Mechanism Design with Opportunity Constraints

Giacomo Rubbini*

Latest version here

Abstract

Traditional mechanism design assumes the planner has almost complete freedom in choosing an implementing mechanism. The paper generalizes the model by requiring the implementing mechanism to satisfy some exogenously imposed constraints on the lotteries in each agent's opportunity set. We show the revelation principle is still valid for a class of these restrictions, and we discuss applications of this framework to mechanism design in network environments.

Keywords: Mechanism Design, Revelation Principle, Networks

JEL Codes: C72, D78, D82

*Department of Economics, Brown University, giacomo_rubbini@brown.edu. I am indebted to Roberto Serrano for his guidance and support. I wish to thank Pietro Dall'Ara, Ben Golub, Santiago Hermo, Zeky Murra Anton, and Rajiv Vohra for their useful comments and suggestions.

1 Introduction

Can a government design a mechanism curbing the probability of infection in a population of agents, when the marginal cost of avoiding infection is private information? At first glance, this may seem like a regular mechanism design problem: as long as it is incentive compatible, it would be enough for the government to ask each agent to report their private information and implement the outcome corresponding to agents' reports.

This approach neglects a crucial detail of the problem: agents can affect each others' effort cost and probability of contagion only if they have social contact in the first place, restricting the set of mechanisms the planner can choose from. In particular, it imposes a restriction on the opportunity sets agents can face in the implementing mechanism. For example, a shop may be unable to affect the level of precautions and risk of infection of agents that are neither customers nor employees.[1]

The literature on mechanism design has not developed, thus far, a good language to describe problems of this kind. This paper develops a novel language to model these problems and characterizes the set of implementable social choice functions for any restrictions the planner may face. We will consider two main motivating examples: implementation over a given network (e.g., a social network or a priority ordering) and network design (e.g., implementation of an innovation-fostering network).

As the set of mechanisms the designer can choose from is now restricted, we do not know whether the revelation principle holds—that is, if it is enough to focus on direct revelation mechanisms only. The revelation principle may be violated, for instance, if the restrictions the planner faces force her to include some specific alternative in the opportunity sets of the players. This is the case in some exercises of network design, such as designing a social network app in which the option if the option to unfollow other users has to be always available.

[1]For simplicity, let us consider a static environment in which the social circles of each agent are taken to be exogenous.

The revelation principle is still valid when the opportunity restrictions are such that, for any allowed opportunity set, all subsets of that set are also allowed. As the opportunity sets of the direct mechanism will always be a subset of the opportunity sets that agents face in any indirect implementing mechanism, any implementable social choice function is implementable via the associated direct mechanism.[2] This is the case in two of the motivating examples mentioned above: mechanism design over a given network and one-sided matching problems.

If the revelation principle holds, the set of (partially) implementable social choice functions is equivalent to the set of functions that are BIC and *admissible*. Admissibility requires that the direct mechanism associated with a given social choice function satisfies the opportunity restrictions the planner faces.

The revelation principle holds in one of the two applications considered in this paper, implementation over an exogenously given network. In particular, focusing on the problem of one-sided matching with priorities, admissibility and the revelation principle entail that any implementable and ex-post efficient SCF must deliver the same allocation as the serial dictatorship algorithm. If we focus instead on the problem of implementing an SCF with one of the graphical games associated with the exogenously given network, we obtain Local Incentive Compatibility as a new necessary condition for implementation. Local Incentive Compatibility strengthens regular BIC by requiring truth-telling to be optimal even if opponents not connected to an agent do not report truthfully their type.

If the revelation principle does not hold, the class of implementable social choice functions coincides instead with the class of social choice functions that can be *extended* to a larger type space and whose extension is Bayesian Incentive Compatible, admissible, and delivers the same outcome as the social choice function of interest. This additional complexity is due to the fact some restrictions may require the planner to gather more information

[2]If we interpret agent's actions as messages, this kind of restriction can be thought of as setting an upper bound to how much private information agents can communicate with their choice of an action.

than the one she needs to implement the SCF—for instance, whenever the planner is always required to include some alternative in all agents' opportunity sets.

The revelation principle does not hold in the second application considered in this paper: network design exercises. For these exercises, we prove implementability is equivalent to Bayesian Incentive Compatibility and a novel condition called *accountability*. Accountability requires that, for all possible links that an agent may want to destroy, there exists an appropriate "punishment" the planner can enact to discourage a deviation in that sense. Even if the revelation principle does not hold, we propose a relatively simple indirect mechanism implementing any accountable SCF.

Other works have considered restrictions imposed on the implementing mechanism. The most recent work in this area is Gavan and Penta (2022) and their notion of *safe implementation*, which restricts the set of possible deviations from the equilibrium profiles. In our paper, restrictions apply instead to *every* profile of actions in the mechanism. This allows us to capture constraints that the mechanism has to satisfy regardless of the state, such as the ones imposed by a pre-existing network. This difference makes the results in this paper not directly comparable to Gavan and Penta (2022) except for the degenerate case in which their *admissibility function* does not depend on the state.

Hayashi and Lombardi (2019) also consider a planner who faces an institutional constraint in the choice of the implementing mechanism. In their work, the planner is able to design a mechanism only for one of the two sectors of society, while taking the mechanism of the other sector as given: for instance, the planner may design a mechanism to share the cost of a public good, while taking as given a market for private goods. Their framework is again unsuitable to capture some of the restrictions considered in this paper—in particular, restrictions on how large the opportunity sets of players can be.

2 Model

The goal of the social planner is to select an alternative from a set A, conditional on some information privately held from the agents in set I. As usual in the literature, we model the incomplete information problem by assuming there exists a finite set of types T_i for each agent $i \in I$ and that each agent knows her type but not the type of other players.[3] Let $T = \times_{i \in I} T_i$ be the set of all possible type profiles.

Let agents share a prior p over the type space, and let them update their beliefs according to Bayes' rule. We assume preferences over lotteries have expected utility form, with Bernoulli utility $u_i : A \times T \to \mathbb{R}$. Abusing notation slightly, let $u_i(a, t)$ for $a \in \Delta(A)$ denote the expected utility agent i derives from lottery a when the type profile is t.

The social planner seeks to implement a social choice function $f : T \to \Delta(A)$, and she does so by designing a mechanism (μ, S) where $S = \times_{i \in I} S_i$ is an action space and $\mu : S \to \Delta(A)$ is an outcome function. Once the planner has committed to a mechanism, agents choose a strategy $\sigma_i : T_i \to \Delta(S_i)$. We will denote the set of all such functions as Σ_i and a profile of strategies $\{\sigma_i\}_{i \in I}$ as $\sigma \in \Sigma$.

