A}) < 1$ for sufficiency, where

$$|\bar{SE}^{\vec{t}^l}| + \mathcal{A} = \begin{bmatrix} 0 & |\bar{\xi}_1| + \alpha_1 & \alpha_1 & \dots & \alpha_1 \\ |\bar{\xi}_2| + \alpha_2 & 0 & \alpha_2 & \dots & \alpha_2 \\ |\bar{\xi}_3| + \alpha & \alpha & 0 & \dots & \alpha \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ |\bar{\xi}_n| + \alpha & \alpha & \alpha & \dots & 0 \end{bmatrix}. \tag{9}$$

Theorem 3. *Let (v,d) be such that d is partially $\mathcal{B}^{id}$-implementable. Then:*

1. *d is fully $\mathcal{B}^{id}$-implemented by $\vec{t}^l$ if: (i) $|\bar{\xi}_1\bar{\xi}_2| + \alpha_1 G_1 + \alpha_2 G_2 + (n-3)\,\alpha(1 - |\bar{\xi}_1\bar{\xi}_2|) < 1$, and (ii) $\frac{1}{3}\left[|\bar{\xi}_1\bar{\xi}_2| + \alpha_1 F_1 + \alpha_2 F_2\right] + \frac{2}{3}(n-3)\,\alpha < 1$.*[22]
2. *d is fully $\mathcal{B}^{id}$-implemented by $\vec{t}^l$ only if: $|\left(|\bar{\xi}_1| - \alpha_{12}\right)\left(|\bar{\xi}_2| - \alpha_{21}\right)| < 1$.*

First of all, notice that if there are no 'approximation errors' (i.e., $\alpha = \alpha_1 = \alpha_2 = 0$), then (i) implies (ii), and the conditions in points 1 and 2 both boil down to $|\xi_1\xi_2| < 1$. Hence, the characterization in Theorem 2 obtains from this theorem for the special case in which $\alpha = \alpha_1 = \alpha_2 = 0$. With positive approximation terms, there are two main changes in the sufficient conditions (part 1): first, the two lowest canonical strategic externalities (ξ_1, ξ_2) (which are constant over the entire domain in the SC-PC setting of Theorem 2), are replaced by the two lowest strategic externalities at the 'midpoint', $(\bar{\xi}_1, \bar{\xi}_2)$; second, the condition that the absolute value of their product is less than one is strengthened in that it is required to hold with sufficient slack so as to accommodate the extra terms that depend on the approximation errors (cf. (i) and (ii)).

Intuitively, the general message is that, as long as the strategic externalities do not vary too much across the entire domain, so that the 'approximation errors' are small, then the basic insights from Theorem 2 extend to non SC-PC, via the adoption of the *generalized loading transfers* defined

[21] We note that this result is significantly stronger than other results that appeared in earlier drafts of this paper. In particular, earlier results (which are available upon request) provided sufficient conditions for the transfers in (6) to achieve full $\mathcal{B}^{id}$-implementation in non-SCPC settings. The next result, in contrast, is based on the *generalized* loading transfers, which tailor the design of the mechanism to the non-SCPC environment.

[22] Where $F_1 = \sum_{i\neq 1} |\bar{\xi}_i| + \alpha_2 + (n-2)\,\alpha$, $F_2 = (n-2)\,\alpha$, $G_1 = (\alpha_2 + 1)\sum_{i\neq 1,2}\left(|\bar{\xi}_i| + \alpha\right) + (\alpha+1)\left(|\bar{\xi}_2| + \alpha_2\right)$, and $G_2 = |\bar{\xi}_1| \sum_{i\neq 1,2} |\bar{\xi}_i| + (\alpha+1)\,|\bar{\xi}_1| + (n-2)\,\alpha$.

in (8). Note that by the continuity of ρ, for any $|\bar{\xi}_1\bar{\xi}_2| < 1$, there exists α sufficiently close to 0 such that full $\mathcal{B}^{id}$-implementation follows. However if externalities vary and α is not very small, then joint sufficient conditions on $\bar{\boldsymbol{\xi}}$ and $\boldsymbol{\alpha}$ can still imply full implementation. To understand the logic of the sufficiency result in Theorem 3, it is instructive to exploring the role that the different components of the matrix in (9) play in the conditions (i) and (ii) in part 1. The next remark gathers the main interesting special cases:

Remark 1. *Consider the sufficient conditions in part 1 of Theorem 3:*

1. *If $\alpha_1 = \alpha_2 = 0$, then (i) and (ii) hold if and only if $(n-3)\,\alpha < 1$, and $|\bar{\xi}_1\bar{\xi}_2| < 1$.*
2. *If $\alpha = 0$, then (i) implies (ii) and it is equivalent to:*
   $|\bar{\xi}_1\bar{\xi}_2| + \alpha_1 \sum_{i\neq 1} |\bar{\xi}_i| + \alpha_1\alpha_2 \left(1 + \sum_{i\neq 1,2} |\bar{\xi}_i|\right) + \alpha_2|\bar{\xi}_1| \left(1 + \sum_{i\neq 1,2} |\bar{\xi}_i|\right) < 1.$
   (a) *If $\alpha = \alpha_2 = 0$, then this reduces to $|\bar{\xi}_1\bar{\xi}_2| + \alpha_1 \sum_{i\neq 1} |\bar{\xi}_i| < 1$.*
   (b) *If $\alpha = \alpha_1 = 0$, then this reduces to $|\bar{\xi}_1\bar{\xi}_2| + \alpha_2|\bar{\xi}_1| \left(1 + \sum_{i\neq 1,2} |\bar{\xi}_i|\right) < 1$.*

5 Other Designs: The *Equal-Externality* Transfers

In this Section we consider alternative transfer schemes to the loading transfers, which have especially relevant structure and properties. As illustrated by the t^e transfers in Example 1.3, these transfers pursue a uniform redistribution of the strategic externalities. As it will be shown, such alternative design principle is still widely applicable and has desirable robustness properties.

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

$$t_i^e(m) := \underbrace{t_i^*(m)}_{\text{canonical transfers}} + \underbrace{L_i^e(m_{-i})\, m_i}_{\substack{\text{redistribution of} \\ \text{canonical externalities}}}, \tag{10}$$

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

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

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

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

5.1 Full Implementation via *Equal-Externality* Transfers

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

Theorem 4. *Under SC-PC, the transfers in (10) achieve full $\mathcal{B}^{id}$-implementation if*

$$\textit{either (i)} \left|\sum_{j\neq i}\frac{\partial^2 v_i}{\partial x\partial\theta_j}\right| < \frac{\partial^2 v_i}{\partial x\partial\theta_i} \textit{ for all } i\textit{; or (ii)} \sum_{j\neq i}\left|\frac{\partial^2 v_j}{\partial x\partial\theta_i}/\frac{\partial^2 v_j}{\partial x\partial\theta_j}\right| < 1 \textit{ for all } i. \tag{11}$$

The proof of this result follows from more general results which we discuss in Appendix C. To appreciate the conditions in (11), it is useful to compare them to the following known sufficient conditions for full implementation via the *canonical* transfers: Under SC-PC, the canonical transfers achieve (belief-free) full implementation if

$$\sum_{j\neq i}\left|\frac{\partial^2 v_i}{\partial x\partial\theta_j}\right| < \frac{\partial^2 v_i}{\partial x\partial\theta_i} \text{ for all } i. \tag{12}$$

Condition (12) requires that the sum of preference interdependencies, across all of opponents' of agent i, to be small relative to the dependence of i's marginal utility on his own type. When this condition is satisfied, the resulting strategic externalities in the canonical direct mechanism are small, and belief-free full implementation follows from the results in Ollár and Penta (2017) and Bergemann and Morris (2009a). Relative to this belief-free benchmark, the $\mathcal{B}^{id}$-restrictions enable the designer to redistribute the strategic externalities, and hence to weaken Condition (12) to part (i) of Theorem 4, in which the absolute value is moved outside of the summation. This means that, by relying on the $\mathcal{B}^{id}$-restrictions, preference interdependencies with opposite signs can be leveraged, to obtain full implementation: Under $\mathcal{B}^{id}$, it is the total amount of *net* preference interdependence that matters, not the total amount of *absolute* interdependence.

The second condition in (11) instead focuses on the total *impact* that agent i's type has on other agents' preferences. Rather than pointing at the way player i's preferences depend on others' information, it measures the total impact of i's information on others' preferences. The reason why this alternative condition is also sufficient is related to the idea of *Limited Strategic Impact*, formalized by part (ii) of Lemma 5 in Appendix C.

Example 3. *In the setting of the leading examples, consider preferences $v_i : X \times \Theta \to \mathbb{R}$ which satisfy the following condition:*

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

It is immediate to check that condition (12) does not hold, and one can also show that belief-free full implementation is not possible in this setting. In fact, for agent 3 condition (i) is also violated. Condition (ii), however, holds and implementation via the equal-externality transfers is possible. (Of course, implementation via the loading transfers is possible too.)

