Mechanism Design and Communication Networks *

Ludovic Renou† & Tristan Tomala‡

February 23, 2009

*We thank Dirk Bergemann, Subir Bose, Gianni De Fraja, Johannes Hörner, Claudio Mezzetti, Jérôme Renault, Karl H. Schlag, Sylvain Sorin, Nicolas Vielle, Yannick Viossat, Piercarlo Zanchettin and seminar participants at the First Northwestern-PSE Transatlantic Theory Workshop and Warwick CRETA Workshop for helpful discussions and comments. We owe this piece of work to a discussion Murali Agastya and one of the author had few years ago. Renou thanks the hospitality of Fuqua Business School at Duke University.

†Department of Economics, Astley Clarke Building, University of Leicester, University Road, Leicester LE1 7RH, United Kingdom. lr78@le.ac.uk

‡Department of Economics and Finance, HEC School of Management, 78351 Jouy-en-Josas Cedex, France. tomala@hec.fr

Abstract

This paper characterizes the communication networks for which, in any environment (utilities and beliefs), every incentive compatible social choice function is (partially) implementable. In environments with *either* common independent beliefs and private values *or* a bad outcome, we show that any incentive-compatible social choice function is implementable on a given communication network if and only if the network is *weakly 2-connected*. A network is weakly 2-connected if each player is either directly connected to the designer or indirectly connected to the designer through one of the disjoint path emanating from another player with two disjoint paths to the designer. We couple encryption techniques together with appropriate incentives to secure the transmission of each player's private information to the designer.

Keywords: Mechanism design, incentives, Bayesian equilibrium, communication networks, encryption, secure transmission, coding

JEL Classification Numbers: C72, D82

1 Introduction

The *revelation principle* is the cornerstone of the theory of mechanism design and its applications. It asserts that the outcome of any communication system can be replicated by a direct revelation mechanism, in which agents directly and privately communicate with a designer, and truthfully report all their information (Gibbard (1973), Dasgupta, Hammond and Maskin (1979), Myerson (1979), Harris and Townsend (1981), Myerson (1982)). As a technical result, the revelation principle is a blessing. It allows to abstract from the very details of communication systems and to focus on the social choice function to be implemented. At the same time, it is slightly disturbing as it implies that no decentralized communication system, however sophisticated, can dominate the centralized (direct) communication system. Real-world organizations seldom take the form of centralized communication systems. The aim of this paper is to characterize the communication systems which replicate the incentive properties of the centralized communication system and, thus, to show that incentive considerations alone can already explain the existence of a large variety of real-world organizations.[1]

An important application of our results is thus the design of optimal organizations. As already mentioned, our results characterize the organizations with the same incentive properties as the centralized organization. We can then appeal to other considerations e.g., computational complexity or size of the message spaces, to discriminate among all these organizations. While this is not the objective of this paper, we can note that, unlike the centralized organization, the designer does not need to communicate with all the players in our model: he only needs to communicate with *two* players. However, he needs to listen to twice as many messages. Another important application is the design of protocols (mechanisms) to guarantee the secure communication of "secrets" between multiple senders and a receiver in environments where senders respond

[1]There is a recent literature, labeled as *algorithmic mechanism design* in the computer science literature, e.g., Nisan and Segal (2006) and Van Zandt (2007), which focus on communication complexity and mechanism design. (See Nisan et al. (2007) for an excellent exposition.) Unlike this literature, we abstract from complexity considerations and entirely focus on the topology of the communication networks.

to incentives. Indeed, imagine a sender and a receiver as distant nodes in a network. The problem is to guarantee the communication of the sender's secret to the receiver without revealing information to others. We can view this problem as a mechanism design problem where the sender is a player, the receiver is the designer, and the social choice function coincides with the sender's secret. Our main results provide conditions on the networks and protocols to achieve such secure communication (see the literature review below).

Communication networks (systems) are naturally modeled as directed graphs, where the nodes represent the players and the designer. A player can directly communicate with another if there exists a directed edge from that player to the other. We then associate communication networks with social environments representing the preferences and beliefs of the players, and characterize the topology of communication networks for which, in any environment, *every* incentive-compatible social choice function is implementable.

A network is 2-connected if each player is either directly connected to the designer or indirectly connected to the designer through at least two disjoint paths. A network is *weakly-2-connected* if each player i is either directly connected to the designer or there exists a player $k(i)$ indirectly connected to the designer through at least two disjoint paths and such that player i belongs to one of these two paths. Our first main result states that in all environments with common independent beliefs and private values, any incentive-compatible social choice function is partially implementable if and only if the communication network is weakly-2-connected. The intuition for this result is as follows. A social choice function is (Bayesian) incentive compatible if, when each player expects the others to tell the truth, then no player has an incentive to lie about his own information. Importantly, players use their prior beliefs to form their expectations. However, in a general communication network, players receive messages from their neighbors. Consequently, their incentives to tell the truth may be altered since posterior beliefs may differ from prior beliefs. To circumvent this problem, we couple encryption techniques and incentives to transfer securely each player's private information to the designer through the network. Our encoding technique guarantees

that no player learns anything about the types of the other players and, therefore, posterior beliefs are equal to prior beliefs. Assume that the network is 2-connected. A player can send a private "encoding" key to the designer through one path, and his type encoded with the key, a "cypher-type" through the other (disjoint) path. However, this is not sufficient: players must also have an incentive to truthfully forward the messages they receive. Our encoding technique guarantees furthermore that a player's expected payoff is independent of the messages he forwards. Therefore, players have an incentive to truthfully forward the messages of their neighbors. Incentive compatibility ensures that players also have an incentive to truthfully report their own private information.

The encoding technique we use is tailored to environments with common independent beliefs and does not extend to more general environments. To do so, we resort to a different encoding technique. Beyond insuring secrecy, this new encoding technique "authenticates" the player's messages in that if a player does not truthfully forward the messages of his neighbors, he is detected with probability one. In environments with a bad outcome, i.e., an alternative that yields a smaller utility to each player at each type profile, the threat to be punished with the bad outcome upon detection of a false report deters players from lying about the messages of their neighbors. Again, incentive compatibility insures that players also have an incentive to truthfully report their own private information. This is our second main result.[2]

Related literature. The computer science literature on secure transmission of messages is closely related to this paper. This literature studies the problem of securely transmitting a secret message from a sender to a receiver in a general network, in which processors (players) might be malicious (active Byzantine adversaries). The transmission of messages is secure if it is both perfectly reliable and secret. Perfect reliability means that the receiver correctly infers the message transmitted by the sender, while perfect secrecy means that no adversary learns anything about the message sent. This literature then characterizes the class of networks, which guarantee secure transmission of messages from the sender to the receiver, *regardless* of the behavior of the adversaries. Considering undirected and unicast networks, Dolev et al. (1990) show

[2]To the best of our knowledge, the encoding technique we use is novel.

that in 1-way problems, i.e. if the information flows only from the sender to the receiver, a sufficient and necessary condition for the secure transmission of information is the 4-connectivity of the network, while in 2-way problems, i.e. when the sender and receiver "converse", a sufficient and necessary condition is the 3-connectivity of the network. Similarly, considering undirected but broadcast networks, Franklin and Wright (2000) show that a necessary and sufficient condition for the secure transmission of messages is the 3-connectivity of the network (see also Renault and Tomala (2008)). Considering directed networks, as we do in this paper, Desmedt and Wang (2002) show that if there are $4 - 3l \geq 2$ disjoint paths from the sender to the receiver and l disjoint paths from the receiver to the sender (these l paths are also disjoint from the $4 - 3l$ paths from the sender to the receiver), then the secure transmission of messages can be achieved.