We will model the restrictions the planner faces through an *opportunity mapping* $\mathcal{O}$, which is a convex correspondence mapping each subset K of the set I of all players to a subset $\mathcal{O}_K$ of the set of lotteries over alternatives $\Delta(A)$. These restrictions should be understood as constraints the planner takes as given when choosing a mechanism to implement the SCF of interest. As the mechanism chosen by the planner is by definition state-independent, it is without loss of generality to assume that $\mathcal{O}$ is not responsive to changes in the state.

We then define the set of *admissible mechanisms* $\mathcal{G}(\mathcal{O})$ as:

$$\mathcal{G}(\mathcal{O}) = \{(\mu, S) : \mu(S_K, s_{-K}) \in \mathcal{O}_K, \text{ for all } \delta_{-K} \in \Delta(S_{-K})\}$$

[3] The finiteness assumption is not necessary for the arguments, but it simplifies exposition.

That is, a mechanism is allowed if all opportunity sets of coalition $K \subseteq I$ belong to $\mathcal{O}_K$.

Could these constraints be more easily captured by removing some elements from the set of possible alternatives A the planner has available, or by imposing suitable restrictions on agents' utility functions? The focus of this paper is on those setups in which only a *subset* of agents is allowed to induce some outcomes: indeed, removing an alternative a from set A is equivalent to imposing the restriction that implies no player can induce a.

Consider again our contagion example. The restriction the planner faces in that case is that agents that are not connected should not be able to affect each other's outcomes: however, this constraint cannot be captured in the classical mechanism design model, which does not put any bounds on the agents' power to induce a given alternative.[4] Moreover, eliminating some alternatives from set A would impose a much stronger restriction on the planner, as it would entail that *no* agent is able to induce such alternative.

A similar argument applies to suitably restricting agents' utility functions as well.

Finally, we say that a SCF f is *(partially) $\mathcal{O}$-implementable* whenever there exists an admissible mechanism that partially implements f in BNE, i.e. there exists $(\mu, S) \in \mathcal{G}(\mathcal{O})$ and a BNE σ of (μ, S) such that $f = \mu(\sigma)$. It is straightforward to check this definition coincides with implementability in BNE whenever the restrictions have no bite, i.e. whenever $\mathcal{O}_K = \Delta(A)$ for all $K \subseteq I$. For the remainder of the paper, let me say f is $\mathcal{O}$-implementable whenever it is partially $\mathcal{O}$-implementable.

As $\mathcal{O}$-implementability is a stricter requirement than implementability, all necessary conditions for BNE implementation will still be necessary for $\mathcal{O}$-implementation.

In particular, Bayesian Incentive Compatibility (BIC) will still be a necessary condition. We say a SCF f is *Bayesian Incentive Compatible* (BIC) whenever for all $i \in I$, $t_i, t_i' \in T_i$:

$$\int_{T_{-i}} u_i(f(t), t) \, \mathrm{d}p(t_{-i}|t_i) \geq \int_{T_{-i}} u_i(f(t_i', t_{-i}), t) \, \mathrm{d}p(t_{-i}|t_i)$$

[4]The canonical mechanism of Maskin (1999) and Jackson (1991) are cases in point, as each agent can induce any outcome she desires by announcing an integer higher than the one reported by their opponents.

Finally, we say a SCF satisfies *admissibility* whenever $f(\Delta(T_K), \delta_{-K}) \in \mathcal{O}_K$ for all $K \subseteq I$ and $\delta_{-K} \in \Delta(T_{-K})$. In other words, admissibility ensures the direct mechanism associated with a given SCF belongs to $\mathcal{G}(\mathcal{O})$.

3 Motivating Examples

In this section, we discuss how the model above can be applied to capture restrictions arising in mechanism design over a fixed network and in network design. In each of these cases, the planner is constrained to choose from a particular class of games: graphical games, games in which lower-priority agents cannot affect the outcome of high-priority ones, and network formation games.

3.1 Implementation over an Exogenous Network

The problem described in the introduction to this paper is an exercise of mechanism design over an exogenously given network: agents are embedded in a social network, and agents who are not linked cannot affect each other's outcome. Other examples from the network literature include management hierarchies and load-balancing negotiations in computer networks (Kearns et al., 2001), provision of local public goods, local market power, innovation spillovers, peer effects, and organization hierarchies. The same framework can capture also problems of mechanism design with priorities.

The following two sections present two different ways of formalizing this intuition.

3.1.1 Network Games

The first way to capture the restrictions imposed by a pre-existing network is to assume agents can affect the outcome of agents they are linked to in the network, i.e. their neighbors.[5]

[5]More permissive definitions can be captured within the same framework—for instance, allowing agents to affect others' outcomes only to a limited degree.

Consider a network $\mathcal{N} = (g, I)$, where g is a collection of (directed) links and I is a set of nodes. In this case, nodes $i \in I$ represent players of the game. Each link $g_{ij} \in g$ captures the fact player i is allowed to affect the outcome of player j in the mechanism. Let us adopt the convention that each agent is linked to herself, i.e. that $g_{ii} \in g$ for all $i \in I$.

Let A have a product structure, i.e. $A = \times_{i \in I} A_i$. We can then define the set of *network games* associated with a network $\mathcal{N}$ as the set of games induced by any mechanism (μ, S) such that $(\mu(s))_i = (\mu(s'_j, s_{-j}))_i$ for all i, j with $g_{ji} \notin g$, $s'_j \in S_J$ and $s \in S$.

This definition formalizes the intuition that i is able to change j's outcome by varying her action only if there is a (directed) link between the two. It is possible to prove the set of all network games associated with network $\mathcal{N}$ can be characterized as the set of mechanisms that are admissible according to a suitably chosen restriction mapping $\mathcal{O}$ capturing the intuition that agents can affect the (distribution over) outcomes of their neighbors only. To this purpose, it is enough to consider any mapping $\mathcal{O}$ such that $O_K \in \mathcal{O}_K^N$ whenever all lotteries in O_K imply the same marginal probability distribution over $A_{-N(K)}$, where $N(K)$ is the set of agents connected to at least one $i \in K$.[6]

This paper's framework can be immediately applied to implementation via network games. In particular, if the network is completely connected (i.e., there is a link between any two players: $g_{ij} \in g$ for every i, j), $\mathcal{O}^N$ becomes trivial and then network implementation boils down to Bayesian Nash implementation. The network structure, however, generally constitutes a non-trivial constraint: Section 5.3 shows that the network structure yields a local incentive compatibility condition that is more restrictive than usual BIC. In particular, $\mathcal{O}^N$-admissibility requires the SCF f to be such that each agent's outcome is not affected by agents that do not belong to her neighborhood.