Proposition 4 in Appendix C further formalizes the sense in which – while still not as applicable as the loading transfers (which, by Theorem 2, achieve full implementation whenever possible) – the logic of the equal-externality transfers is still widely applicable.

5.2 Environments with Symmetric Aggregators

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

Definition 7 (Symmetric Aggregators in Valuations)**.** *An environment has symmetric aggregators in valuations if for all i, there exist $w : X \times \mathbb{R} \to \mathbb{R}$ and $h_i : \Theta \to \mathbb{R}$ strictly increasing in θ_i such that $v_i(x, \theta) = w(x, h_i(\theta))$, $\partial h_i(\theta)/\partial\theta_i = \partial h_j(\theta)/\partial\theta_j$ and $\sum_{k\neq i}(\partial h_i(\theta)/\partial\theta_k) = \sum_{k\neq j}(\partial h_j(\theta)/\partial\theta_k)$ for all i, j and θ.*

Proposition 2 (Full $\mathcal{B}^{id}$-Implementation with Symmetric Aggregators)**.** *Consider an SC-PC environment with symmetric aggregators in valuations.*

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

Figure 5.2 summarizes the relations between different transfer design strategies. This figure summarizes the implications of the counterexamples above; and the results on full implementability under identical distributions (via the loading transfers in Theorem 2, via designing diagonal dominance in Lemma 5 (in the Appendix), via the equal-externality transfers in Theorem 4, and in environments with symmetric aggregators in Proposition 2).

5.3 Sensitivity Results

In this section we explore the sensitivity of the loading and equal-externality mechanisms to various forms of model deviations from the baseline model of rationality, first with respect to the possibility of 'slightly faulty' agents, and then with respect to lower-orders of rationality.

5.3.1 Lower Orders of Rationality and Robust Level-k Implementation

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

Theorem 5. *Let t be any $\mathcal{B}^{id}$-IC transfer scheme. Then: $R_i^{id,k}(\theta_i|t^l) \subseteq R_i^{id,k}(\theta_i|t)$ for all k.*

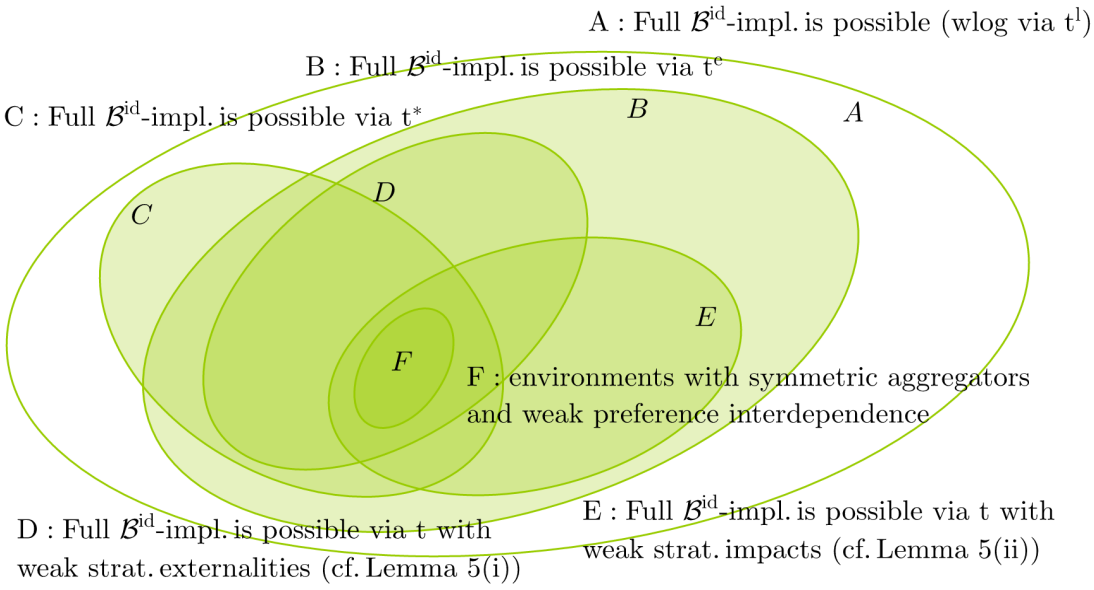

Figure 2: Viability of alternative design strategies for full implementation. This diagram illustrates the relationship between the possibility of achieving full $\mathcal{B}^{id}$-implementability via different transfer schemes, for sets of environments (v, d) which satisfy the SC-PC restriction in Def. 6.

This result is interesting because it points at a different notion of robustness, with respect to lower order beliefs in rationality: the loading transfers are the most efficient at minimizing the possible misreports which could arise due to failures of common belief in rationality. This is an important property, because common belief in rationality (which is implicit in the notion of rationalizability) is often very demanding, and need not be satisfied in a given environment. If the designer is concerned with agents' sharing lower orders of mutual belief in rationality, then he would not only consider the sets $R_i^{id}(\theta_i)$, but also $\left(R_i^{id,k}(\theta_i)\right)_{k\in\mathbb{N}}$.[23] Hence, among two fully implementing transfers (i.e., both such that $R^{id}(\theta_i) = \{\theta_i\}$), he should prefer the one which also induces the smaller $R_i^{id,k}(\theta_i)$ at every k. The loading tranfers are optimal in this respect.

As further explained in Appendix D.2, this notion of robustness is connected to the literature on level-k implementation (and, particularly, to de Clippel et al. (2018), which has explicitly considered designing mechanisms for players who don't share common knowledge of rationality.

5.3.2 *Slightly Faulty* Players and Preference Misspecification

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

[23] Saran (2016) makes a similar point in the context of complete information environments, and studies full implementation when common belief in rationality is relaxed and arbitrary level-0 anchors are allowed.

due to a misspecification of agent's preferences in the model.[24]

Formally, let $F \subseteq I$ be an arbitrary set of possibly *slightly faulty* agents, in the sense that they may report messages up to $\varepsilon > 0$ away from their optimal response. For any $\varepsilon > 0$, $\theta_i \in \Theta_i$ and $\mu_i \in \Delta(\Theta_{-i} \times M_{-i})$, we let $BR^{\varepsilon}_{\theta_i}(\mu_i) = \{m_i \in \Theta_i : |m_i - m'_i| \leq \varepsilon \text{ for some } m'_i \in BR_{\theta_i}(\mu_i)\}$ denote the set of possible responses of a possibly ε-faulty agent with type θ_i and conjectures μ_i. The next solution concept characterizes the behavioral implication of assuming common knowledge that a subset F of players may be ε-faulty in the sense above:

Definition 8 (F_ε-rationalizability)**.** *Fix a direct mechanism (d,t), $\varepsilon \geq 0$ and $F \subseteq I$. For any $\theta_i \in \Theta_i$ and $\mu_i \in \Delta(\Theta_{-i} \times M_{-i})$, we let $BR^{F_\varepsilon}_{\theta_i} = BR^{\varepsilon}_{\theta_i}(\mu_i)$ if $i \in F$, and $BR^{F_\varepsilon}_{\theta_i} = BR_{\theta_i}(\mu_i)$ otherwise. For every $i \in I$, and $k = 1, 2, ...,$ let $R_i^{F_\varepsilon,0} = \Theta_i \times M_i$, $R_{-i}^{F_\varepsilon,k-1} = \times_{j \neq i} R_i^{F_\varepsilon,k-1}$,*

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

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

$R_i^{F_\varepsilon}$ represents our model of strategic interaction when players consider the possibility that agents in F may be ε-faulty. The next definition formalizes our notion of robustness to 'slightly faulty' agents. In words, the idea is that the designer does not know how many or which agents might be potentially faulty, and the criterion with which he/she assesses the robustness of the mechanism is the worst-case scenario across all possible configurations of sets of faulty agents. The measure of the fragility of the mechanism is therefore provided by the largest misreport consistent with $R_i^{F_\varepsilon}$, across all agents and all configurations of the set of faulty agents:

Definition 9 (Sensitivity to ε-Faulty Agents)**.** *Fix a direct mechanism (d,t). For any $n_f = 1, ..., n$, let $N(k) := \{F \subseteq I : |F| = n_f\}$, and $\eta^t(\varepsilon, n_f) := \sup_{F \in N(n_f)} \sup_{i \in I} \sup_{\theta_i \in \Theta_i} \sup_{m_i \in R_i^{F_\varepsilon}(\theta_i)} |m_i - \theta_i|$ be t's sensitivity to n_f agents who are ε-faulty, and let $\eta^t(\varepsilon) = (\eta^t(\varepsilon, 1), ..., \eta^t(\varepsilon, n))$*

The next result shows that, in SC-PC environments with symmetric aggregators, the equal-externality transfers are more robust than the loading transfers in this sense:

Theorem 6. *[Sensitivity to ε-Faulty Players]Under SC-PC and Symmetric Aggregators, for all $\varepsilon > 0$, $\eta^{t^e}(\varepsilon) \leq \eta^{t^l}(\varepsilon)$, moreover for all $n_f < n$, $\eta^{t^e}(\varepsilon, n_f) < \eta^{t^l}(\varepsilon, n_f)$.*

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.