As in the computer science literature, the implementation on a communication network of any incentive-compatible social choice functions requires to construct strategies (protocols) that guarantee the secure transmission of messages. However, and crucially, in our setting, players (nodes) do not play arbitrarily, but rather optimally respond to incentives. As a consequence, encryption techniques together with appropriate incentives weaken the conditions necessary for the secure transmission of messages (private information) to the designer: the weak 2-connectivity of the communication network is necessary and sufficient in a large class of environments. Furthermore, we do not need multiple rounds of communication: a player "speaks" only once in our model.

A paper closely related to our work is Monderer and Tennenholtz (2001), who study the same problem as ours. Our paper substantially generalizes their results. In particular, these authors assume the existence of a worst outcome, common independent beliefs and private values, and show that 2-connectivity of the network is a sufficient condition. Our encoding techniques apply to a broader set of networks and environments, and we provide necessary and sufficient conditions on the networks.

The literature on organizations and hierarchies (see Mookherjee (2006) and references therein) is also related to this paper. This literature considers simple organizational arrangements and provides conditions for the equivalence between central-

ized and decentralized organizations for the partial implementation of *particular* social choice functions in particular environments. The focus of the present paper is different, however. We characterize the topology of communication networks necessary for the partial implementation of *all* incentive-compatible social choice functions, regardless of the environment.

The paper is organized as follows. Section 2 presents a simple example to illustrate our main results. Section 3 introduces our formal model, while Section 4 presents the main theorems and proofs. Section 5 discusses our results and offers several extensions and open issues. Finally, Section 6 concludes.

2 Illustration of the results

This section illustrates our main results within the context of a simple example. There are three players, labeled 1, 2 and 3, two types for player 2, labeled θ and θ', and two alternatives a and b. Player 2's preferences over these alternatives depend on his type (In all examples, preferences are strict). Player 2 prefers a to b if his type is θ and prefers b to a if his type is θ'. Player 1 always prefers a to b, while player 3 always prefers b to a. The designer aims at implementing the social choice function f^* that selects the preferred alternative of player 2 for each of his type: player 2 is dictatorial.

If player 2 can securely and directly communicate with the designer, f^* is clearly implementable: the designer can ask player 2 to directly report his type and select the alternative accordingly. Suppose now that player 2 cannot directly communicate with the designer and consider the communication network $\mathcal{N}_1$ in Figure 1 (player 0 is the designer).

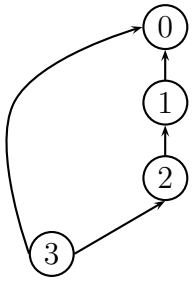

Figure 1: Communication network $\mathcal{N}_1$

With the communication network $\mathcal{N}_1$, player 2 can indirectly communicate with the designer through player 1. Moreover, player 3 has two disjoint paths of communication to the designer with player 2 on one of them: the network $\mathcal{N}_1$ is *weakly* 2*-connected*. The idea is then to use the two disjoint paths so as to guarantee the communication of player 2's type to the designer, without revealing information to the other players. So, suppose that players 1 and 3 share a common prior and believe that player 2's type is θ with probability $1/3$, independently of their own types. The goal is to design a mechanism and an equilibrium such that the designer implements a in state θ and b in state θ'.

The mechanism allows player 3 to send a real number in $[0,1]$ to player 2 and another real number in $[0,1]$ to player 0 (the designer). Player 2 can send a real number in $[0,1]$ to player 1. In turn, player 1 can send also a real number in $[0,1]$ to the designer. An informal description of the strategies is as follows. Regardless of his type, player 3 draws an "encoding key" y uniformly on $[0,1]$ and sends it to both players 0 and 2. Player 2 of type θ (resp., θ') draws a "pseudo-type" $\tilde{x}$ (resp., $\tilde{x}'$) uniformly on $[0,1/3)$ (resp., $[1/3,1]$), and encodes his pseudo-type $\tilde{x}$ with the encoding key y received from player 3 to obtain the "cypher-type" $x = (\tilde{x}+y) \,\mathsf{mod}_{0,1}$ (resp., $x' = (\tilde{x}'+y) \,\mathsf{mod}_{0,1}$).[3] Player 2 of type θ (resp., θ') sends x (resp., x') to player 1. Player 1 has to correctly forward the message of player 2 to the designer. Let $(\hat{x},\hat{y})$ be a pair of messages received by the designer. The allocation rule is the following: If $(\hat{x}-\hat{y}) \,\mathsf{mod}_{0,1} \in [0,1/3)$, the designer implements a and, otherwise, implements b.

If the players follow the prescribed strategies, the designer correctly learns player 2's type and implements the desired social choice function f^*. We now show that the players do not have an incentive to deviate from the prescribed strategies. Suppose that player 1 deviates and sends a message $\hat{x}$ to the planner instead of x (or x'). The designer implements the alternative a if $(\hat{x}-y) \,\mathsf{mod}_{0,1} \in [0,1/3)$ and b otherwise. Since y is uniformly distributed, so is $(\hat{x}-y) \,\mathsf{mod}_{0,1}$ (see Lemma 1 below). Accordingly, player 1 expects the designer to implement a with probability $1/3$ and b with probability $2/3$. Thus, player 1's expected payoff does not depend on the message sent $\hat{x}$ and has,

[3]For a real number r, $r \,\mathsf{mod}_{0,1} = r - \lfloor r \rfloor$, with $\lfloor r \rfloor$ the highest integer less or equal to r.

therefore, no incentive to deviate. A similar argument applies to player 3. As for player 2, he has no incentive to deviate since f^* is incentive-compatible.

It is worth stressing that the essential feature of the network is its weak 2-connectedness. For instance, if in addition to the links shown in Figure 1, player 3 has also a link to player 1, the result remains valid. (Note that the network remains weakly 2-connected.) Indeed, we can construct a "babbling equilibrium" in which player 3 sends an uninformative message to player 1, and player 1 plays independently of player 3's message. Alternatively, and more simply, we may let the message space from player 3 to player 1 be a singleton. In fact, we show that the weak 2-connectedness of the network is a necessary and sufficient condition for the implementation of any incentive-compatible social choice functions in environments with independent common beliefs and private values.

A further and important feature of the proposed mechanism and strategies is that players 1 and 3 learn nothing about player 2's type. This is clearly true for player 3 as he does not receive a message from player 2. As for player 1, we prove that the message x (or x') he receives is uniformly distributed on $[0, 1]$, independently of player 2's type. This feature is crucial for the implementation of incentive-compatible social choice functions, which depend on the private information of all players. It guarantees that posterior beliefs are equal to prior beliefs and, consequently, that each player's incentives to truthfully reveal their own private information are not altered.

Another important aspect is that the mechanism and strategies are tailored to environments with common independent beliefs and private values. Firstly, the partition of $[0, 1]$ into $\{[0, 1/3), [1/3, 1]\}$ is such that the Lebesgue measure of each subset exactly matches the *common* prior beliefs of players 1 and 3. Now, suppose that player 1 believes that player 2's type is θ with probability $2/3$. With the above strategies, player 1 expects the designer to decode player 3's type as being θ with probability $1/3$, which is different from his prior belief $2/3$. Consequently, player 1's incentive to truthfully report his private information might be altered. Secondly, to understand the importance of the private values assumption, suppose that player 1 prefers b to a when player 2's type is θ and a to b when player 2's type is θ' (interdependent values).

If player 1 truthfully forwards the message x he received from player 2, the alternative a is implemented if and only if player 2's type is θ and b is implemented if and only if player 2's type is θ'. However, if he sends a message $\hat{x}$ independently of the message received from 2, both alternatives a and b are implemented with positive probability, regardless of player 2's type, a profitable deviation for player 1. In sum, the problem with more general environments is not to guarantee that no information is revealed, but to guarantee that the other players have incentives to truthfully communicate their private information and the messages they receive.