An example of a problem in which the network structure plays a role comes from the constraints imposed by a pre-existing priority system. Let agents be ranked according to a particular criterion, such as test scores in college choice or seniority in dormitory allocation

[6]Formally, $\mathcal{O}_K$ is the set of all $O_K \subseteq \Delta(A)$ such that $\pi, \pi' \in O_K$ entails that π and π' have the same marginal distribution over $A_{-N(K)}$.

problems. If the planner operates under such a constraint, the implementing mechanism will have to be such that agents with lower priority can not affect the matching or allocation of those ranked above them.

Assume agents with lower indexes have a higher priority order than agents with larger indexes and that the set of alternatives consists of all possible allocations of objects in a set $A_i = X$ (without repetition).

We can then capture the fact that agents with lower priority are not allowed to affect the object agents with higher priority receive by assuming $g_{ij} \in g$ if and only if $i \geq j$ by assuming that for all $K \subseteq I$.[7] In this case, $\mathcal{O}$ will be such that $\mathcal{O}_K$ contains any subset O_K of $\Delta(A)$ such that all lotteries in O_K have the same distribution over outcomes for agents with priority higher than the agent with the highest priority in K.[8]

Admissibility then entails the information agent i reports should not affect the allocation received by agents with higher priority: the planner cannot use the information provided by i to determine the object received by any $j < i$. In particular, only self-reported information will matter for the first agent in the priority ranking: if f is BIC, it must be strategy-proof for this agent. A similar argument applies for remaining agents: if f is BIC, truthful reporting must be optimal no matter what agents with lower priority report.

3.1.2 Graphical Games

An alternative way to model the restrictions imposed from a pre-existing network is to assume an agent can affect an opponent's payoff only if she is linked to that opponent.

Consider again a network $\mathcal{N} = (g, I)$, without imposing any restriction on the structure of A. We define the set *graphical game* associated with network $\mathcal{N}$ as the set of games induced by any mechanism (μ, S) such that $u_i(\mu(s), t) = u_i(\mu(s'_j, s_{-j}), t)$ for all j with

[7] While we focus on the simplest case in which agents are totally ordered, more complex priority systems can be captured by appropriately defining the network associated to it.

[8] Formally, $\pi, \pi' \in O_K$ implies π and π' have the same marginal distribution over outcomes for all agents i with $i < \min_{j \in K} j$.

$g_{ji} \notin g$, $s'_j \in S_j$, $s \in S$ and $t \in T$.[9] Differently from the network games described above, graphical games allow an agent to directly affect an opponent's payoff only if she is linked to that opponent.

It is again possible to prove the set of all graphical games associated with the pair $(\mathcal{N}, T)$ can be characterized as the set of mechanisms that are admissible according to a suitably chosen restriction mapping. To this purpose, it is enough to consider $\mathcal{O}^G$ to be such that $O_K \in \mathcal{O}^G_K$ whenever for all $a, b \in O_K$ we have $u_i(a,t) = u_i(b,t)$ for all $t \in T$ and i with $g_{ji} \notin g$ for all $j \in K$.

Similarly to network games, restricting attention to graphical games generally imposes a non-trivial constraint on the planner whenever the network is not complete. As for network games, admissibility requires a local incentive compatibility condition to hold (Section 5.3). In addition, if the domain of preferences is rich enough and the network is not completely connected, the only implementable SCF are the ones in which a subset of players acts as a "dictator".

3.2 Network Design

The social planner may be tasked with the design of large infrastructures robust to failures and perturbations (such as power grids and computer networks) or a network that slows down contagion across nodes (e.g. to prevent epidemics) or accelerates it (e.g. to foster innovation).[10]

We can see this network design exercise as a mechanism design problem in which the planner has to pick a mechanism inducing the right network in each state. Denote the set of all possible graphs as $G = \times_{i \in I} G_i$, where G_i is the set of all (directed) links g_{ij} that agent i could form with her opponents. We will assume A is the space of all pairs $(g, x) \in (G, X)$, where x is a vector of other outcomes (for example, allocations or transfers).

[9] This definition of *graphical game* is used, among the others, from Kearns et al. (2001) in Computer Science.

[10] See Jackson et al. (2017) for a longer discussion.

The first kind of restriction the planner may face in these environments is that an agent is typically not allowed to affect the links between other agents that do not involve herself—that is, her actions can affect the network only locally.[11] This is the case, for instance, in the network formation models of Jackson and Wolinsky (1996) and Bala and Goyal (2000): both works assume the set of strategies available to each agent consists of all linking decisions with other players in the network formation game. We can capture this feature by considering an opportunity mapping $\mathcal{O}^D$ such that, for all $K \subseteq I$, $\mathcal{O}^D_K$ contains all $O_K \subseteq \Delta(G \times X)$ such that all $\pi \in O_K$ imply the same marginal distribution over $(G_i)_{i \notin K}$.

In this setup, admissibility entails an agent's report can only affect the target network locally, as all networks in i's opportunity set will differ only in i's neighborhood.[12] We can also notice this restriction function satisfies the conditions of Theorem 2, and we can use the revelation principle to argue that any BIC function such that agents' information has only local effects on the network is implementable.

In some network design environments, the planner may also be unable to force an agent to link or not to link with another agent. For instance, a dating app cannot force two users to date, and a regulation authority cannot compel two firms to innovate by sharing their know-how.[13]

We can capture this intuition by defining $\mathcal{O}^{D'}$ as follows. For each graph g, let $g_{iK} \subseteq g$ denote the set of links that i forms with agents in K. For all $K \subseteq I$, let then $\mathcal{O}^{D'}_K$ contain all sets O_K such that for all $\pi \in \mathcal{O}_K$ and collections $(L_i)_{i \in I}$ with $L_i \subseteq I/\{i\}$ there exists $\pi' \in \mathcal{O}_K$ that assigns $g - \cup_{i \in K} g_{iL_i}$. In other words, any $O_K \in \mathcal{O}_K$ is such that agent $i \in K$ could choose to sever the links connecting her with any subset of agents by changing the action she chooses in the mechanism.