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

6 Application: Utilitarian Public Good Problems with Network Effects and Private Information.

In this section we illustrate how the general results above can be easily applied to a class of environments which is especially relevant for the networks literature.[25] Specifically, we consider an economy with n angents, $I = \{1, ..., n\}$, with private information $\theta_i \in [\underline{\theta}, \overline{\theta}] \in \mathbb{R}_+$. We let $x \in \mathbb{R}_+$ denote the quantity of a public good, and agents' preferences are such that, for each $i \in I$,

$$v_i(x; \theta_1, ..., \theta_n) = \left(K_i + \sum_{j \in I} \gamma_{ij} \theta_j \right) \cdot x \tag{13}$$

The term $K_i \in \mathbb{R}$ is a type-independent component of i's marginal utility for the public good; the rest is determined by a network of preference interdependence. For convenience, we describe the latter component by means of a *matrix of preference interdependence*:

$$\Gamma = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1n} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{n1} & \gamma_{32} & \cdots & \gamma_{nn} \end{bmatrix}.$$

This is an environment with network effects and private information. The *private values* case holds if and only if $\gamma_{ij} = 0$ for all $i \in I$ and $j \neq i$, otherwise we have *interdependent values*. We assume that the social planner aims to maximize a utilitarian social welfare functional, with generalized weights $(\lambda_i)_{i \in I}$ s.t. $\lambda_i \geq 0$ and $\sum_{i \in I} \lambda_i = 1$, net of the production costs. Formally, for each $\theta \in \Theta$, he wishes to implement the following quantity of public good:

$$d(\theta) \in argmax_{x \in \mathbb{R}_+} = \sum_{i \in I} \lambda_i \cdot v_i(x; \theta) - c(x), \tag{14}$$

where $c : \mathbb{R}_+ \to \mathbb{R}_+$ is twice continuously differentiable, strictly increasing and strictly convex function. Letting $h_i(\theta) = \left(K_i + \sum_{j \in I} \gamma_{ij} \theta_j \right)$ denote i's marginal utility for the public good, we also assume that $min_{\theta \in \Theta} \sum_{i \in I} h_i(\theta) \geq lim_{x \to 0^+} c'(x)$ and $lim_{x \to \infty} c'(x) \geq max_{\theta \in \Theta} \sum_{i \in I} h_i(\theta)$.[26]

We will refer to environments (v, d) which satisfy (13) and (14) and the assumptions above as *utilitarian public good problems with network effects*. First we note that any such environment satisfies the public-concavity condition in part 2 of Def. 6. Hence, this class of public good problems in networked economies are an SC-PC environment if and only if they satisfy the standard single-crossing condition in part 1 of Def. 6: the restriction on v_i in this case holds if and only if (i) $\gamma_{ii} > 0$, and strict monotonicity of the allocation rule holds if and only if (ii) $\overline{\gamma}_i := \sum_{l \in I} \gamma_{il} \cdot \lambda_l > 0$

[25] E.g., Leister, Zenou and Zhou (2020), Calvó-Armengol and De Martí (2007), Galeotti et al. (2010), Calvó-Armengol et al. (2015), Blume et al. (2015), De Martí and Zenou (2015), Golub and Morris (2017), Myatt and Wallace (2019), Leister (2020).

[26] The latter assumptions merely ensure existence of an interior optimum in the planner's problem in (14) for all θ. In conjunction with the twice continuous differentiability of c, this also ensures that d is differentiable and responsive. These conditions follow, for instance, from standard Inada conditions on the cost function (i.e., $lim_{x \to \infty} c'(x) = \infty$ and $lim_{x \to 0^+} c'(x) = 0$), as soon as the sum of the marginal utilities for the public good is non-negative ($min_{\theta \in \Theta} \sum_{i \in I} h_i(\theta) \geq 0$). These assumptions could be changed, as long as the argmax in (14) is non-empty and interior for all θ, so that the allocation rule is well-defined and differentiable. The one-dimensionality of the public good could be relaxed too, at the expense of heavier notation. Note that the quadratic case in the examples of Section 2.1 is a special case of this class of environments.

for each $i \in I$. Hence, the SC-PC assumption in these environments only has bite in that it imposes a standard single-crossing condition, which requires that: (i) marginal utilities for the public good are increasing in own type; and (ii) the total *weighted* preference externality from each agent to all individuals, each weighted by his social weight, is strictly positive.

In these environments, the matrix of *canonical strategic externalities*, SE^*, is promptly obtained from the matrix of preference interdependence as follows: $SE^*_{ij} = \frac{\gamma_{ij}}{\gamma_{ii}}$ for all $j \neq i$, and $SE^*_{ii} = 0$. Relabel agents if necessary so that $\sum_{j\neq 1} \frac{\gamma_{1j}}{\gamma_{11}} < \sum_{j\neq 2} \frac{\gamma_{2j}}{\gamma_{21}} < ... < \sum_{j\neq n} \frac{\gamma_{nj}}{\gamma_{nn}}$. Then, the following results follow directly from Theorems 1, 2 and 4:

Proposition 3. *Consider a utilitarian public good problem with network effects,* Γ. *Then:*

1. *Partial* $\mathcal{B}^{id}$*-implementation is possible (w.l.o.g., via the* canonical transfers*) if* $\gamma_{ii}, \overline{\gamma}_i > 0$ $\forall i \in I$ *and only if* $\gamma_{ii}, \overline{\gamma}_i \geq 0$ $\forall i \in I$.

2. *If* $\gamma_{ii}, \overline{\gamma}_i > 0$ $\forall i \in I$, *Full-*$\mathcal{B}^{id}$ *implementation is possible (w.l.o.g., via the* loading transfers*) if and only if* $|\sum_{j\neq 1} \gamma_{1j} \sum_{j\neq 2} \gamma_{2j}| < \gamma_{11} \cdot \gamma_{22}$. *It is also possible via the* equal-externality transfers *if either (i)* $\left|\sum_{j\neq i} \gamma_{ij}\right| < \gamma_{ii}$ *for all* i, *or (ii)* $\sum_{j\neq i} \left|\frac{\gamma_{ji}}{\gamma_{jj}}\right| < 1$ *for all* i.

7 Conclusions

This paper continues a long tradition of works on implementation, which have taken up Wilson (1987) and Jackson (1992) call for a greater 'relevance' of full implementation theory, through a repeated weakening of common knowledge assumptions n the environment, and the exploration of restricted classes of mechanisms.[27] 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 ('common knowledge of identicality', CKI). Our main results characterize the transfer schemes which achieve, respectively, partial and full implementation under CKI whenever possible, as well as the conditions on agents' preferences and on the allocation rules 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 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 to characterize the transfers for full implementation) is to minimize the spectral radius of the matrix of strategic externalities; the constraints (which are dictated by incentive compatibility under the CKI restriction) require preserving the total level of such externalities.

[27] 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,b) and Ollár and Penta (2017); etc.

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, Golub and Goyal (2020) recently studied the important 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. 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 closely connected to Elliott and Golub (2019) 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 favors 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 (recent papers include Leister, Zenou and Zhou (2020), Galeotti et al. (2010), De Martí and Zenou (2015), Golub and Morris (2017), as well as others we listed in footnote 25).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 characterization of uniqueness of rationalizable solutions) for incomplete information games, with both private and interdependent values.[28]

With respect to robust mechanism design, this paper contributes to the literature which has explored environments with limited information about agents' beliefs, intermediate between the standard Bayesian settings (e.g., Postlewaite and Schmeidler (1986), Jackson (1991)), and the

[28]On a more technical note, Lemma 5 in Appendix C may also prove especially valuable, in that it provides easier to check conditions for uniqueness, in terms of both 'outgoing' and 'incoming' externalities, formalized respectively by a row- and a column-condition on the strategic externality matrix.

belief-free benchmark studied by the first-wave of the modern literature on robust mechanism design (e.g., Bergemann and Morris (2005, 2009a)). Relative to 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, this paper represents an example of a specific belief restriction based on an interesting class of economic environments (namely, the CKI assumption). As discussed, these restrictions turn out to induce a tractable mathematical structure, and also enable 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 assumptions. For instance, it would be interesting to study implementation under qualitative restrictions such as independence, or affiliation (or other ways of formalizing the idea of 'positivie correlation'), without imposing standard common knowledge assumptions of classical 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.[29] 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 A.2), is based on general game theoretic principles which may be applied to any kind of baseline mechanism.[30] The logic of our construction may thus provide useful guiding principles also for indirect implementation.