With more elaborated encryption techniques, our result remain valid in environments with a worst alternative (Theorem 2). The intuition is as follows. Consider again the network $\mathcal{N}_1$. Player 3 can draw a large number of independent encoding keys $y_1, \ldots, y_\eta$ and sends them to players 0 and 2. Player 2 privately draws one of the keys and uses it to encrypt his type. He then sends to player 1 the encrypted type and the unused keys, *without telling him which key was used for coding.* Player 1 correctly forwards player 2's message to the designer. The designer compares the two vectors he received. If these vectors differ in exactly one component η^*, he infers that the key y_{η^*} transmitted by player 3 was used for coding, and decodes player 2's type accordingly. Otherwise, the designer implements the worst alternative. This encoding technique guarantees that players 1 and 3 learn nothing about player 2's type and allow the designer to detect unilateral deviations with arbitrarily high probability.[4] In turn, the threat to implement the worst alternative upon detection of a deviation deters players from lying.

To conclude this section, we preview some secondary aspects of our analysis. Firstly, the use of probabilistic coding implies that our equilibria are in mixed strategies. This point is crucial as the social choice function f^* of our example is not implementable in pure equilibria. Section 5.2 elaborates on this issue. Secondly, unlike the computer science literature, our encoding technique relies on transmitting real numbers, which may not have a finite binary expansion. This choice is well in accordance with the implementation literature: a social choice function is implementable if there exists a

[4]To the best of our knowledge, this encoding technique is novel.

mechanism, possibly with continuous action spaces, and an equilibrium which corresponds to the social choice function at each state. With some modifications, our main results can be obtained with finite message spaces (see Section 5).

3 Definitions

The primitives of the model consist of two essential ingredients: social environments (players, outcomes and preferences) and communication networks.

A ***social environment*** $\mathcal{E}$ is a tuple $\langle N, A, (\Theta_i, P_i, u_i)_{i\in N}\rangle$ where $N := \{1, \ldots, n\}$ is the set of players, A the finite set of alternatives, and Θ_i the finite set of types of player $i \in N$.[5] Let $\Theta := \times_{i\in N}\Theta_i$ and $\Theta_{-i} := \times_{j\in N\setminus\{i\}}\Theta_j$, with generic elements θ and θ_{-i}, respectively. Each player knows his own type and player i of type θ_i holds a probabilistic belief $P_i(\cdot|\theta_i)$ over Θ_{-i}. Throughout the paper, we assume $P_i(\theta_{-i}|\theta_i) > 0$ for all $(\theta_i, \theta_{-i}) \in \Theta$ and for all $i \in N$. Each player has a preference relation over alternatives, which is representable by the type-dependent utility function $u_i : A\times\Theta \to \mathbb{R}$. Players are expected utility maximizers. Four properties of an environment are of particular importance to our analysis:

- The environment has a *common prior* if there exists a probability distribution P on Θ such that $P_i(\theta_{-i}|\theta_i)$ is the conditional distribution of θ_{-i} given θ_i derived from P. The common prior is *independent* if P is the product of its marginal distributions.

- The environment has *private values* if for each player i, his utility function does not depend on the types θ_{-i} of his opponents.

- The environment has a *bad outcome* if there exists an alternative $\underline{a} \in A$ such that for each player i, each type profile θ and each alternative $a \in A$, $u_i(\underline{a}, \theta) \leq u_i(a, \theta)$.

- The environment has a *worst outcome* if there exists an alternative $\underline{a} \in A$ such that for each player i, each type profile θ and each alternative $a \in A \setminus \{\underline{a}\}$, $u_i(\underline{a}, \theta) < u_i(a, \theta)$.

[5] In Section 5, we extend our analysis to environments with infinite type spaces.

A social choice function $f : \Theta \to A$ associates with each type profile θ an alternative $f(\theta) \in A$. A social choice function is *incentive compatible* if for each player $i \in N$, for each type θ_i of player i:

$$\sum_{\theta_{-i}} u_i(f(\theta_i, \theta_{-i}), \theta_i, \theta_{-i}) P_i(\theta_{-i}|\theta_i) \geq \sum_{\theta_{-i}} u_i(f(\theta'_i, \theta_{-i}), \theta_i, \theta_{-i}) P_i(\theta_{-i}|\theta_i),$$

for all types θ'_i.

Note that our definition of a bad (resp., worst) outcome is stronger than actually required; it would be enough to consider an alternative worse than any alternative in the range of the social choice function we aim to implement. Exchange economies with free disposal are examples of environments with bad (resp., worst) outcome: the zero allocation is a bad (resp., worst) outcome if preferences are monotonic (resp., strictly monotonic) and the social choice function selects positive vectors of goods.

A ***communication network*** captures the possibilities of communication between the players and the designer. A communication network is a *directed* graph with $n+1$ vertices representing the n players and the designer (henceforth player 0). There is a directed edge from player i to player j, denoted ij, if i can send a message to j. Formally, the network, denoted by $\mathcal{N}$, is defined as a set of edges $\mathcal{N} \subseteq (N \cup \{0\}) \times (N \cup \{0\})$. By convention, we assume $ii \notin \mathcal{N}$ for each i. We denote $C(i) = \{j \in N \cup \{0\} : ij \in \mathcal{N}\}$ the set of players to whom player i can directly send a message. Similarly, we denote $D(i) = \{j \in N \cup \{0\} : ji \in \mathcal{N}\}$ the set of players who can directly send a message to player i. A *path* in $\mathcal{N}$ is a finite sequence of vertices $(i_1, \ldots, i_m)$ such that $i_k i_{k+1} \in \mathcal{N}$ for each $k = 1, \ldots, m-1$. A communication network $\mathcal{N}$ is *m-connected* if for each player $i \in N \setminus D(0)$, there exist m disjoint paths (i.e., having no common vertex except i and 0) from player i to the designer. By convention, the communication network is n-connected if $N \setminus D(0) = \emptyset$. A network of particular importance is the star network $\mathcal{N}^\star$ with the designer as the center and $D(i) = \emptyset$, $C(i) = \{0\}$ for all player $i \in N$. With the star network, each player communicates directly and privately with the designer; the star network is n-connected. Throughout the paper, we assume that networks are 1-connected: for each player $i \in N$, there exists a path from i to 0. This assumption ensures that the designer may receive information from each player.

Now, we describe the interaction between a social environment and a communication network. In the standard theory of mechanism design, players know their type and communicate directly with the designer, who selects an outcome. In our model with communication networks, player i can only send messages to players in $C(i)$. The social interaction (the extensive-form) unfolds as follows.

- Knowing his type, each player i reads the messages he receives from players in $D(i)$. Then, he sends messages to players in $C(i)$ (he may send different messages to different players).
- The designer "reads" the messages he receives from players in $D(0)$ and selects an alternative.

Note that if $\mathcal{N} = \mathcal{N}^\star$, this corresponds to the classical model where each player communicates directly and privately with the designer. We make the following assumptions on the network. Firstly, we assume that $C(0) = \emptyset$, i.e., the designer cannot send messages to the players. In other words, as in the classical models of mechanism design, the designer is not active in the game: he merely collects information and implements outcomes accordingly. Secondly, we assume that the restriction of the graph to the set of players N is acyclic, that is, for each $i \in N$, there is no path from i to himself. Acyclicity and 1-connectedness of the graph implies that communication as described above gives rise to a well-defined extensive-form. With acyclicity, the communication rule stating that "*a player sends his messages after having received all his messages*" generates a well-defined timing structure, where each player i is assigned a stage $t(i)$ at which he sends his messages. This statement is proved in Appendix. For instance, in Figure 2, player 3 can communicate with player 1, but not with player 2 and the designer. In the associated extensive-form, player 3 communicates first with player 1, and after observing player 3's message, player 1 communicates with the designer. The assumptions of inactive designer and acyclicity (i.e., each player "speaks" only once) make our problem of implementation the hardest. In Section 5, we discuss several extensions. In particular, we discuss an extension of the model where the designer may send messages and show that this simplifies our problem.