[11] In other applications, such as school choice problems, we may relax this assumption and allow agents to affect the links of agents they are directly or indirectly connected to. While we do not consider these more complex restrictions in this section, they can be accommodated similarly to the ones presented here.

[12] While i's action affects the network only locally, it could still indirectly affect the incentives to form a link of non-neighboring nodes.

[13] This does not entail the decision not to comply does not have other consequences: for example, a government may offer incentives (captured by x) to firms that come together to innovate.

Finally, notice the restrictions captured by $\mathcal{O}^D$ and $\mathcal{O}^{D'}$ are not mutually exclusive: in that case, it will be enough to take restriction mapping $\mathcal{O}^{D*}$ to be given by $\mathcal{O}^{D*}_K = \mathcal{O}^D_K \cap \mathcal{O}^{D'}_K$ for all $K \subseteq I$.

4 Results

In this section, we show the set of implementable social choice function coincides with the set of social choice functions that admit an extension that is both BIC and admissible (Theorem 1). In some interesting cases, this extension coincides with the social choice function of interest (Theorem 2).

Due to restrictions on the opportunity sets the planner can present to agents, the revelation principle does not generally hold. This happens, in particular, when the planner has to include in agents' opportunity sets lotteries that are not part of the agents' opportunity sets in the direct mechanism. For instance, the designer of a dating app has to offer users the possibility to decline a specific date—even if the designer would want that date to happen. From an informational standpoint, this entails that the planner has to ask agents for more information than the one she needs to implement the SCF of interest.[14]

We can capture this intuition by extending the direct mechanism accordingly. Denoting the pair (T, p) as a type space, let $\chi : \hat{T} \to T$ be an onto mapping, $f : T \to \Delta(A)$, and $\hat{f} : \hat{T} \to \Delta(A)$. We then say triple $(\hat{T}, \hat{p}, \chi)$ *extends* type space (T, p) whenever for all $T' \subseteq T$:

$$p(T) = \int_{\hat{t} \in \hat{T}} \mathbf{1}_{\{\hat{t} \in \chi^{-1}(T)\}} \, \mathrm{d}\hat{p}(\hat{t})$$

That is, $(\hat{T}, \hat{p}, \chi)$ extends type space (T, p) whenever the probability of each state in T is the same as the sum of the probabilities of all states $\hat{t} \in \hat{T}$ that χ associates to t. We can then interpret χ as a function that associates each piece of information an agent could reveal to the planner in this "extended direct mechanism" with a standard type.

[14] Here we interpret "information" in terms of revealed preference data: the planner elicits agents' preferences over a set of lotteries that is larger than their opportunity set in the direct mechanism.

For a given extended type space $(\hat{T}, \hat{p}, \chi)$, we say $\hat{f}$ is an *extension* of f whenever for all $t \in T$:

$$f(t) = \int_{\hat{t} \in \chi^{-1}(t)} \hat{f}(\hat{t}) \, \mathrm{d}\hat{p}(\hat{t})$$

That is, $\hat{f}$ extends f with respect to a given expanded type space $(\hat{T}, \hat{p}, \chi)$ whenever the distribution over alternatives f prescribes in state t is the same as the one expected as the sum over all states $\hat{t}$ that χ associates to t.

Theorem 1. *A SCF f is $\mathcal{O}$-implementable if and only if there exists an extended type space $(\hat{T}, \hat{p})$ of (T, p), a $\chi : \hat{T} \to T$ and an extension $\hat{f} : \hat{T} \to \Delta(A)$ of f that is BIC and admissible.*

Theorem 2 below highlights they are instead enough in the special case of opportunity restrictions only imposing an upper bound to how large the opportunity sets are.

Theorem 2. *Let $\mathcal{O}$ be such that $O_K \in \mathcal{O}_K$ implies $O'_K \in \mathcal{O}_K$ for all $O'_K \subseteq O_K$ and $K \subseteq I$. Then f is partially implementable if and only if it is BIC and admissible. Moreover, if f is partially implementable, it is partially implementable via a direct mechanism.*

The revelation principle reduces any (potentially complex) implementing mechanism to a direct mechanism in which agents have only to report their own type. As long as opportunity restrictions do not impose a constraint on how small opportunity sets can be, any implementable social choice functions will be implementable via the associated direct mechanism.

5 Applications

5.1 Implementation over an Exogenous Network

In network games, only the neighborhood of each agent is able to affect her outcome or payoffs. As both $\mathcal{O}^N$ and $\mathcal{O}^G$ satisfy the condition of Theorem 2, the revelation principle holds.

In either case, admissibility imposes non-trivial restrictions on the class of implementable SCF whenever the underlying network is not complete.

5.2 Network Games

For the remainder of this section, let me assume agents' utility depends only on the object they receive and their type to keep our results comparable with standard models in the matching literature. Abusing notation, let us denote agents' Bernoulli utilities as $u_i(x, t_i)$, where x is the object the agent is allocated. Let us also assume that preference reversals are possible for each agent. That is, assume there exists no $i \in I$ and no pair of objects x, y such that $u_i(x, t_i) > u_i(y, t_i)$ for all $t_i \in T_i$.

We will prove that, in this environment, ex-post Pareto efficiency is possible only if the SCF allocates objects the same way a serial dictatorship algorithm would do. This follows from the fact that agents with a priority ordering lower than i cannot affect the object she is allocated: if x is not i's preferred object among those left over by agents with higher priority in state t, then x must be assigned to i even if type t'_j of $j > i$ would prefer x to the object y she is assigned, generating an inefficiency.

Formally, we say f is *(ex-post) efficient* whenever for all $t \in T$, there exists no $a \in A$ such that $u_i(a, t_i) \geq u_i(f(t), t_i)$ for all $i \in I$, with strict inequality for at least one i. We will moreover say f is a *serial dictatorship* whenever it assigns to player 1 her top choice, to player 2 her top choice among the remaining objects, and so on and so forth.

We can then derive the following result.

Theorem 3. *An SCF f is efficient and $\mathcal{O}^N$-admissible only if f is a serial dictatorship.*

Notice $\mathcal{O}^N$ satisfies the conditions of Theorem 2, entailing $\mathcal{O}^N$-admissibility is necessary for implementation. Together with Theorem 3, this implies any implementable SCF f will yield the same outcome as the serial dictatorship algorithm in this kind of setup.