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 $m \in M \equiv \Theta$,*

$$t_i(m) = \underbrace{t_i^*(m) + \tau_i(m_{-i})}_{\substack{\textit{belief-free transfers} \\ \textit{(ep-IC characterization)}}} + \underbrace{\int_{\underline{\theta}_i}^{m_i} K_i(s_i, m_{-i})\, ds_i}_{\textit{belief-based component}} \tag{15}$$

[29] 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.

[30] For instance, in the papers mentioned in the previous footnote, the extra desiderata are achieved by adding a belief-dependent component to some baseline payments, much as Ollár and Penta (2017) or the results above attain full implementation appending an extra term to the canonical transfers.

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

$$\mathbb{E}^{b_{\theta_i}}\left(K_i\left(\theta_i, \theta_{-i}\right)\right) = 0 \text{ for all } \theta_i \text{ and for all } b_{\theta_i} \in B^{id}_{\theta_i}.^{31} \tag{16}$$

[Sufficiency:] *If* (d,t) *is twice differentiable,* t *satisfies (15) and (16), and the resulting payoffs are such that* $E^{b_{\theta_i}}\left(\partial^2 U_i\left(m_i, \theta_{-i}; \theta\right)/\partial^2 m_i\right) < 0$ *for all* m_i *and* $b_{\theta_i} \in \mathcal{B}^{id}_{\theta_i}$*, then* (d,t) *is* $\mathcal{B}^{id}$*-IC.*

Equation (15) 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 $K_i : M \to \mathbb{R}$. This result therefore extends Ollár and Penta (2017)'s characterization of ex-post incentive compatible transfers in belief-free settings to the belief restrictions $\mathcal{B}^{id}$. The sense in which the extra component is 'belief-dependent' is clarified by the condition in equation (16), 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) $E^{b_{\theta_i}}\left(\partial U_i\left(\theta_i, \theta_{-i}; \theta_i, \theta_{-i}\right)/\partial m_i\right) = 0$ and (ii) $E^{b_{\theta_i}}\left(\partial^2 U_i\left(\theta_i, \theta_{-i}; \theta_i, \theta_{-i}\right)/\partial^2 m_i\right) \leq 0$. But if t partially implements d, then by Lemma 3 it can be written as in (15), and hence – letting U^* denote the payoff function of the canonical direct mechanism – for any $\theta_i \in (\underline{\theta}, \bar{\theta})$ and $b_{\theta_i} \in B^{id}_{\theta_i}$, we have:

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

Condition (16) 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 (16) 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 (16) of Lemma 3 can be gathered by looking at the special case in which the K_i function can be written as $K_i(m) = L_i(m_{-i}) - f_i(m_i)$, for some $L_i : \Theta_{-i} \to \mathbb{R}$ and $f_i : \Theta_i \to \mathbb{R}$. Then, the expected value condition (16) can be written as

$$E^{b_{\theta_i}}\left(L_i\left(\theta_{-i}\right)\right) = f_i\left(\theta_i\right) \text{ for all } \theta_i \text{ and for all } b_{\theta_i} \in B^{id}_{\theta_i}. \tag{17}$$

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 (17) 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:

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

Definition 10. *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 (17) 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.(15) specializes to

$$t_i(m) = \underbrace{t_i^*(m) + \tau_i(m_{-i})}_{\substack{\text{characterization} \\ \text{of ep-IC transfers}}} + \underbrace{L_i(m_{-i})\, m_i - \int^{m_i} f_i(s_i)\, ds_i}_{\text{moment condition-based term}}. \tag{18}$$

This is precisely the class of transfers for which Ollár and Penta (2017) provide sufficient conditions for full implementation.[32] By Lemma 3, there may exist incentive compatible transfers which cannot be written as in equation (18), since not all functions $K_i : \Theta \to \mathbb{R}$ in that Lemma are equivalent to moment conditions in the sense of Definition 10. 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. ■

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

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

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

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

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

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

Let the difference between t_i and t_i^* be $D_i\left(m\right):=t_i\left(m\right)-t_i^*\left(m\right)$. 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 $\tau_i\left(m_{-i}\right):=D_i\left(m\right)-\int_{\underline{\theta}}^{m_i}\frac{\partial D_i}{\partial m_i}\left(s_i,m_{-i}\right)ds_i$, and further let $K_i\left(m\right):=\partial D_i\left(m\right)/\partial m_i$ for all m. Then, the transfer t_i takes the form $t_i\left(m\right)=t_i^*\left(m\right)+\tau_i\left(m_{-i}\right)+\int_{\underline{\theta}}^{m_i}K_i\left(s_i,m_{-i}\right)ds_i$ for all m and K_i satisfies the expected value condition in (16). 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 (15) and the expected value condition in (16), then

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

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

Proof of Theorem 1.

Step 1: If $K_i:M\to\mathbb{R}$ satisfies condition (16), then for all θ_i $E^{b_{\theta_i}}\left(K_i\left(m_i,\theta_{-i}\right)\right)=0$ 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 16, $\mathbb{E}^{b_{\theta_i}}\left(K_i\left(\theta_i,\theta_{-i}\right)\right)=0$ for all θ_i and for all $b_{\theta_i}\in B^{id}_{\theta_i}$. Fix $p\in B^{id}_{\theta_i}$. It is a consequence of identicality that if $p\in B^{id}_{\theta_i}$, then $p\in B^{id}_{m_i}$ for all $m_i\in[\underline{\theta},\overline{\theta}]$, that is $\mathbb{E}^{p}\left(K_i\left(m_i,\theta_{-i}\right)\right)\equiv 0$ as a function of m_i, and this holds for any $p\in B^{id}_{\theta_i}$, which proves this Step.[33] □

To show the Theorem, if (d,t) partially implements d, then by Lemma 3, t can be written as in (15), and hence – letting U^* denote the payoff function of the canonical direct mechanism – for

[33] 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.

any θ_i and $b_{\theta_i} \in \mathcal{B}^{id}$:

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

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. ■

B Proofs of Results from Section 4

Proof of Lemma 1.[34] **(i)** (Sufficiency: Eigenvalue Condition for Full Implementation.)[35] Fix θ_i *in* $\left(\underline{\theta},\overline{\theta}\right)$ and examine the k-th round of eliminations: fix $m_i \in R_i^k\left(\theta_i\right)$. Thus for m_i, there exists a conjecture which supports m_i as a best reply and is concentrated on R_{-i}^{k-1}. Let this conjecture be μ_L. 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 $E^{\mu}U_i\left(m_i;\theta_i\right)$ 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

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

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

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

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

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,

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

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

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

After examining the signs of $\partial_{ii}^2 E^{\mu_L}U_i\left(s_i;\theta_i\right)$ and the respective signs of $\left(m_i-\theta_i\right)$ in the latter

[34] 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.

[35] 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 $\rho(A) := \lim_{K\to\infty}\rho\left(A_K\right)$. Beyond the standard extensions of operators, we adopt the understanding that $0/0 = \infty$ and $\infty/\infty = \infty$.

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

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

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

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

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

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

Since $\rho(|SE_{max}^t|) < 1$, we have $|SE_{max}^t|^k \to \mathbf{0}$, 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 $|SE_{min}^t|$ is irreducible and features no 0 eigenvalues. (Such subset $I_E \subseteq I$ of the agents exists and, since $\rho(|SE_{min}^t|) > 1$ 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 $|SE_{min}^t|$. We will show next, that for each round k for some $i \in I_E$, there is a best reply outside the open set $\left(\theta_i \pm \left[E \cdot \mathbf{l}_{min,E}^{k-1}\right]_i\right) \cap$ int cl $R_i^{k-1}(\theta_i)$. The notation $\mathbf{l}_{min,E}^k$ is such that: for each agent $j \in I_E$ and round k, let $l_{j,min,E}^k := \inf_{\theta_j} \min\left\{\sup_{m_j \in R_j^k(\theta_j); m_j \leq \theta_j}(\theta_j - m_j); \sup_{m_j \in R_j^k(\theta_j); m_j > \theta_j}(m_j - \theta_j)\right\}$, and let $l_{j,min}^0 := l_j = \overline{\theta}_j - \underline{\theta}_j$.[36]

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 μ_L^E that is consistent with $\mathcal{B}$, with round $k-1$ rationalizability, and is such that for all $j \in I_E$, $\text{marg}_{M_j}\mu_L^E$ places probability one on positive misreports that are $l_{j,min,E}^k$ apart from θ_j if the absolute smallest $\partial_{ij}^2 U_i^t$ is positive and on negative misreports if it is negative; and for all $j \notin I_E$, $\text{marg}_{M_j}\mu_L^E$ places probability one on the true type θ_j being reported. Now, if the considered m_i is an extremal point of cl $R_i^{k-1}(\theta_i)$, then we are done. However, if it is an internal point, then $\partial_{ii}^2 E^{\mu_L^E} U_i^t(m_i) \leq 0$ and there is a small ε such that the modified function $E^{\mu_L^E} U_i^{t,\varepsilon} := E^{\mu_L^E} U_i^t(s_i) - \varepsilon(s_i - m_i)^2$ 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

[36]The intuition for $\mathbf{l}_{min,E}^k$ 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.