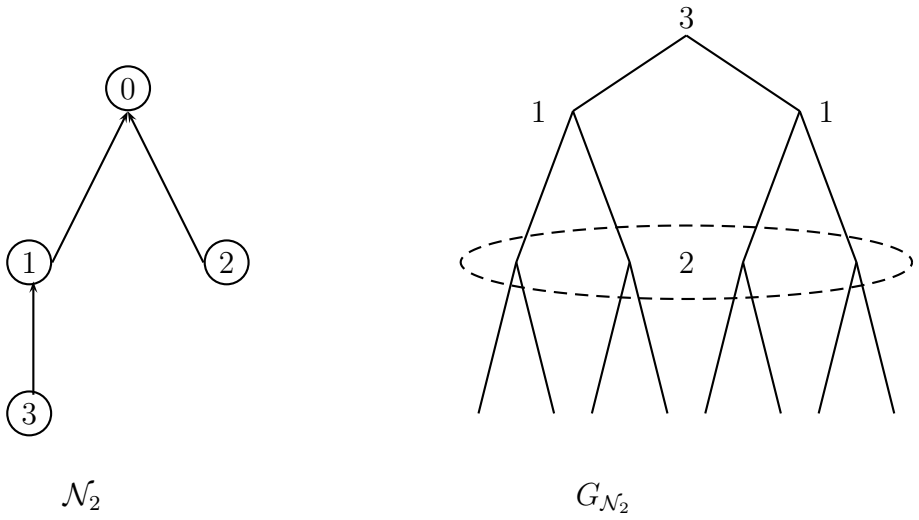

Figure 2: Network $\mathcal{N}_2$ and a consistent extensive-form $G_{\mathcal{N}_2}$

A mechanism is a pair $\langle (M_{ij})_{ij\in\mathcal{N}}, g\rangle$ where for each edge ij, M_{ij} is the set of messages that player i can send to player j, and $g : \times_{i\in D(0)} M_{i0} \to A$ is the allocation rule. Note that the allocation rule depends only on the messages the designer can receive. The next step is to define the Bayesian game induced by a mechanism, a communication network and an environment.

Fix an environment $\langle N, A, (\Theta_i, P_i, u_i)_{i\in N}\rangle$, a communication network $\mathcal{N}$ and a mechanism $\langle (M_{ij})_{ij\in\mathcal{N}}, g\rangle$. Define $M_{D(i)} := \times_{j\in D(i)} M_{ji}$ as the set of messages player i can receive and $M_{C(i)} := \times_{j\in C(i)} M_{ij}$ as the set of messages that player i can send. A *pure strategy* s_i for player i is a mapping from $M_{D(i)} \times \Theta_i$ to $M_{C(i)}$. We denote by S_i the set of player i's pure strategies and by $s_{ij}(m_{D(i)}, \theta_i)$ the message player i sends to player $j \in C(i)$ conditional on receiving the messages $m_{D(i)}$ and being of type θ_i. A *behavioral strategy* σ_i for player i maps $M_{D(i)} \times \Theta_i$ to $\Delta(M_{C(i)})$, the set of probability distributions over $M_{C(i)}$[6]. We denote by $\mathbb{P}_{\sigma,\theta}$ the probability distribution over profiles of messages (i.e., over $\times_{ij\in\mathcal{N}} M_{ij}$) induced by the strategy profile $\sigma = (\sigma_i)_{i\in N}$ at state θ. The Bayesian game $G_{\mathcal{N}}$ induced by an environment, a mechanism and a network is

[6]We also find it convenient to view a behavioral strategy as a measurable mapping from $M_{D(i)} \times \Theta_i \times Y_i$ to $M_{C(i)}$, where $(Y_i, \mathcal{Y}_i, \mu_i)$ is a probability space independent of types and messages, i.e., a private randomization device.

defined as follows:

- The set of players is N, the set of player i's types is Θ_i and his beliefs are given by P_i.
- The set of strategies of player i is S_i.
- The payoff of player i is his expected payoff conditional on his type and given that the outcomes are selected by the allocation rule g.

Definition 1 *The social choice function f is partially implementable on the communication network $\mathcal{N}$ if there exist a mechanism $\langle (M_{ij})_{ij\in\mathcal{N}}, g\rangle$ and a Bayesian-Nash equilibrium σ^* of $G_{\mathcal{N}}$ such that for all $\theta \in \Theta$, $g((m^*_{i0})_{i\in D(0)}) = f(\theta)$ for all profiles of messages $(m^*_{i0})_{i\in D(0)}$ received by the designer in the support of $\mathbb{P}_{\sigma^*,\theta}$.*

Denote $F_{\mathcal{N}}(\mathcal{E})$ the set of social choice functions partially implementable on the communication network $\mathcal{N}$ when the environment is $\mathcal{E}$. From the revelation principle, $F_{\mathcal{N}}(\mathcal{E}) \subseteq F_{\mathcal{N}^\star}(\mathcal{E})$ for every environment $\mathcal{E}$, and $F_{\mathcal{N}^\star}(\mathcal{E})$ is precisely the set of incentive compatible social choice functions. The aim of this paper is to characterize the communication networks $\mathcal{N}$ for which $F_{\mathcal{N}}(\mathcal{E}) = F_{\mathcal{N}^\star}(\mathcal{E})$ for every environment $\mathcal{E}$.

4 The main results

This section presents our main results regarding the partial implementation of social choice functions on communication networks. We introduce the following connectivity condition.

Definition 2 *The communication network $\mathcal{N}$ is* weakly 2-connected *if for each player $i \in N \setminus D(0)$, there exist a player k (possibly i) and two disjoint paths π^1_k and π^2_k from player k to the designer such that i belongs to either π^1_k or π^2_k.*

In words, a network is weakly 2-connected if for each player not directly connected to the designer, either there exist two disjoint paths from this player to the designer or this player is connected to the designer via a path, which is part of a pair of disjoint

paths linking another player to the designer. For instance, in Figure 3, the network $\mathcal{N}_3$ is weakly 2-connected while the network $\mathcal{N}_3'$ is not. Note that in both networks, player 2 has a unique path to the designer and, therefore, neither network is 2-connected.

Importantly, if a network is not weakly connected, there exists a pair of players (i, i^*) such that all paths from player i to the designer goes through player i^* and, furthermore, all paths from any player $j \neq i$ to the designer that go through player i also go through player i^*. Player i^* "controls" all the possible messages that player i can use to communicate his private information. For instance, with the network $\mathcal{N}_3'$, player 1 controls all messages that player 2 can send. In turn, this simple observation suggests that there is no hope to implement all incentive-compatible social choice functions on a network that is not weakly 2-connected. We will show that it is indeed the case.

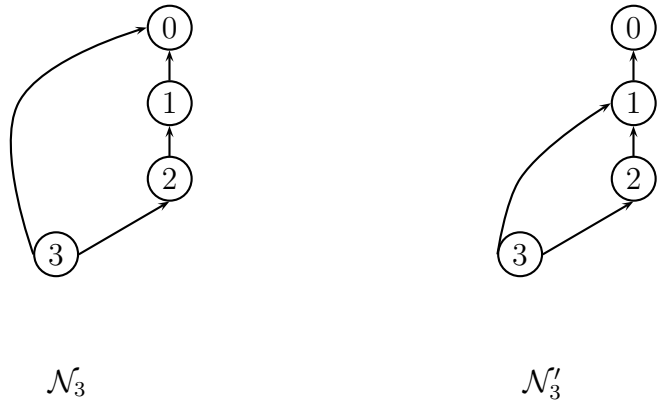

Figure 3: $\mathcal{N}_3$ is weakly 2-connected, $\mathcal{N}_3'$ is not

4.1 Common independent beliefs and private values

We first consider environments with common independent beliefs and private values. This assumption is common in several applications of the theory of mechanism design, e.g., auction theory (Krishna (2002)) or contract theory (Salanie (2000)). Our first result states that any incentive compatible social choice function is implementable on $\mathcal{N}$ for all such environments if and only if $\mathcal{N}$ is weakly 2-connected.