5.3 Graphical Games

The fact the revelation principle holds immediately delivers a restriction on f: as any $\mathcal{O}^G$-implementable f is also admissible, $u_i(f(t), t') = u_i(f(\tilde{t}_j, t_{-j}), t')$ for all $t, t' \in T$, $\tilde{t}_j, t_j \in T_j$ and j not connected to i.

This translates into a necessary condition stronger than BIC: truth-telling will have to be optimal for each agent as long as her neighborhood is reporting truthfully, regardless of the reports of agents outside her neighborhood.[15] Formally, we say a SCF f is Locally Incentive Compatible (LIC) whenever for all $i \in I$, $t_i, t_i' \in T_i$ and deception $\alpha_{-N(i)} : T_{-N(i)} \to T_{-N(i)}$:

$$\int_{T_{-i}} u_i(f(t_i, t_{N(i)}, \alpha_{-N(i)}(t_{-i}), t) \, \mathrm{d}p(t_{-i}|t_i) \geq \int_{T_{-i}} u_i(f(t_i', t_{N(i)}, \alpha_{-N(i)}(t_{-i}), t) \, \mathrm{d}p(t_{-i}|t_i)$$

In this sense, we can interpret LIC as a weaker form of strategy-proofness, requiring i's report to be optimal no matter what agents outside her neighborhood report. We then derive the following result.

Corollary 1. *If f is $\mathcal{O}^G$-admissible, then f is BIC if and only if it is LIC.*

While LIC implies BIC, it is not enough to imply $\mathcal{O}^G$-admissibility of f, as the latter requires reports of agents outside of i's neighborhood to not affect i's outcome in any way. Instead, LIC does not exclude this possibility: reports of agents outside of i's neighborhood may still affect her outcome as long as truth-telling remains optimal.

Notice that any LIC f is BIC even if agents do not expect opponents outside their neighborhood to report truthfully. Therefore, f would be implementable even if agents pay attention or correctly anticipate only the strategies of their neighbors. Similarly, LIC entails each agent has an incentive to truthfully report her type even if the agents that are not linked to her are *faulty* (in the sense of Eliaz (2002)).

[15] Notice a similar result obtains with network games as well, as long as agents only care about their own allocation. As that allocation can be affected only by their neighbors, the same is true for their payoff.

If the domain of preferences is rich enough with respect to the range of f, $\mathcal{O}^G$-admissibility becomes even more restrictive. Let us say T is *rich* with respect to SCF f if for all $i \in I$ and pair of alternatives $a, b \in f(T)$ there exists at least one state $t \in T$ such that i is not indifferent between a and b.[16]

If richness holds, any agent who is not connected to all other agents will not be able to affect the outcome of the mechanism—in other words, f will not respond to the report of this agent. This is the case for most agents in a variety of network shapes (lines, circles, stars, trees) when there are at least three agents.

Formally, we will say f is *unresponsive* to agent i whenever $f(t_i, t_{-i}) = f(t'_i, t_{-i})$ for all $t_{-i} \in T_{-i}$ and $t_i, t'_i \in T_i$. The following result then follows as a corollary to Theorem 1.

Corollary 2. *If T is rich with respect to f and there exist $j \in I$ such that $g_{ij} \notin g$ for some $i \in I$, any $\mathcal{O}^N$-admissible f is unresponsive to i.*

This result follows from the fact agents cannot affect the payoff of the opponents they are not linked to in any state (by admissibility), and richness entails they will not be able to affect the outcome at all. Formally, consider any $j \neq i$ with $g_{ij} \notin g$ and $f(t) \in O_i$. By admissibility, $u_j(f(t), t') = u_j(f(t''_i, t_{-i}), t')$ for all $t' \in T$, $t''_i \in T_i$. As T is rich with respect to f, it follows $f(t_i, t_{-i}) = f(t'_i, t_{-i})$ for all $t_i, t'_i \in T_i$ and $t_{-i} \in T_{-i}$, i.e. f is unresponsive to i. Intuition for this result is quite straightforward: agents cannot affect the payoff of the opponents they are not linked to in any state by admissibility, and richness entails they will not be able to affect the outcome at all.

This has important implications. For example, any SCF that is implementable over a star network is *dictatorial* when T is rich with respect to f. Formally, say f is *dictatorial* whenever there exists $i \in I$ such that $u_i(f(t), t) \geq u_i(f(t'), t)$ for all $t, t' \in T$.[17] This result is easy to derive if we consider that any two peripheral players in the network are

[16] This condition is reminiscent of the universal domain assumption in the impossibility theorems of Arrow and Gibbard and Satterthwaite.

[17] This is a slightly weaker definition of dictatorship, that coincides with the classical one whenever f is onto.

not connected: due to richness, those players are therefore unable to affect the outcome. A similar (weaker) argument holds for other core-periphery networks: peripheral players will be unable to affect the outcome, which would be determined in the connected core[18]. Generalizing, for these graphs Corollary 2 implies information revelation to the planner can only affect other agents' incentives *locally*: i can therefore reveal payoff-relevant information only on her neighbors.

5.4 Network Design

In network design environments, whether the revelation principle holds or not will depend on what restrictions the planner is facing. For the remainder of this section, we will restrict attention to any SCF $f = (\gamma, \xi)$, where $\gamma : T \to G$ and $\xi : T \to \Delta(X)$.[19]

If we assume the restrictions the planner faces are captured by mapping $\mathcal{O}^D$, the revelation principle holds by an argument similar to the one of previous sections and a SCF f is implementable if and only if it is BIC and admissible. Moreover, an agent's report has only local effects on the network arising from the strategic interaction whenever A is the set of all possible graphs over the nodes in I. That is, the planner cannot use the information contained in i's report to decide whether agents $j, k \neq i$ should be linked.

As $\mathcal{O}^{D'}$ does not satisfy the conditions laid down in Theorem 2, not all $\mathcal{O}^{D'}$-implementable and $\mathcal{O}^{D*}$-implementable SCF are implementable via a direct mechanism. For instance, a constant SCF that prescribes all agents are connected for every state $t \in T$ would not be implementable via the associated direct mechanism as there has to exist at least one action for each agent allowing her to destroy each of her links.