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

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

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

which further implies for such m_i that

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

Thus, summarizing this, for each k, there is a kth round rationalizable message that is outside the set $\left(\theta_i \pm \left[E\cdot \mathbf{l}_{min,E}^{k-1}\right]_i\right) \cap \text{int cl} R_i^{k-1}(\theta_i)$, which when iterated gives that it is outside the set $\left(\theta_i \pm \left[E^k\cdot \mathbf{l}_{min,E}^{0}\right]_i\right) \cap \left(\underline{\theta}_i, \overline{\theta}_i\right)$. Iteratively, one can see that $\mathbf{l}_{min,E}^0, \mathbf{l}_{min,E}^1$ are strictly positive. Assuming that $\mathbf{l}_{min,E}^{k-1}$ is strictly positive, and by the irreducibility of the non-negative E, we have that $\mathbf{l}_{min,E}^{k}$ is strictly positive. From this, we can see that if the spectral radius $\rho(|SE_{min}^t|) \geq 1$, then the sequence $\left\{E^k\right\}_{k=1}^{\infty}$ of nonnegative 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 (16) 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 $\left\{H_{ij}^k\right\}_{j\neq i, k\in\mathbb{N}}$ 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 (16) 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 (16) 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 $b_{\theta_i} \in B_{\theta_i}^{id}$ 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.[37]

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

$$H_k\left(\theta^1, \theta^2, \ldots, \theta^2\right) - H_k\left(\theta^2, \theta^2, \ldots, \theta^2\right) \neq H_k\left(\theta^1, \theta^1, \ldots, \theta^1\right) - H_k\left(\theta^2, \theta^1, \ldots, \theta^1\right). \tag{19}$$

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

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

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

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

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

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

The last negation follows from Equation (19), which recall was the consequence of non-separability, and this negation implies that $\mathbb{E}^{p^{indep}}\left(H_k\left(\theta_{-i}\right)\right) \neq \mathbb{E}^{p^{corr}}\left(H_k\left(\theta_{-i}\right)\right)$, which would imply the contradiction that K_i does not satisfy the expected value condition. And therefore, H_k must be separable.

Step 1b: (Each H_k gives 0 at identical profiles.) Fix a type θ_i. Consider beliefs of i which are identical point-distributions; distributions which are concentrated on the same type of all other agents. Formally, consider a belief b_{θ_i} such that, for some $\theta \in [\underline{\theta}, \overline{\theta}]$, the probability $b_{\theta_i}\left(\{\theta_j = \theta \text{ for all } j \neq i\}\right)$ is 1 for all $j \neq i$. Then, b_{θ_i} is included in $B^{id}_{\theta_i}$, moreover such point-

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

beliefs exist for all θ. Fix this (independent) point belief b_{θ_i}. The expected value condition implies that for the polynomial format $0 \equiv \sum_{k=1}^{\infty} \mathbb{E}^{b_{\theta_i}}\left(H_k\left(\theta_{-i}\right)\right)\theta_i^k$ and thus for any k $\mathbb{E}^{b_{\theta_i}}\left(H_k\left(\theta_{-i}\right)\right)=0$. At identical profiles as represented by b_{θ_i}, this latter means that $H_k(\theta,\theta\ldots,\theta)=0$ for all $\theta\in[\underline{\theta},\overline{\theta}]$, which proves that the H_k are 0 at identical profiles.

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

$$\begin{aligned}\mathbb{E}^{b_{\theta_i}}\left(K_i\left(\theta\right)\right) &= \int_{\Theta_{-i}}\sum_{k=1}^{\infty}H_k\left(\theta_{-i}\right)\theta^k db_{\theta_i} = \int_{\Theta_{-i}}\sum_{k=1}^{\infty}\sum_{j\neq i}H_{kj}\left(\theta_j\right)\theta^k db_{\theta_i}\\ &= \sum_{k=1}^{\infty}\sum_{j\neq i}\left[\int_{\Theta_j}H_{kj}\left(\theta_j\right)d\operatorname*{marg}_{\Theta_j}b_{\theta_i}\right]\theta^k \qquad (20)\end{aligned}$$

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 (20) is

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

and the two conditions,

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

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

If K_i satisfies the expected value condition in 15, then based on the characterization in Step 1 of Proof of Lemma 1, we have

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

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

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 15; and is such that $\partial U_i^t\left(m;\theta\right)/\partial m_i=\partial U_i^*\left(m;\theta\right)/\partial m_i+K_i\left(m_i,m_{-i}\right)$. 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. 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 $R_i^{id,k}\left(\theta_i|t^l\right)$ is a closed interval around θ_i.[38]

[38]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.

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

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

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

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

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

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

In the second round of iterations:

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

In the third round of iterations:

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

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

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

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

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

In words, this means that in every *even round* of iteration, for each type of agent 1, the rationalizable set is either given by the previous rationalizable set or it is shrank to $|\xi_2|$ of this set and, for each type of agent $j\neq 1$, the rationalizable set is either the previous rationalizable set or it is shrank to $|\xi_1|$ of this set. 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 $R_i^{id}\left(\theta_i|t^l\right)=\{\theta_i\}$ for all i for all θ_i. On the other hand, if $|\xi_1\xi_2|\geq 1$, then $|\xi_2|\geq 1$ and in every round k, $R_2^{id,k}\left(\theta_2|t^l\right)=\left[\theta_2\pm\left(\overline{\theta}-\underline{\theta}\right)\right]\cap\left[\underline{\theta},\overline{\theta}\right]=\left[\underline{\theta},\overline{\theta}\right]$, 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, 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 6. 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

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

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

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

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

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

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

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

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

where the first equality holds because of the canonical 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 $\left(\theta_i,s_{-i}^o\right)$.

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

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

and there exists s_o such that

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

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

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

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

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

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

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

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

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

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

which gives that, for some profile,

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

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

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

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

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

which, completes the proof of this Step. □

Step 3 is the key step in establishing the if and only if result. In words, it implies that in any $\mathcal{B}^{id}$-implementing direct mechanism, the externalities can not be reduced beyond the sum of externalities in the canonical direct mechanism. The consequence of such irreducibility of externalities is reflected in each k-rationalizable set of the step-by-step iterations; for all $\mathcal{B}^{id}$-IC t. 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 nessecary externalitites and therefore leads to full implmenetation 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 loaded direct mechanism $R_i^{id,k}\left(\theta_i | t^l\right)$ are contained in $R_i^{id,k}\left(\theta_i | t\right)$, for any partially implementing direct mechanism (d, t).

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

following set:

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

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

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

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

For the second round of iteration,

$$[\theta_1-\xi_1\xi_2,\theta_1+\xi_1\xi_2]\cap\left[\underline{\theta},\overline{\theta}\right]\subseteq\ R_i^{id,2}\left(\theta_i|t\right)\ \text{ if } i=1 \text{ and}$$
$$[\theta_i-\xi_i\xi_1,\theta_i+\xi_i\xi_1]\cap\left[\underline{\theta},\overline{\theta}\right]\subseteq\ R_i^{id,2}\left(\theta_i|t\right)\ \text{ if } i\neq 1.$$

For the third round of iteration,

$$[\theta_1-\xi_1\left(\xi_1\xi_2\right),\theta_1+\xi_1\left(\xi_1\xi_2\right)]\cap\left[\underline{\theta},\overline{\theta}\right]\subseteq\ R_i^{id,3}\left(\theta_i|t\right)\ \text{ if } i=1 \text{ and}$$
$$[\theta_i-\xi_i\left(\xi_1\xi_2\right),\theta_i+\xi_i\left(\xi_1\xi_2\right)]\cap\left[\underline{\theta},\overline{\theta}\right]\subseteq\ R_i^{id,3}\left(\theta_i|t\right)\ \text{ if } i\neq 1.$$

For the forth round of iteration,

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

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

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

Proof of Theorem 3. Fix an environment (v,d), and consider $\bar{t}^l$. Note that these transfers ensure $\mathcal{B}^{id}$-IC, moreover, the resulting strategic externalitites are such that for all i,j, $SE_{ij}^{\bar{t}^l}(m;\theta)\in\bar{SE}_{ij}^{\bar{t}^l}\pm\alpha_{ij}$. From this, for all $(m;\theta)$, the following matrix inequalities hold element-wise: $||\bar{SE}^{\bar{t}^l}|-A|\leq|SE^{\bar{t}^l}(m;\theta)|\leq|\bar{SE}^{\bar{t}^l}|+\mathcal{A}$.