Theorem 1 *For all environments $\mathcal{E}$ with common independent beliefs and private values, $F_{\mathcal{N}}(\mathcal{E}) = F_{\mathcal{N}^\star}(\mathcal{E})$ if and only if $\mathcal{N}$ is weakly 2-connected.*

The proof of Theorem 1 proceeds as follows. We first show how to implement the dictatorial social choice function of player i. Since the communication network is weakly 2-connected, there exist a player k and two disjoint paths from player k to the designer, one of which contains player i. The main idea is then to use these two disjoint paths to securely transmit player i's type to the designer without revealing information to the other players. Intuitively, our construction requires player k to send an encryption key to the designer and player i through the two disjoint paths. Player i can then use the encryption key to encode his type and send his encoded type to the designer via the path from player k to the designer that goes through him. The strategies require all other players on these two paths to truthfully forward the message they receive (see Section 2 for an example). In a second part, we show how to generalize our construction to implement any social choice function. Finally, we show that weak 2-connectedness is a necessary condition to implement all incentive compatible social choice functions. To get some intuition for this result, let us consider a simple example. There are two players, 1 and 2, two alternatives, a and b, and two types, θ and θ' for each player. Regardless of his type, player 1 prefers a over b, and player 2 of type θ prefers a over b, while player 2 of type θ' prefers b over a. Consider the social choice function f for which player 2 is dictatorial and the communication network in Figure 4. The issue with this network and, more generally, with any communication network that is not weakly 2-connected, is that player 1 controls all the information sent by player 2, and there is no way for the designer to detect a false report by player 1.

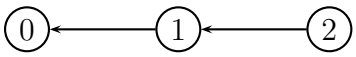

Figure 4: Communication networks $\mathcal{N}$ is 1-connected

Clearly, f is implementable on the star network $\mathcal{N}^\star$, but not on $\mathcal{N}$. By contradiction, suppose that f is implementable on $\mathcal{N}$ by the mechanism $\langle M_1, M_2, g\rangle$. There must exist a message $m_1 \in M_1$ such that $g(m_1) = b$. However, regardless of his type and

message received, player 1 has no incentives to send any message m_1 with $g(m_1) = b$, so that f cannot be implemented. The proof of Theorem 1 generalizes this argument to any network that is not weakly 2-connected.

Two further remarks are worth making. Firstly, our encoding technique extends to environments with continuous type spaces (see Subsection 5.3). Secondly, the strategies we consider are behavioral strategies. In Subsection 5.2, we prove that our result does not hold if we restrict ourself to pure equilibria, a frequently used solution concept in the mechanism design literature.

Let us now turn to the proof of Theorem 1. We first present three important properties about the modular manipulations of real numbers in $[0, 1]$. All our coding techniques rest on these properties. For a real number x, we denote $\lfloor x \rfloor$ the greatest integer less than or equal to x. We define $x \,\mathsf{mod}_{0,1}$ as $x - \lfloor x \rfloor$, the fractional part of x.

Lemma 1 *1. For each $x \in [0, 1]$, $y \in [0, 1]$,*

$$((x + y) \,\mathsf{mod}_{0,1} - y) \,\mathsf{mod}_{0,1} = x$$

2. Let Y be a random variable in $[0, 1]$ and $x \in [0, 1]$. If Y is uniformly distributed, then so are $(x + Y) \,\mathsf{mod}_{0,1}$ and $(x - Y) \,\mathsf{mod}_{0,1}$.

3. Let X, Y be independent random variables in $[0, 1]$. If Y is uniformly distributed, then so are $Z = (X + Y) \,\mathsf{mod}_{0,1}$ and $W = (X - Y) \,\mathsf{mod}_{0,1}$. Further, (X, Y, Z) (resp., (X, Y, W)) are pairwise-independent.

Proof of Lemma 1. (1) Consider any pair $(x, y) \in [0, 1] \times [0, 1]$. If $x + y \leq 1$ the statement is clear. If $x+y > 1$, $(x+y) \,\mathsf{mod}_{0,1} = x+y-1$. Thus $(x+y) \,\mathsf{mod}_{0,1}-y = x-1$ and $(x - 1) \,\mathsf{mod}_{0,1} = x$.

(2) For each $z \in [0, 1]$, we have

$$\begin{aligned}
\mathbb{P}((x + Y) \,\mathsf{mod}_{0,1} \leq z) &= \mathbb{P}((x + Y) \leq z, Y \in [0, 1 - x]) + \\
&\quad \mathbb{P}(x + Y - 1 \leq z, Y \in (1 - x, 1]) \\
&= \begin{cases} z - x + x & \text{if } z \geq x \\ z + 1 - x - (1 - x) & \text{if } z < x \end{cases} \\
&= z
\end{aligned}$$

Thus, $(x+Y)\,\mathsf{mod}_{0,1}$ is uniformly distributed. Similarly, for each $z \in [0,1]$,

$$\begin{aligned}\mathbb{P}((x-Y)\,\mathsf{mod}_{0,1} \leq z) &= \mathbb{P}(x-Y \leq z, Y \in [0,x]) + \\ &\quad \mathbb{P}(x-Y+1 \leq z, Y \in (x,1]) \\ &= \begin{cases} x+1-(x+1-z) & \text{if } z \geq x \\ z+0 & \text{if } z < x \end{cases} \\ &= z\end{aligned}$$

Thus, $(x-Y)\,\mathsf{mod}_{0,1}$ is uniformly distributed.

(3) We only show that X and Y are independent, the rest being similar. For each $z \in [0,1]$, $\mathbb{P}(Z \leq z \mid X = x) = \mathbb{P}(x+Y \leq z) = z$. □

Proof of Theorem 1. We first prove the "if" part. The proof is constructive and proceeds in two parts. In the first part, we design a "sub-mechanism" to implement any dictatorial social choice function. In the second part, we concatenate these "sub-mechanisms" to implement any incentive compatible social choice function.

Fix an environment $\mathcal{E}$ with common independent beliefs and private values. Denote P^i the marginal distribution of the common belief P on Θ_i, i.e., this is the common belief of any player $j \neq i$ on Θ_i. Without loss of generality, assume that $\Theta_i := \{1, \ldots, t_i, \ldots, T_i\}$ for each player $i \in N$ and denote $\overline{P}^i(t_i) = \sum_{\theta_i \leq t} P^i(\theta_i)$, the cumulative distribution function of P^i. Define a partition $\Pi_i = \{\Pi_i(1), \ldots, \Pi_i(T_i)\}$ of $[0,1]$ into T_i subsets with $\Pi_i(t_i) = [\overline{P}^i(t_i - 1), \overline{P}^i(t_i))$ (with $\overline{P}^i(0) = 0$). Note that if X is uniformly distributed on $[0,1]$, the event $\{X \in \Pi_i(t_i)\}$ has probability $P^i(t_i)$. In what follows, Y_i, $X_i^1, \ldots, X_i^{t_i}, \ldots, X_i^{T_i}$ denote $T_i + 1$ independent random variables, uniformly distributed on $[0,1]$, $\Pi_i(1), \ldots, \Pi_i(t_i), \ldots, \Pi_i(T_i)$ respectively, and $y_i, x_i^1, \ldots, x_i^{t_i}, \ldots, x_i^{T_i}$ denote realizations.

Since the communication network $\mathcal{N}$ is weakly 2-connected, for each player $i \in N \setminus D(0)$, there exists a player $k(i) = k$ and two disjoint paths $\pi_k^1 = (k = i_0, i_1, \ldots, i_Q, i_{Q+1} = 0)$ and $\pi_k^2 = (k = j_0, j_1, \ldots, j_R, j_{R+1} = 0)$ from player k to the designer 0 such that $i_{q^*} = i$ for some $q^* \in \{0, \ldots, Q\}$.