Define $\gamma^{-iK} : T \to G$ as $\gamma^{-iK}(t) = \gamma(t) - g_{-iK}$ for all $t \in T$, $i \in I$ and $K \subseteq I$. We say a SCF f is *accountable* whenever for all $i \in I$ and non-empty $K \subseteq I$ there exists

[18] In this sense, we could argue the core is the "dictator" in this case.

[19] Assuming the SCF assigns a deterministic graph to each state in T does not affect our results, but it considerably simplifies notation.

$\xi_i^K : T \to \Delta(X)$ such that for all $t_i, t_i' \in T_i$:

$$\int_{T_{-i}} u_i(\gamma(t), \xi(t)), t) \, \mathrm{d}p(t_{-i}|t_i) \geq \int_{T_{-i}} u_i(\gamma^{-iK}(t_i', t_{-i}), \xi_i^K(t_i', t_{-i})), t) \, \mathrm{d}p(t_{-i}|t_i)$$

For each $K \subseteq I$, we can interpret ξ_i^K as "punishing" agent i of type t_i for severing links with agents in K.

Theorem 4. *A SCF f is $\mathcal{O}^{D'}$-implementable if and only if it is BIC and accountable. Moreover, if f is $\mathcal{O}^{D}$-admissible, then it is $\mathcal{O}^{D^*}$-admissible.*

To implement any SCF satisfying accountability and BIC, it is enough to design a simple indirect mechanism asking agents to report their type and a set of agents they do not want to be linked to. If all agents report a type and do not ask to destroy any link, the outcome of the mechanism is the same as the SCF would prescribe for that type profile. Otherwise, the graph resulting from the mechanism is the same as the SCF would prescribe minus the links agents asked not to form and agents will be "punished" with the function x_i^K derived from the accountability condition above.

Notice that, whenever if A consists of the set of all graphs over I (that is, $X = \emptyset$), accountability holds only if the SCF is such that the expected value of each set of links is be non-negative.

References

Bala, V. and Goyal, S. (2000). A noncooperative model of network formation. *Econometrica*, 68(5):1181–1229.

Eliaz, K. (2002). Fault tolerant implementation. *The Review of Economic Studies*, 69(3):589–610.

Gavan, M. J. and Penta, A. (2022). Safe implementation. Bse working paper 1363.

Hayashi, T. and Lombardi, M. (2019). Constrained implementation. *Journal of Economic Theory*, 183:546–567.

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

Jackson, M. O., Rogers, B. W., and Zenou, Y. (2017). The economic consequences of social-network structure. *Journal of Economic Literature*, 55(1):49–95.

Jackson, M. O. and Wolinsky, A. (1996). A strategic model of social and economic networks. *Journal of Economic Theory*, 71(1):44–74.

Kearns, M. J., Littman, M. L., and Singh, S. P. (2001). Graphical models for game theory. In *Proceedings of the 17th Conference in Uncertainty in Artificial Intelligence*, UAI '01, page 253–260, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Maskin, E. (1999). Nash equilibrium and welfare optimality. *Review of Economic Studies*, 66:23–38.

Appendix A Proofs

Proof of Theorem 1. Suppose f is $\mathcal{O}$-implementable. Then there exists a mechanism $(\mu, S) \in \mathcal{G}(\mathcal{O})$ with an equilibrium σ such that $\mu(\sigma) = f$. Construct then an extended type space as follows. Let $\hat{T}_i = T_i \times \Delta(S_i)$ and $\chi : \hat{T}_i \to T_i$ be such that $\chi(t_i, \delta_i) = t_i$ for all $\delta_i \in \Delta(S_i)$. Clearly, χ is onto as for all $t_i \in T_i$ there exists $\hat{t}_i = (t_i, \delta_i)$ for $s_i \in \Delta(S_i)$ such that $\chi_i(\hat{t}_i) = t_i$. Define then $\hat{p}$ be such that for all $i \in I$, $t_i \in T_i$, $s_i \in S_i$, $T_i' \subseteq T_i$ and $S_i' \subseteq S_i$:

$$\hat{p}((T_{-i}', S_{-i}')|(t_i, s_i)) = \int_{T_{-i}'} \sigma_{-i}(t_{-i})[S_{-i}'] \, \mathrm{d}p_i(t_{-i}|t_i)$$

As $\sigma_{-i}(t_{-i})[S_{-i}] = 1$, it follows that $\hat{p}((T_{-i}', S_{-i})|(t_i, s_i)) = p(T_{-i}'|t_i)$ for all $t_i \in T_i$ and $T_{-i}' \subseteq T_{-i}$. Therefore for all $T' \subseteq T$:

$$\int_{\hat{t} \in \hat{T}} \mathbf{1}_{\{\hat{t} \in \chi^{-1}(T')\}} \, \mathrm{d}\hat{p}(\hat{t}) = \hat{p}((T',)) = p(T')$$

Set now $\hat{f}(\hat{t}) = \hat{f}((t, \delta)) = \mu(\delta)$ for all $\hat{t} \in \hat{T}$. We prove $\hat{f}$ extends f. By implementability we have that for all $t \in T$:

$$f(t) = \mu(\sigma(t)) = \int_{s \in S} \mu(s) \, \mathrm{d}\sigma(t)[s] = \int_{s \in S} \hat{f}((t, s)) \, \mathrm{d}\hat{p}((t, s)) = \int_{\hat{t} \in \chi^{-1}(t)} \hat{f}(\hat{t}) \, \mathrm{d}\hat{p}(\hat{t})$$

Let us now prove $\hat{f}$ is BIC. Notice first that for all $\hat{t}_i' \in \hat{T}_i$ there exists $\delta_i' \in \Delta(S_i)$ such that $\hat{t}_i' = (t_i', \delta_i')$. The equilibrium condition states that for all $i \in I$, $t_i \in T_i$ and $\delta_i' \in \Delta(S_i)$:

$$\int_{T_{-i}} u_i(\mu(\sigma(t)), t) \, \mathrm{d}p_i(t_{-i}|t_i) \geq \int_{T_{-i}} u_i(\mu(\delta_i', \sigma_{-i}(t_{-i})), t) \, \mathrm{d}p_i(t_{-i}|t_i)$$

Rewriting this inequality using linearity of expected utility:

$$\int_{T_{-i}} \int_{S_{-i}} u_i(\mu(s), t) \, \mathrm{d}\sigma_{-i}(t_{-i}) \, \mathrm{d}p_i(t_{-i}|t_i) \geq \int_{T_{-i}} \int_{S_{-i}} u_i(\mu(s_i', s_{-i}), t) \, \mathrm{d}\sigma_{-i}(t_{-i}) \, \mathrm{d}p_i(t_{-i}|t_i)$$