For part 1, by Gelfand's formula[39] we have that $\rho\big(|SE^{\bar{t}^l}(m;\theta)|\big)\leq\rho\big(|\bar{SE}^{\bar{t}^l}|+\mathcal{A}\big)$ for all (m,θ), and thus $\rho\big(|SE_{max}^{\bar{t}^l}|\big)\leq\rho\big(|\bar{SE}^{\bar{t}^l}|+\mathcal{A}\big)$. Recall that $\bar{\xi}_i$ is the row-sum of the midpoints of strategic

[39]Gelfand's formula characterizes the spectral radius of a matrix A such that $\rho(A)=\lim_{k\to\infty}||A^k||^{1/k}$. Using Gelfand's formula, if a non-negative matrix A is element-wise dominated by a non-negative matrix B, that is if $A_{ij}\leq B_{ij}$ for all i,j, then $\rho(A)\leq\rho(B)$.

externalitites affecting agent i and that the upper bounding matrix is such that

$$|\bar{SE}^{\bar{t}^l}| + \mathcal{A} = \begin{bmatrix} 0 & |\bar{\xi}_1| + \alpha_1 & \alpha_1 & \dots & \alpha_1 \\ |\bar{\xi}_2| + \alpha_2 & 0 & \alpha_2 & \dots & \alpha_2 \\ |\bar{\xi}_3| + \alpha & \alpha & 0 & \dots & \alpha \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ |\bar{\xi}_n| + \alpha & \alpha & \alpha & \dots & 0 \end{bmatrix}.$$

We need to study the eigenvalues of this matrix. Assume that $n > 2$.[40] Notice that this matrix has $n-3$ independent eigenvectors of the form $(0, \dots, 0, 1, -1, 0, \dots, 0)$, with eigenvalues $-\alpha$. Let $\lambda_1, \lambda_2, \lambda_3$ denote the other three eigenvalues and let λ_1 denote the leading eigenvalue. From the determinant of $|\bar{SE}^{\bar{t}^l}| + \mathcal{A}$,[41] the trace of $|\bar{SE}^{\bar{t}^l}| + \mathcal{A}$ and the trace of $\left(|\bar{SE}^{\bar{t}^l}| + \mathcal{A}\right)^2$, we have that

$$\lambda_1 \lambda_2 \lambda_3 = H \tag{21}$$

$$\lambda_1 + \lambda_2 + \lambda_3 = (n-3)\,\alpha \tag{22}$$

$$\lambda_1^2 + \lambda_2^2 + \lambda_3^2 = 2K + (n-3)^2\,\alpha^2 \tag{23}$$

where

$$H = \left(\alpha_2 |\bar{\xi}_1| + \alpha_1 \alpha_2\right) \sum_{i \neq 1,2} |\bar{\xi}_i| + \left(\alpha_2 |\bar{\xi}_1| + \alpha_1 |\bar{\xi}_2|\right) \alpha + (n-1)\,\alpha_1 \alpha_2 \alpha - (n-3)\,\alpha |\bar{\xi}_1 \bar{\xi}_2|$$

$$K = |\bar{\xi}_1 \bar{\xi}_2| + \alpha_2 |\bar{\xi}_1| + \alpha_1 \sum_{i \neq 1} |\bar{\xi}_i| + \alpha_1 \alpha_2 + (n-2)\,(\alpha_1 + \alpha_2)\,\alpha.$$

The difference between the square of (22) and (23) gives that $\lambda_1\lambda_2 + \lambda_1\lambda_3 + \lambda_2\lambda_3 = -K$. Using Vieta's formulas we can relate the (possibly complex) eigenvalues to the three complex roots of the following cubic:[42]

$$x^3 - (n-3)\,\alpha x^2 - Kx - H = 0.$$

By the Perron-Frobenius theorem, λ_1 is real and $\lambda_1 \geq 0$. And thus λ_1 is the largest root of the cubic. Observe that the derivative of this cubic is decreasing at $x = 0$ and that the cubic is infinite at infinity. This means that $\lambda_1 < 1$ if and only if $x = 1$ is on the increasing positive side and the value at 1 is positive. That is $\lambda_1 < 1$ if and only if $LHS'(1) = 3 - 2(n-3)\,\alpha - K > 0$ and $H + K + (n-3)\,\alpha < 1$. These two inequalities are equivalently given in part 1 and thus they imply that $\lambda_1 < 1$.[43] Thus part (i) of Lemma 1 implies that $\bar{t}^l$ ensures full $\mathcal{B}^{id}$-implementation.

For part 2, the lower bounding inequality $||\bar{SE}^{\bar{t}^l}| - A| \leq |SE^{\bar{t}^l}(m;\theta)|$ for all $(m;\theta)$ implies that $\rho\left(||\bar{SE}^{\bar{t}^l}| - A|\right) \leq \rho\left(|SE^{\bar{t}^l}_{min}|\right)$. The lower bounding matrix has an upper left 2 by 2 block

[40] If $n = 2$, then $\rho\left(|\bar{SE}^{\bar{t}^l}| + \mathcal{A}\right) = \sqrt{\left(|\bar{\xi}_1| + \alpha_1\right)\left(|\bar{\xi}_2 + \alpha_2|\right)}$.

[41] For this calculation, use Schur's determinant identity such that $\det\left(|\bar{SE}^{\bar{t}^l}| + \mathcal{A}\right) = \det(D)\det\left(A - BD^{-1}C\right)$, where A is the upper left 2 by 2 block, D is the lower right, B is the upper right and C is the lower left block.

[42] Vieta's formulas for cubic polynomials give that $(x-\lambda_1)(x-\lambda_2)(x-\lambda_3) = x^3 - (\lambda_1+\lambda_2+\lambda_3)\,x^2 + (\lambda_1\lambda_2 + \lambda_1\lambda_3 + \lambda_2\lambda_3)\,x - \lambda_1\lambda_2\lambda_3$.

[43] In the special case when $\alpha_1 = \alpha_2 = 0$, we have $K = |\bar{\xi}_1\bar{\xi}_2|$ and $H = -(n-3)\,\alpha|\bar{\xi}_1\bar{\xi}_2|$. The inequalitites in part 1 are $2(n-3)\,\alpha + |\bar{\xi}_1\bar{\xi}_2| < 3$ and $|\bar{\xi}_1\bar{\xi}_2| - (n-3)\,\alpha|\bar{\xi}_1\bar{\xi}_2| + (n-3)\,\alpha < 1$; which are equivalent to the system $(|\bar{\xi}_1\bar{\xi}_2| < 1,\ (n-3)\,\alpha < 1)$. Further, regarding Remark 1 – using Vieta's formulas in the two inequalities – notice that if $H > (n-3)\,\alpha - 2$, then inequality (i) implies (ii).

whose spectral radius is $\sqrt{|\left(|\bar{\xi}_1|-\alpha_{12}\right)\left(|\bar{\xi}_2|-\alpha_{21}\right)|}$ which by assumption is at least as large as 1. Thus by part (ii) of Lemma 1, $\vec{t}^{\,l}$ fails $\mathcal{B}^{id}$-implementation, and part 2 follows. ■

C Further Design Strategies for Full Implementation

In this Section we consider alternative design strategies for full implementation, and we prove the results in Section 5. The next Lemma provides general sufficient conditions which will be useful for the following discussion:

Lemma 5. *The $\mathcal{B}^{id}$-IC transfer scheme t achieves full $\mathcal{B}^{id}$-implementation if either: (i) it ensures* limited strategic externalities from other agents – *that is, if $\sum_{j\neq i}|SE^t_{max}|_{ij}<1$ for all i; or (ii) it ensures* limited strategic impact on other agents – *that is, if $\sum_{j\neq i}|SE^t_{max}|_{ji}<1$ for all i.*

Proof of Lemma 5. By the Gershgorin circle theorem, both under condition (i) and (ii) the absolute value of all eigenvalues of $|SE^t_{max}|$ are smaller than 1, which by Lemma 1 ensures full $\mathcal{B}^{id}$-implementation. ■

The condition in the first point of this Lemma resembles the design principle in Ollár and Penta (2017), in that it requires 'not too strong' *strategic externalities.*[44] Formally, it is a row-wise condition on the $|SE^t_{max}|$-matrix. The second condition instead is a column-wise restriction on $|SE^t_{max}|$, which can be interpreted as requiring that any agent's *strategic impacts* on others is not too strong.