Part I. Assume that f selects a preferred alternative of player i at any state (θ_i, θ_{-i}), i.e., for any θ_i, define $f^*(\theta_i) \in \arg\max_a u_i(a, \theta_i)$ and let $f(\theta_i, \theta_{-i}) = f^*(\theta_i)$

for all θ_{-i}: player i is dictatorial. If $i \in D(0)$, f is clearly implementable. Assume that $i \notin D(0)$. The problem is then to design a mechanism and an equilibrium such that player i has an incentive to truthfully reveal his type and no other player has an incentive to manipulate the transmission of information from player i to the designer. We now describe the mechanism $\langle M, g\rangle$.

Messages M**.** Player $k(i) = k$'s message space is given by $M_{ki_1} = [0,1] = M_{kj_1}$, and $M_{kl} = \{\emptyset\}$ for all $l \in C(k) \setminus \{i_1, j_1\}$. The message space of any player i_q or j_r, $q = 1, \ldots, Q$, $r = 1, \ldots, R$, is $M_{i_q i_{q+1}} = M_{j_r j_{r+1}} = [0,1]$, and $M_{i_q l} = M_{j_r l'} = \{\emptyset\}$ for all $l \in C(i_q) \setminus \{i_{q+1}\}$, for all $l' \in C(j_r) \setminus \{j_{r+1}\}$. All other players have no messages.

Allocation rule g**.** Let $(\hat{x}_i, \hat{y}_i)$ be the messages received by the designer from players i_Q and j_R. If $(\hat{x}_i - \hat{y}_i) \,\mathsf{mod}_{0,1} \in \Pi_i(t_i)$, let $\hat{\theta}_i = t_i$ and $g(\hat{x}_i, \hat{y}_i) = f^*(\hat{\theta}_i)$.

Strategies σ**.** The strategies are as follows. Assume that $k \neq i$.

Player k*.* Regardless of his type and of his messages, player k draws an encoding key y_i (a realization of Y_i) and sends it to players i_1 and j_1. Formally, his behavior strategy is defined by $\sigma_k(\theta_k, \cdot)[y_i] = (y_i, y_i) \in M_{ki_1} \times M_{kj_1}$.[7]

Player $i = i_{q^*}$*.* Player i of type $\theta_i = t_i$ draws $x_i^{t_i}$, a realization of $X_i^{t_i}$. If he receives the message y_i from i_{q^*-1}, he sends the cypher-type $x_i = (x_i^{t_i} + y_i) \,\mathsf{mod}_{0,1}$ to player i_{q^*+1}. The behavioral strategy of player i of type $\theta_i = t_i$ is thus $\sigma_i(\theta_i, y_i)[x_i^t] = (x_i^t + y_i) \,\mathsf{mod}_{0,1} \in M_{i_{q^*} i_{q^*+1}}$.

If $k = i$, player i performs both the tasks of player k and of player i_{q^*}. That is, player i of type $\theta_i = t_i$ draws y_i and $x_i^{t_i}$ and sends y_i to j_1 and $x_i = (x_i^{t_i} + y_i) \,\mathsf{mod}_{0,1}$ to i_1.

Other players. Regardless of his type, player i_q (resp., j_r) truthfully forwards the message he receives to player i_{q+1} (resp., j_{r+1}), for $q = 1, \ldots, Q$, $q \neq q^*$, (resp., $r = 1, \ldots, R$). The behavioral strategy of player i_q is thus $\sigma_{i_q}(\theta_{i_q}, x_i')[\cdot] = x_i'$ for all $x_i' \in [0,1]$, for all $\theta_{i_q} \in \Theta_{i_q}$.

Let us prove that σ is a Bayesian equilibrium that implements f.

[7] We denote $\sigma_k(\theta_k, m)[y]$ the action played by player k of type θ_k who has observed the message m, and drawn the random element y from a private randomization device.

Claim 1. *Under σ, the decoded type $\hat{\theta}_i$ coincides with the true type θ_i.*

Assume that $\theta_i = t_i$, thus $x_i^{t_i} \in \Pi_i(t_i)$. The designer receives $\hat{x}_i = (x_i^{t_i} + y_i) \,\mathsf{mod}_{0,1}$ and $\hat{y}_i = y_i$. From part 1 of Lemma 1,

$$((x_i^{t_i} + y_i) \,\mathsf{mod}_{0,1} - y_i) \,\mathsf{mod}_{0,1} = x_i^{t_i},$$

and thus $\hat{\theta}_i = \theta_i$.

Claim 2. *For each player $j \neq i$, the distribution of θ_i conditional on player j's type and on the messages received by j is P^i.*

This clearly holds true for players who receive no messages and for players i_q, $q < q^*$, j_r, $r = 1, \ldots, R$, who observe the realization of Y_i which is independent of θ_i. For players i_q, $q = q^* + 1, \ldots, Q$, this follows from part 3 of Lemma 1.

Claim 3. *If a player other than player i does not truthfully forward his message, his expected payoff does not change.*

Under σ, the expected payoff of a player i_q, $q = q^* + 1, \ldots, Q$, is

$$\sum_{\Theta_i} u_{i_q}(f^*(\theta_i), \theta_{i_q}) P^i(\theta_i),$$

since player i_q's posterior is equal to his prior (by Claim 2), and the designer correctly decodes the type of player i under σ. Suppose that player i_q has received the message x_i, but forwards the message x_i' instead. Under σ_{-i_q}, the designer decodes $\theta_i = t_i$ if $(x_i' - y_i) \,\mathsf{mod}_{0,1} \in \Pi_i(t_i)$. Player i_q evaluates the probability of this event to be $\mathbb{P}((x_i' - Y_i) \,\mathsf{mod}_{0,1} \in \Pi_i(t_i)) = P^i(t_i)$ since $(x_i' - Y_i) \,\mathsf{mod}_{0,1}$ is uniformly distributed on $[0, 1]$ by part 2 of Lemma 1. Therefore, player i_q's expected payoff does not depend on the value of the message he forwards.

A similar argument applies to players j_r, $r = 1, \ldots, R$. Note that the distribution of the message x_i received by any player i_q, $q > q^*$, given player j_r's types and message y_i is the distribution of the random variable

$$X_i = \sum_{t_i=1}^{T_i} \mathbf{1}_{\{\theta_i = t_i\}} X_i^{t_i},$$

and since $X_i^{t_i}$ is uniformly distributed on $\Pi(t_i)$, X_i is uniformly distributed on $[0, 1]$. Suppose that player j_r has received the message y_i, but forwards the message y_i' instead.

Under σ_{-j_r}, the designer decodes $\theta_i = t_i$ if $(x_i - y_i') \,\mathsf{mod}_{0,1} \in \Pi_i(t_i)$. Player j_r evaluates the probability of this event to be $\mathbb{P}((X_i - y_i') \,\mathsf{mod}_{0,1} \in \Pi_i(t_i)) = P^i(t_i)$ since $(X_i - y_i') \,\mathsf{mod}_{0,1}$ is uniformly distributed on $[0,1]$ by part 2 of Lemma 1. The exact same argument applies to players i_q, $q < q^*$.

Lastly, player i has clearly no incentive to deviate from σ_i since f satisfies player i's incentive constraints.

Part II. Let f be a social choice function implementable on $\mathcal{N}^\star$ i.e., f is incentive compatible. To implement f on $\mathcal{N}$, we construct a mechanism by concatenating N sub-mechanisms as constructed in Part I. A player directly connected with the designer simply transmits his type to the designer along with all the messages he receives. Each player $i \in N \setminus D(0)$ uses the sub-mechanism constructed in Part I to securely transmit his type and forward the messages he receives on the appropriate paths. More precisely, if player i is on the path $\pi^1_{k(j)}$ (resp., $\pi^2_{k(j)}$) from player $k(j)$ to 0, he forwards the message about j's type to the next player l on the path $\pi^1_{k(j)}$ (resp., $\pi^2_{k(j)}$). The mechanism is constructed in such a way that player i cannot forward a message about j's type to a player other than player l.