The inequality above directly entails:

$$\int_{\hat{T}_{-i}} u_i(\hat{f}(\hat{t}), \chi(\hat{t}))\ \mathrm{d}\hat{p}(\hat{t}_{-i}|\hat{t}_i) \geq \int_{\hat{T}_{-i}} u_i(\hat{f}(\hat{t}'_i, \hat{t}_{-i}), \chi(\hat{t}))\ \mathrm{d}\hat{p}(\hat{t}_{-i}|\hat{t}_i)$$

Finally, we prove $\hat{f}$ is admissible. Notice that, by construction, for all $K \subseteq I$ and $\tau_{-K} \in \Delta(\hat{T}_{-K})$ there exists $\delta_{-K} \in \Delta(S_{-K})$ such that:

$$\hat{f}(\Delta(\hat{T}_K), \tau_{-K}) = \mu(\Delta(S_K), \delta_{-K}) \in \mathcal{O}_K$$

As our choice of τ_{-K} was arbitrary and (μ, S) is an admissible mechanism, this concludes the necessity part of the proof.

As for the sufficiency part of the proof, set $(\mu, S) = (\hat{f}, \hat{T})$. As $\hat{f}(\Delta(\hat{T}_K), \tau_{-K}) \in \mathcal{O}_K$ for all $K \subseteq I$ and $\tau_{-K} \in \Delta(\hat{T}_{-K})$, $(\mu, S) \in \mathcal{G}(\mathcal{O})$. As $\hat{f}$ is BIC and preferences are determined only by t, for all $i \in I$, $\hat{t}_i, \hat{t}'_i \in \hat{T}_i$:

$$\int_{\hat{T}_{-i}} u_i(\hat{f}(\hat{t}), \chi(\hat{t}))\ \mathrm{d}\hat{p}(\hat{t}_{-i}|\hat{t}_i) \geq \int_{\hat{T}_{-i}} u_i(\hat{f}(\hat{t}'_i, \hat{t}_{-i}), \chi(\hat{t}))\ \mathrm{d}\hat{p}(\hat{t}_{-i}|\hat{t}_i)$$

Moreover, for all $i \in I$ and $\hat{t}_i, \hat{t}'_i \in \chi_i(t_i)$:

$$\int_{\hat{T}_{-i}} u_i(\hat{f}(\hat{t}), \chi(\hat{t}))\ \mathrm{d}\hat{p}(\hat{t}_{-i}|\hat{t}_i) = \int_{\hat{T}_{-i}} u_i(\hat{f}(\hat{t}), \chi(\hat{t}'_i, \hat{t}_{-i}))\ \mathrm{d}\hat{p}(\hat{t}_{-i}|\hat{t}_i)$$

Consider $(\mu, S) = (\hat{f}, \hat{T})$ and σ such that, for each $t \in T$ and $i \in I$, $\sigma_i(t_i)[\hat{T}'_i] = \hat{p}(\hat{T}'_i|\chi_i(\hat{t}_i) = t_i)$ for all $\hat{T}'_i \subseteq \hat{T}_i$. By our definition of σ, the inequality and equality above together imply:

$$\int_{T_{-i}} u_i(\hat{f}(\sigma(t)), t)\ \mathrm{d}p(t_{-i}|t_i) \geq \int_{T_{-i}} u_i(\hat{f}(\hat{t}'_i, \sigma_{-i}(t_{-i})), t)\ \mathrm{d}p(t_{-i}|t_i)$$

Therefore σ is an equilibrium of $(\hat{f}, \hat{T})$. Moreover, for all $t \in T$:

$$\int_{T_{-i}} \hat{f} \circ \sigma(t)\ \mathrm{d}p(t) = \int_{\hat{t} \in \chi^{-1}(t)} \hat{f}(\hat{t})\ \mathrm{d}\hat{p}(\hat{t}) = f(t)$$

Where the last equality follows from the fact $\hat{f}$ extends f. Then $(\hat{f}, \hat{T})$ has an equilibrium σ such that $\hat{f}(\sigma) = f$. This concludes the proof. □

Proof of Theorem 2. For the if part, by Theorem 1, we have just to find an extended type space $(\hat{T}, \chi, \hat{p})$ and a SCF $\hat{f}$ that is BIC and that extends f. Let $\hat{T} = T$, χ be the identity function and $\hat{p} = p$. It is then immediate to notice $\hat{f} = f$ extends f, and that it is admissible and BIC as f is admissible and BIC. From the sufficiency part of Theorem 1, it then follows f is implementable via the associated direct mechanism (f, T).

As for the converse, suppose f is implementable. As $\mathcal{O}$-implementability implies standard implementability, f is BIC. Moreover, as f is implementable, there exists a mechanism $(\mu, S) \in \mathcal{G}(\mathcal{O})$ and an equilibrium σ such that $f = \mu(\sigma)$. As $(\mu, S) \in \mathcal{G}(\mathcal{O})$, it follows that $\mu(\Delta(S_K), \delta_{-K}) \in \mathcal{O}_K$ for all $\delta_{-K} \in S_{-K}$ and $K \subseteq I$. Notice that for all $\tau_{-K} \in \Delta(T_{-K})$:

$$f(\Delta(T_K), \tau_{-K}) = \mu(\Delta(\sigma_K(T_K)), \sigma_{-K}(t_{-K})) \subseteq \mu(\Delta(S_K), \sigma_{-K}(t_{-K})) \in \mathcal{O}_K$$

As $O_K \in \mathcal{O}_K$ whenever $O_K \subseteq O'_K$ for $O'_K \in \mathcal{O}_K$, it follows $f(\Delta(T_K), \tau_{-K}) \in \mathcal{O}_K$ for all $K \subseteq I$ and $\tau_{-K} \in \Delta(T_{-K})$. This entails f is admissible.