Theorem 4 in Section 5 follows directly from Lemma 5:

Proof of Theorem 4. Under the SC-PC assumption, the equal-externality transfer scheme t^e is $\mathcal{B}^{id}$-IC. Moreover, t^e induces a strategic externality matrix which is such that for all $i,j\neq i$, $SE^e_{ij}=\left(\sum_{j\neq i}\frac{\partial^2 v_i}{\partial x\partial\theta_j}/\frac{\partial^2 v_i}{\partial x\partial\theta_i}\right)\frac{1}{n-1}$. For this externality matrix, notice that condition (i) of this Proposition implies condition (i) of Lemma 5; and condition (ii) of this Proposition implies condition (ii) of Lemma 5, and thus by Lemma 5, full $\mathcal{B}^{id}$-implementation follows. ■

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

Proposition 4. *Under SC-PC, if one of the conditions in Lemma 5 are satisfied by* some *$\mathcal{B}^{id}$-IC transfer scheme t , then the equal-externality transfers $(t^e_i)_{i\in I}$ achieve full $\mathcal{B}^{id}$-implementation.*

Proof of Proposition 4. Under SC-PC, t^e ensures $\mathcal{B}^{id}$-incentive compatibility. Next, we show that t^e ensures full $\mathcal{B}^{id}$-implementation too.

First, assume that there exists a transfer scheme t which ensures full $\mathcal{B}^{id}$-implementation and limited strategic externalities as in (i) of Lemma 5.

By the characterization of belief-based terms for $\mathcal{B}^{id}$-IC in Lemma 2, there exists (m,θ) for which $\sum_{j\neq i}SE^t_{ij}(m;\theta)=\sum_{j\neq i}SE^*_{ij}$. Next we show that t^e induces an externality matrix which satisfies the conditions of the eigenvalue lemma in part (i) of Lemma 1. By construction of t^e,

[44] Unlike here, the analysis in Ollár and Penta (2017) was limited to transfers based on linear moment conditions, a special case of transfers which are linear in own report.

$\sum_{j\neq i}|SE^e_{ij}| = \sum_{j\neq i}|\sum_{j\neq i}SE^*_{ij}/(n-1)| = |\sum_{j\neq i}SE^*_{ij}|$. And thus, there exists (m,θ) such that $\sum_{j\neq i}|SE^e_{ij}| = |\sum_{j\neq i}SE^t_{ij}(m;\theta)| \leq \sum_{j\neq i}|SE^t_{ij}(m;\theta)| < 1$. The latter strict inequality holds by (i) of Lemma 5 and thus by the Gershgorin circle theorem, $\rho(|SE^e|) < 1$ and thus by (i) of Lemma 1, t^e too ensures full $\mathcal{B}^{id}$-implementation.

Second, assume that there exists a transfer scheme t which ensures full $\mathcal{B}^{id}$-implementation and limited strategic impacts as in (ii) of Lemma 5.

By the characterization of belief-based terms for $\mathcal{B}^{id}$-IC in Lemma 2, there exists (m,θ) for which $|\sum_{i\in I}\sum_{j\neq i}SE^*_{ij}| = |\sum_{i\in I}\sum_{j\neq i}SE^t_{ij}(m;\theta)| \leq \sum_{i\in I}\sum_{j\neq i}|SE^t_{ij}(m;\theta)|$ and thus, by t satisfying (ii) of Lemma 5, $|\sum_{i\in I}\sum_{j\neq i}SE^*_{ij}| < n$ and writing this with the total externality notation, $\sum_{i\in I}\xi_i < n$. Now, consider the absolute externality matrix induced by the equal-externality transfers t^e. In what follows, using the Perron-Frobenius theorem, we show that this matrix has a spectral radius which is less than 1. $|SE^e|$ is a non-negative matrix, with zeros in its diagonal and by its construction, for all i and $j \neq i$, $|SE^e|_{ij} = |\sum_{j\neq i}SE^*_{ij}|/(n-1)$, in other notation, $|SE^e|_{ij} = |\xi_i|/(n-1)$. Let ρ denote the largest eigenvalue of this matrix. (Assume that ξ_is, the absolute total canonical externalities, are ordered as before, based on their absolute values, from the smallest to the largest.) By the Perron-Frobenius theorem, there is a positive 1-norm vector $v \in \mathbb{R}^n$ such that $\rho v = |SE^e|v$. The componentwise consequence of this is that, for all i, $\rho\frac{v_i}{|\xi_i|} = \frac{\sum_{j\neq i}v_j}{n-1}$, which also implies that if $|\xi_i| \leq |\xi_j|$, then $v_i \geq v_j$. Adding up these n equations and expressing ρ, gives that $\rho = \frac{\sum_{j\in I}v_i}{\sum_{j\in I}\frac{v_i}{|\xi_i|}}$, which is a weighted harmonic mean of ξ_is with weights v_is. And thus from a weighted harmonic mean – arithmetic mean inequality, $\rho = \frac{\sum_{j\in I}v_i}{\sum_{j\in I}\frac{v_i}{|\xi_i|}} \leq \sum_{i\in I}\frac{v_i}{\sum_{j\in I}v_j}|\xi_i|$. Since larger $|\xi_i|$s have smaller weights, this latter expression is bounded by the average of $|\xi_i|$s, and thus $\rho \leq \sum_{i\in I}\frac{|\xi_i|}{n} < 1$. Recall from above that, the latter strict inequality is a consequence of t satisfying (ii) of Lemma 5. Therefore, by (i) of Lemma 1, t^e ensures full $\mathcal{B}^{id}$-implementation. ■

Corollary 2. *If Condition (12) holds, then both t^* and t^e ensure full $\mathcal{B}^{id}$-implementation.*

Hence, whenever there is an implementing transfer scheme which satisfies the easy-to-check conditions of Lemma 5, then the equal-externality transfers t^e also achieve full $\mathcal{B}^{id}$-implementation. There are, however, environments in which the canonical transfers t^* achieve full $\mathcal{B}^{id}$-implementation, but the equal-externality transfers t^e do not:

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

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

In this example, the $|SE^*|$-matrix has spectral ratio less than 1, however the $|SE^e|$-matrix has an eigenvalue larger than 2. Here the canonical transfers coincide with the loading transfers, and so achieve full implementation, but the equal-externality transfers do not.[45] □

[45] For cases in which, contrary to this example, the canonical transfers fail full implementation but the transfers with uniformly redistributed externalities work well, see Examples 1.1 and 3.

We conclude this Section of the Appendix C with the proof of the last result in Section 5:

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

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

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

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

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

To prove part 1 of this proposition, note that from Theorem 2, if the equal-externality mechanism (d,t^e) achieves full $\mathcal{B}^{id}$-implementation, then the loaded direct mechanism $\left(d,t^l\right)$ achieves this too. To prove the other direction, note that if the loaded direct mechanism $\left(d,t^l\right)$ achieves full $\mathcal{B}^{id}$-implementation, then by the previous part of this proof, we have $\left|\sum_{k\neq i}\frac{\partial^2 v_i}{\partial x\partial\theta_k}\right|<\frac{\partial^2 v_i}{\partial x\partial\theta_i}$ for every i, and (i) of Theorem 4 implies that the equal-externality transfers ensure full $\mathcal{B}^{id}$-implementation too, which completes the proof of this part. ■

D Sensitivity Results

D.1 Proof of Theorem 6

The proof of Theorem 6 relies on the following lemma, which characterizes the set of possible misreports at each iteration of the F_ε-rationalizability procedure:

Lemma 6. *Consider an SC-PC environment and linear moment conditions in the fully $\mathcal{B}^{id}$-implementing t. For given ε and F, the largest set of reports in $R_i^{F_\varepsilon}$ is the largest element of the vector $\left[\ I-|SE^t|\ \right]^{-1}C\boldsymbol{\varepsilon}^F l$, where C is the matrix such that $C_{ij}=1/|\partial^2_{ii}U_i|$ if $i=j$ and $C_{ij}=0$ otherwise, $\boldsymbol{\varepsilon}^F=(\varepsilon_i)_{i\in I}$ is the vector such that $\varepsilon_i=\varepsilon$ if $i\in F$ and $\varepsilon_i=0$ if $i\notin F$ and $l=\overline{\theta}-\underline{\theta}$.*

Proof of Lemma 6 Consider an arbitrary non-negative vector $\varepsilon=(\varepsilon_i)_{i\in I}$ .Recall that it is the consequence of SC-PC and linear moment conditions that in the utility functions of the direct mechanism (d,t), the second order derivatives are constants. This latter combined with the characterization of the best reply sets in Step 3 of Proof of Theorem 2 implies that

for the first round of iterative eliminations with ε-faulty agents, with notation $A_1 := |SE^t| + \varepsilon^T C$ and $l := \overline{\theta} - \underline{\theta}$, we have that for all θ_i

$$R_i^{\mathcal{B}^{id},\varepsilon,1}(\theta_i) = [\theta_i \pm [A_1 \mathbf{1} l]_i] \cap [\underline{\theta}, \overline{\theta}] .$$