Messages. For each player $j \in N \setminus D(0)$, fix two disjoint paths $\pi^1_{k(j)}$ and $\pi^2_{k(j)}$ from $k(j)$ to the designer. For each player i on $\pi^1_{k(j)}$ (resp., $\pi^2_{k(j)}$), we denote $\tau^1_j(i)$ (resp., $\tau^2_j(i)$) the successor of i on the path $\pi^1_{k(j)}$ (resp., $\pi^2_{k(j)}$). The messages sets are formally defined as follows.

- For all $i \in D(0)$, for all $l \in C(i) \setminus \{0\}$,

$$M_{il} = [0,1]^{|\{j : l = \tau^1_j(i)\} \cup \{j : l = \tau^2_j(i)\}|},$$

  and

$$M_{i0} = [0,1]^{|\{j : 0 = \tau^1_j(i)\} \cup \{j : 0 = \tau^2_j(i)\}|} \times \Theta_i.$$

- For all $i \in N \setminus D(0)$, for all $l \in C(i)$,

$$M_{il} = [0,1]^{|\{j : l = \tau^1_j(i)\} \cup \{j : l = \tau^2_j(i)\}|}.$$

- If $\{j : l = \tau^1_j(i)\} \cup \{j : l = \tau^2_j(i)\} = \emptyset$, then

$$[0,1]^{|\{j : l = \tau^1_j(i)\} \cup \{j : l = \tau^2_j(i)\}|} = \{\emptyset\}.$$

Allocation rule. Let us now define the allocation rule $g : \times_{i \in D(0)} M_{i0} \to X$. By construction of the message space, the designer receives exactly two messages $(\hat{x}_i, \hat{y}_i)$ in $[0,1] \times [0,1]$ about each player $i \in N \setminus D(0)$ and a unique message θ_i in Θ_i from each player $i \in D(0)$. If $(\hat{x}_i - \hat{y}_i) \,\mathsf{mod}_{0,1} \in \Pi_i(t_i)$, let $\hat{\theta}_i = t_i$, i.e., the designer decodes the type of player i to be t_i. For any profile of messages $(m_i)_{i \in D(0)}$, $g(m) = f((\theta_i)_{i \in D(0)}, (\hat{\theta}_i)_{i \in N \setminus D(0)})$.

Strategies. The strategies are as in Part I and we only provide an informal description. Firstly, if player i of type θ_i is directly connected to the designer, he directly sends θ_i to the designer and truthfully forwards all messages he has received. Secondly, suppose that player i is not directly connected to the designer. Since the network is weakly 2-connected, there exists player $k(i)$ such that $k(i)$ has two disjoint paths to the designer with i on one path. If $k(i) \neq i$, the strategies requires $k(i)$ to draw, independently of his type, an encoding key y_i (a realization of Y_i), to send it on both paths and to truthfully forward all messages he has received on the appropriate paths. Player $k(i)$ must also transfer his type to the designer. To do so, the strategy is identical to the one for player i, which we now describe (with player $k(i)$ receiving his encoding from player $k(k(i))$). Player i of type t_i draws $x_i^{t_i}$ (a realization of $X_i^{t_i}$) and sends the cypher-type $x_i = (x_i^{t_i} + y_i) \,\mathsf{mod}_{0,1}$ on the appropriate path from $k(i)$ to the designer. Moreover, he truthfully forwards all messages he has received on the appropriate paths. If $k(i) = i$, then regardless of his type, player i sends an encoding key y_i (a realization of Y_i) on one path and the encoded type on the other path. He also truthfully forwards all messages he has received on the appropriate paths. (Note that since the network is acyclic, the strategies are well-defined. In particular, there exists a player for which $k(i) = i$.)

The rest of the proof is completed as in Part I.

Now, we prove the "only if" part. The proof proceeds by contradiction. We assume that $\mathcal{N}$ is not weakly 2-connected and construct an environment with common independent belief and private values and an incentive compatible social choice function, which is not implementable on $\mathcal{N}$.

Denote $2C$ the set of players with at least two disjoint paths to the designer. Since $\mathcal{N}$ is not weakly two-connected, there exists a player $i \notin D(0)$ such that there is no player $k \in 2C$ with player i being on one of the two disjoint paths from k to the designer. This has the following consequences: (i) there exists a player $i^* \neq i$ such that all paths from player i to the designer go through i^*, and (ii) for each player k who has a path to player i, all paths from k to the designer go through i^*.

Part (i) is clear. To see why (ii) holds true, assume to the contrary that player k has a path to the designer, which does not go through player i^*. Hence, this path cannot go through player i because of (i). But then, player k has two paths to the designer, one that goes through players i and i^* and one that contains neither players i nor i^*, a contradiction. Note also that (i) implies that for each player on a path from player i to player i^*, all paths from this player to the designer go through player i^*.

An important implication is the following: All players on a path from player i to the designer and who play after player i^* observe player i^*'s message only.

Let us now construct the environment and the social choice function. Assume that all players but player i have a single type and that player i has two types θ_i and θ_i'. Let a and b be two alternatives. The utilities are as follows: $u_i(a, \theta_i) = u_{i^*}(a, \cdot) = 1$, $u_i(b, \theta_i) = u_{i^*}(b, \cdot) = 0$; $u_i(a, \theta_i') = 0$, $u_i(b, \theta_i') = 1$. All other players are indifferent (get a utility of 0) between a and b. Any other alternative gives a utility of -1 to players i and i^* regardless of their types. The common prior is the uniform distribution on the set of types. The social choice function is the dictatorial social choice function of player i.

We claim that for every mechanism on $\mathcal{N}$, there is no equilibrium that implements this social choice function. By contradiction, assume that there exists such an equilibrium σ. Fix a profile of messages $\bar{m}_{i^*} \in M_{D(i^*)}$ for player i^* in the support of $\mathbb{P}_{\theta_i,\sigma}$, i.e., this is a message compatible with θ_i and the equilibrium strategies. Consider the deviation σ_{i^*}' for player i^* which consists in playing $\sigma_{i^*}(\bar{m}_{i^*})$ regardless of his type and messages received.

By construction of the deviation, $\sigma_{i^*}(\bar{m}_{i^*})$ is compatible with the messages sent by players who have no path to player i^*, i.e.,

$$\operatorname{supp} \mathbb{P}_{\theta,(\sigma'_{i^*},\sigma_{-i^*})} \subseteq \operatorname{supp} \mathbb{P}_{\theta_i,\sigma} \quad \forall \theta \in \{\theta_i, \theta'_i\}.$$

Since the strategies are assumed to implement f, it follows that the outcome is almost surely a under the deviation, regardless of the type of player i. Since player i^* prefers a to any other alternative, this deviation is profitable for player i^*. □

Before going further, it is important to stress that the encoding technique used in the proof of Theorem 1 is tailored to environments with common independent beliefs and does not apply to more general environments (even with private values). Intuitively, consider two players j and j' on two disjoint paths from player $k(i)$ to the designer and suppose that they disagree about the likelihood of the two possible types θ_i^0 and θ_i^1 of player i. Now, if player i encrypts his type with an encoding key tailored to player j's prior beliefs, he neither affects the incentives of player j, nor reveals information about his type. However, player j' now evaluates the likelihood of the designer decoding the type of player i as either θ_i^0 or θ_i^1 according to *player j's beliefs*. Thus, the incentives of player j' to truthfully reveal his type have been altered. With general beliefs, different encoding techniques have to be used. This is the object of the next section.