Moreover, as implementability of f entails it is BIC and admissible, the first part of the argument implies it is implementable via the associated direct mechanism (f, T). This establishes the revelation principle holds, concluding the proof. □

Proof of Corollary 1. It is immediate to see LIC implies BIC by considering $\alpha_{-N(i)}$ as the identity function. For the other side of the argument, recall that BIC entails that for all $i \in I$ and $t_i, t'_i \in T_i$ we have:

$$\int_{T_{-i}} u_i(f(t), t) \, \mathrm{d}p(t_{-i}|t_i) \geq \int_{T_{-i}} u_i(f(t'_i, t_{-i}), t) \, \mathrm{d}p(t_{-i}|t_i)$$

As f is $\mathcal{O}^N$-admissible, i is indifferent between any of the reports of opponents she is not

connected to. This implies that, for all $i \in I$, $t'_i \in T_i$ and $\alpha_{-N(i)} : T_{-N(i)} \to T_{-N(i)}$:

$$u_i(f(t'_i, t_{-i}), t) = u_i(f(t'_i, t_N(i), \alpha_{-N(i)}(t_{-i}), t)$$

Therefore, for all $i \in I$, $t_i, t_i \in T_i$ and $\alpha_{-N(i)} : T_{-N(i)} \to T_{-N(i)}$:

$$\int_{T_{-i}} u_i(f(t_i, t_{N(i)}, \alpha_{-N(i)}(t_{-i}), t)\ \mathrm{d}p(t_{-i}|t_i) \geq \int_{T_{-i}} u_i(f(t'_i, t_{N(i)}, \alpha_{-N(i)}(t_{-i}), t)\ \mathrm{d}p(t_{-i}|t_i)$$

Completing the proof. □

Proof of Theorem 3. For the sake of contradiction, suppose f is efficient and $\mathcal{O}^P$-implementable but it does not generate the same allocation as the serial dictatorship algorithm. For the remainder of the proof, define as $H_i = \{j \in I : j < i\}$ and $L_i = \{j \in I : j > i\}$ the set of agents who have (respectively) higher and lower priority than i. Also denote the object that agent i receives in state t according to SCF f as $f_i(t)$.

As f does not generate the same allocation as the sequential dictatorship algorithm, there exists a state $t \in T$, a player $i \in I$, and an object $x \neq f_j(t)$ for all $j \in H_i$ such that $u_i(x, t) > u_i(f_i(t), t)$.

By the rich domain assumption, we know there exists a type profile t'_{L_i} such that $u_{j^*}(f_i(t), t'_{j^*}) > u_{j^*}(x, t'_{j^*})$ for $j^* \in L_i$ and $u_j(x, t'_j) > u_j(f_i(t), t'_{j^*})$ for all $j \in L_i$ different from j^*. That is, there exists a state in which every agent j different from j^* with lower priority than i prefers x to $f_i(t)$.

As f is admissible, agents with lower priority than i cannot affect her outcome: $f_i(t) = f_i(t_{H_i}, t_i, t'_{L_i})$. This contradicts the assumption f is efficient: in state (t_{H_i}, t_i, t'_{L_i}), allocating x to i and $f_i(t)$ to j (leaving other agents' allocation unchanged) would strictly improve i's and j's welfare while leaving other agents' utility unaffected. This leads to a contradiction, concluding the proof. □

Proof of Theorem 4. If $f = (\gamma, \xi)$ is $\mathcal{O}^{D'}$-implementable, it is BIC and the implementing

mechanism belongs to $\mathcal{G}(\mathcal{O}^{D'})$. Let σ be the equilibrium implementing f. As the implementing mechanism is admissible, then for all $K \subseteq I$ there exists $s_i^K \in S_i$ and $\xi_i^K : T \to \Delta(X)$ such that:

$$\int_{T_{-i}} \mu(s_i^K, \sigma_{-i}(t_{-i}))\ \mathrm{d}p(t_{-i}|t_i) = \int_{T_{-i}} (\gamma^{-iK}(t_i', t_{-i}), \xi_i^K(t_i', t_{-i}))\ \mathrm{d}p(t_{-i}|t_i)$$

As σ is an equilibrium it follows:

$$\int_{T_{-i}} u_i(\mu(\sigma(t)), t)\ \mathrm{d}p(t_{-i}|t_i) \geq \int_{T_{-i}} \mu(s_i^K, \sigma_{-i}(t_{-i}), t)\ \mathrm{d}p(t_{-i}|t_i)$$

Substituting both sides then yields the result:

$$\int_{T_{-i}} u_i(f(t), t)\ \mathrm{d}p(t_{-i}|t_i) \geq \int_{T_{-i}} (\gamma^{-iK}(t_i', t_{-i}), \xi_i^K(t_i', t_{-i}))\ \mathrm{d}p(t_{-i}|t_i)$$

As for the sufficiency part, consider the following implementing mechanism. Agents choose an action $s_i = (s_i^1, s_i^2) \in T_i \times 2^I$. Define the outcome function μ as follows:

1. if $s_i = t_i \times \emptyset$ for all $i \in I$, then $\mu(s) = f(t)$
2. $\mu(s) = (\tilde{\gamma}(t), \bar{x}(t))$ for all other profiles s, where $\tilde{\gamma}(t) = \gamma(t) - \cup_{i \in I} \cup_{j \in s_i^2} g_{ij}$ for all $t \in T$ and $\tilde{\xi} = \xi_K^{-is_i^2}$ where k is the player with the lowest index such that $s_i^2 \neq \emptyset$

Notice this mechanism is $\mathcal{O}^{D'}$-admissible: no matter what (mixed) action her opponents are playing, there is always an action in each agent's opportunity set that allows her not to form links with any subset of her opponents. The same is true for each coalition of agents $K \subseteq I$ as well.

To show this mechanism implements f, let us consider strategy profile σ such that $\sigma_i(t_i) = t_i \times \emptyset$ for all $i \in I$ and $t_i \in T_i$. As f is BIC, it is not profitable for any agent of type t_i to deviate to $t_i' \times \emptyset$. Deviating to any $t_i' \times K$ for $K \neq \emptyset$ would not be profitable either. In fact, accountability implies that for all $t_i' \neq t_i$ and $K \subseteq I$ as for all $i \in I$, $t_i \in T_i'$ and

$K \subseteq I$:

$$\int_{T_{-i}} u_i(\gamma(t), \xi(t)), t) \ \mathrm{d}p(t_{-i}|t_i) \geq \int_{T_{-i}} u_i(\gamma^{-iK}(t_i', t_{-i}), \xi_i^K(t)), t) \ \mathrm{d}p(t_{-i}|t_i)$$

This completes the proof for the first part of the theorem.

Finally, suppose f is $\mathcal{O}^D$-admissible. Notice that, for any (mixed) action player by her opponents, player i can not generate any links that she could not generate by changing her report in the direct mechanism associated with f. A similar argument holds for all $K \subseteq I$, concluding the proof. □