In the second round of iterative eliminations, for any $m_i \in R_i^{\mathcal{B}^{id},\varepsilon,2}(\theta_i)$,

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

Moreover, applying Step 3 of Proof of Theorem 2, in this second round of iterative eliminations, with notation $A_2 := |SE^t|^2 + \varepsilon^T |SE^t| C + \varepsilon^T C$ for all θ_i the rationalizable messages are

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

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

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

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

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

Applying this formulat to ε-faulty agents with F_ε complates the proof of this Lemma.■

Proof of Theorem6. For the loading transfers t^l, the inverse of $I - |SE^l|$ is as follows:

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

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

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

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

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

Applying Lemma 6 to these inverses, with the notation that $1/c := \partial^2_{x,\theta_i} v_i$, we have that for the loading transfers, for all $n_f > 1$,

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

and for the equal-exgternality transfers,

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

Comparing

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

is equivalent to comparing

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

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

Proof of Theorem 5. See Step 4 in the Proof of Theorem 2. ■

D.2 Sensitivity to Lower Orders of Rationality and Robust Level-k Implementation

In an important recent paper, de Clippel et al. (2018) have studied a notion of level-k implementation which, for the class of environments and the direct mechanisms we consider, can be described as follows: Let $p \in \Delta(\Theta)$ denote a common prior, and $\Theta = \times_{i \in I} \Theta_i$. For any direct mechanism (d, t), let Σ_i denote the set of strategies $\sigma_i : \Theta_i \to M_i$ ($M_i = \Theta_i$ in the direct mechanism). Each player is characterized by an anchor, $\alpha_i : \Theta_i \to \Delta(M_i)$, which specifies the message chosen in the

mechanism by the non-strategic level-0 type. As usual, let α and α_{-i} denote the profiles of anchors (with independent randomization across players).

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

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

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

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

Compared to the previous literature on level-k implementation, de Clippel et al. (2018)'s notion is more robust in that it doesn't rely on the designer's knowledge of the agents' levels of sophistication: implementation is required to be achieved for all k. Their results are also more general than previous analysis in that they provide results for various anchors α. Their analysis, however, maintains the classical assumption of a commonly known prior $p \in \Delta(\Theta)$.[46] But this notion of implementation can be easily adapted to our belief restrictions, $\mathcal{B}^{id} = ((B_{\theta_i}^{id})_{\theta_i \in \Theta_i})_{i \in I}$, by replacing the solution concept in the above definition to the following weaker version:

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

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

As de Clippel et al. (2018) remark, the behavioral anchors are completely arbitrary, they may be mechanism specific and may differ across agents. In their setting, however, it is still the case that anchors $\alpha_j : \Theta_j \to M_j$ are common knowledge among the agents, and also known to the designer. A natural strengthening of the implementation requirement would be to allow for different players to have different views about others' anchors, or be uncertain over them, or on others' views about anchors, and so on. And, most importantly, without requiring that the designer knows each player's anchor, nor his beliefs about others', at any order. If we let possible anchors in each player i's mind to be any $\alpha_{-i} : \Theta_{-i} \to \Delta(M_{-i})$ – i.e., also allowing for possible correlations – then we obtain the following solution concept for robust level-k implementation:

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

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

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

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

It is easy to verify that, for every k, $\sigma_i \in RL_i^k$ if and only if $\sigma_i(\theta_i) \in R_i^{id,k}(\theta_i)$ for every θ_i. Hence, if one wishes to obtain full implementation for every k – i.e., level-k implementation à la de Clippel et al. (2018), but in the much more robust specification for what concerns agents' anchors – then one needs to obtain implementation in $\mathcal{B}$-dominant strategies, because it requires $R_i^{id,1}(\theta_i) = \{\theta_i\}$ for every θ_i. If that can be obtained, as for instance Ollár and Penta (2017) show in SC-PC environments with independent or affiliated common priors, then the result follows for all levels, and hence *interim $\mathcal{B}$-Dominant Strategy Incentive Compatibility* (iDSIC) characterizes this notion of *robust level-k implementation.*[47] But iDSIC is very demanding, and in particular under the $\mathcal{B}^{id}$-restrictions it cannot be satisfied outside of the very special case of private values. It is then natural to ask what is the best that one could obtain, if such stricter notion of implementation cannot be obtained for every k. One possibility is to ensure that, for each k, the $R_i^{id,k}$-sets are as small as possible around the truthful revelation profile. The result in Theorem 5 addresses precisely this question, and implies that the loading transfers introduced above are optimal with respect to this notion of *robust level-k $\mathcal{B}^{id}$-implementation.*

References

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

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

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

Ballester, C., A. Calvó-Armengol and Y. Zenou, *Who's Who in Networks. Wanted: The Key Player*, Econometrica 74(5), 1403-1417, 2006.

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

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

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

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

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

[47] Albeit in a complete information setting, the above mentioned paper by Saran (2016) shows that, even without the restriction to the class of mechanisms, strategy-proofness is necessary for implementation in his setting, which is in line with our observation that $\mathcal{B}^{id}$-dominant strategy IC is necessary in our setting.

Blume, L. E., W. A. Brock, S. N. Durlauf and R. Jayaraman, *Linear Social Interactions Models*, Journal of Political Economy, 123(2), 444-496, 2015.

Bramoullé, Y. and R. Kranton, *Public Goods in Networks*, Journal of Economic Theory 135(1), 478-494, 2007.

Bramoullé, Y., R. Kranton, M. D'amours, *Strategic Interaction and Networks*, American Economic Review 104(3), 898-930, 2014.

Calvó-Armengol, A. and J. De Martí, *Communication Networks: Knowledge and Decisions*, American Economic Review 97(2), 86-91, 2007.

Calvó-Armengol, A., J. De Martí and A. Prat, *Communication and Influence*, Theoretical Economics 10(2), 649-690, 2015.

Catonini, E., *Self-enforcing Agreements and Forward Induction Reasoning*, The Review of Economic Studies 88(2), 610?642, 2021.

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

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

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

Duggan, J. and J. Roberts, *Implementing the Efficient Allocation of Pollution*, American Economic Review, 92, 1070–1078, 2002.

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

de Clippel, G., R. Saran and R. Serrano, *Continuous Level-k Mechanism Design*, working paper.

De Martí, J. and Y. Zenou, *Network Games with Incomplete Information*, Journal of Mathematical Economics 61, 221-240, 2015.

Cremer, J. and R.P. McLean, *Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions*, Econometrica, 1247–1257, 1988.

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

Elliott, M. and B. Golub, *A Network Approach to Public Goods*, Journal of Political Economy, 127(2), 2019.

Fainmasser and Galeotti, *Pricing Network Effects*, Review of Economic Studies, 1-36, 2015.

Galeotti, A., S. Goyal, M. O. Jackson, F. Vega-Redondo, and L. Yariv, *Network Games*, Review of Economic Studies 77 (1), 218-244, 2010.

Galeotti, A., B. Golub and S. Goyal, *Targeting Interventions in Networks*, Econometrica 88(6), 2445-2471, 2020.

Golub, B., and S. Morris, *Expectations, Networks, and Conventions*, arxiv, 2017.

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

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

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

Jackson, M. O., *Bayesian Implementation*, Econometrica, 59(2), 461–477, 1991.

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

Jehiel, P., L. Lamy, *A Mechanism Design Approach to the Tiebout Hypothesis*, Journal of Political Economy 126(2), 735-760, 2018.

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

Kunimoto, T., R. Saran and R. Serrano, *Interim Rationalizable Implementation of Functions*, mimeo.

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

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

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

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

Leister, C.M., Y. Zenou and J. Zhung, *Social Connectedness and Contagion*, *mimeo*, 2020.

Leister, C.M., *Information Acquisition and Welfare in Network Games*, Games and Economic Behavior, 122, 453-475, 2020.

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

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

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

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

Müller, C., *Robust Implementation in Weakly Perfect Bayesian Strategies*, Journal of Economic Theory 189, 105038, 2020.

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

Myatt, D. P. and C. Wallace, *Information Acquisition and Use by Networked Players*, Journal of Economic Theory 182, 360-401, 2019.

Myerson, R.B., *Optimal Auction Design*, Mathematics of Operations Research, 6(1), 58–73, 1981.

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

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

Ollár, M. and A. Penta, *Incentive Compatibility and Belief Restrictions*, mimeo, 2021.

Oury, M. and O. Tercieux (2012), *Continuous Implementation*, Econometrica, Vol. 80, pp. 1605-1637 (2012).

Postlewaite A. and D. Schmeidler, *Implementation in Differential Information Economies*, Journal of Economic Theory, 39- 14-33.

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

Wilson, R., *Auctions of Shares*, The Quarterly Journal of Economics, 675-689, 1979.

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

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