4.2 Bad outcome

In concrete applications of the theory of mechanism design, players often hold different and correlated beliefs about states of the world either because they have received different signals (information) or on purely subjective grounds. For instance, in auction models with interdependent values, bidders often have different information about the value of the good for sale (e.g., mineral or oil rights). To handle these more general beliefs, we resort to a different encoding technique. Our new technique consists in coding the type of player such that no information is revealed to the other players and, if a player does not truthfully forward the messages he receives, the designer detects it with probability 1.

Theorem 2 *For all environments $\mathcal{E}$ with a bad outcome, $F_{\mathcal{N}}(\mathcal{E}) = F_{\mathcal{N}^\star}(\mathcal{E})$ if and only if $\mathcal{N}$ is weakly 2-connected.*

The intuition for Theorem 2 is as follows. Consider the network $\mathcal{N}_1$ of Section 2 and the dictatorial social choice function of player 2. With this communication network, player 2 has a *unique* communication path to the designer through player 1, who can "filter" any information received from player 2 to his own advantage. The idea is then to use player 3 to simultaneously detect a false report by player 1 and allow player 2 to transmit his type to the designer. More precisely, the strategies are the following. Player 3 sends a large number of encoding keys to player 2 and to the designer. Player 2 then select one of the keys at random and encrypts his type with this key. He then substitutes the selected key by the cypher-type and sends it to player 1 along with all the other keys. Player 1 truthfully forwards the message received. The designer can then detect a false report by comparing the two vectors of messages received from players 1 and 3. Namely, if player 1 truthfully forwards the message he receives, the two vectors should coincide but for one component. In such a case, the designer decodes the type of player 2 according to this component and implements the appropriate outcome. Otherwise, the designer implements the worst outcome. By construction, only player 2 knows the key selected to encrypt his type. Thus, any deviation by player 1 or 3 induces the worst outcome with arbitrarily high probability: this deters them from lying. A refinement of this technique makes it possible to detect a deviation with probability 1 and, thus, to only require the existence of a bad outcome. Essentially, the refinement consists in duplicating the above detection test an infinite number of times. To the best of our knowledge, this encoding technique is novel.

Proof of Theorem 2. (if) The proof is constructive and proceeds in two parts. In the first part, we design a "sub-mechanism" to implement any dictatorial social choice function. We first do the construction for an environment $\mathcal{E}$ with a worst outcome $\underline{a}$. We explain then how to extend our construction to an environment with a bad outcome. In the second part, we concatenate these "sub-mechanisms" to implement any incentive-compatible social choice function. We stress the first part; the second part being similar to the second part of the proof of Theorem 1. Without loss of

generality, we assume that Θ_i is a finite subset of the open interval $(0,1)$ for each player $i \in N$.

Part I with worst outcome. Assume that f selects a preferred alternative of player i at any state (θ_i, θ_{-i}), i.e., for any θ_i, define $f^*(\theta_i) \in \arg\max_a \sum_{\theta_{-i}} u_i(a, \theta_i, \theta_{-i}) P_i(\theta_{-i}|\theta_i)$ and let $f(\theta_i, \theta_{-i}) = f^*(\theta_i)$ for all θ_{-i}. If $i \in D(0)$, f is clearly implementable. Assume that $i \notin D(0)$. The problem is to design a mechanism and an equilibrium such that player i has an incentive to truthfully reveal his type and no other player has an incentive to manipulate the transmission of information from player i to the designer. Since the network is weakly 2-connected, there exist a player $k(i) = k$ and two disjoint paths $\pi_k^1 = (k = i_0, i_1, \ldots, i_Q, i_{Q+1} = 0)$ and $\pi_k^2 = (k = j_0, j_1, \ldots, j_R, j_{R+1} = 0)$ from player k to the designer such that $i = i_{q^*}$ for some $q^* = 0, \ldots, Q$. Assume $k \neq i$.

Messages. Let η be a large integer. Player k's message space is given by $M_{ki_1} = [0,1]^\eta = M_{kj_1}$, and $M_{kl} = \{\emptyset\}$ for all $l \in C(i) \setminus \{i_1, j_1\}$. The message space of any player i_q or j_r, $q = 1, \ldots, Q$, $r = 1, \ldots, R$, is $M_{i_q i_{q+1}} = M_{j_r j_{r+1}} = [0,1]^\eta$, and $M_{i_q l} = M_{j_r l'} = \{\emptyset\}$ for all $l \in C(i_q) \setminus \{i_{q+1}\}$, for all $l' \in C(j_r) \setminus \{j_{r+1}\}$. All other players have no messages.

Allocation rule. The designer receives two η-vectors of messages: $(\hat{x}_i^1, \ldots, \hat{x}_i^\eta)$ from i_Q and $(\hat{y}_i^1, \ldots, \hat{y}_i^\eta)$ from j_R. If the two vectors differ by a unique component $\hat{\eta}$, the designer decodes $\hat{\theta}_i = (\hat{x}_i^{\hat{\eta}} - \hat{y}_i^{\hat{\eta}}) \,\mathsf{mod}_{0,1}$ and implements $f^*(\hat{\theta}_i)$. Otherwise, the designer implements the worst outcome $\underline{a}$.

Strategies σ. Let $Y_i^1, \ldots, Y_i^\eta$ be η random variables, independently and uniformly distributed on $[0,1]$, and denote $y_i^1, \ldots, y_i^\eta$ a realization.

Player k. Regardless of his type, player k draws η encoding keys $(y_i^1, \ldots, y_i^\eta)$ and sends them to players i_1 and j_1.

Player $i = i_{q^*}$. Regardless of his type, player i draws uniformly at random an integer η^* in $\{1, \ldots, \eta\}$, and encodes his type θ_i with the encoding key $y_i^{\eta^*}$ to obtain the cypher-type $x_i = (\theta_i + y_i^{\eta^*}) \,\mathsf{mod}_{0,1}$. Player i then sends the vector $(y_i^1, \ldots, y_i^{\eta^*-1}, x_i, y_i^{\eta^*+1}, \ldots, y_i^\eta)$ to player i_{q^*+1}.

Players j_r, $r = 1, \ldots, R$, i_q, $q \neq q^*$. Regardless of his type, player j_r (resp., i_q, $q \neq q^*$) truthfully forwards the message he received to player j_{r+1} (resp., i_{q+1}).

(If $k = i$, the construction is almost identical. The main difference is that player i also performs the task of player k, that is, player i draws a large number of keys, selects one at random for coding and sends the vector of keys to j_1 and the vector containing the cypher-type to i_1.)

Now, we prove that the strategies define a Bayesian equilibrium which implements f.

Claim 1. *Under* σ*, the decoded type* $\hat{\theta}_i$ *coincides with the true type* θ_i*.*

Under σ, the designer receives $(y_i^1, \ldots, y_i^{\eta^*-1}, x_i, y_i^{\eta^*+1}, \ldots, y_i^\eta)$ from i_Q and $(y_i^1, \ldots, y_i^\eta)$ from j_R, with $x_i = (\theta_i + y_i^{\eta^*}) \,\mathsf{mod}_{0,1}$. These two vectors differ by at most one component. Furthermore, $x_i = y_i^{\eta^*}$ if and only if $\theta_i = 0$ or 1 which is ruled out by the assumption $\Theta_i \subset (0,1)$. Thus, the designer decodes $\hat{\theta}_i = (x_i - y_i^{\eta^*}) \,\mathsf{mod}_{0,1} = \theta_i$.

Claim 2. *The message received by any active player conveys no information on the type of player* i*.*

Clearly, if $k \neq i$, the message of player k does not convey information about player i's type. The statement is thus clear for players i_q, $q = 1, \ldots, q^* - 1$ and players j_r, $r = 1, \ldots, R$. Players i_q, $q > q^*$ receive vectors of random variables which are independent, uniformly distributed and independent from θ_i by application of Lemma 1. If $k = i$, the same arguments apply.

Claim 3. *If a player other than player* i *forwards a false message, the worst outcome is implemented with probability at least* $1 - 1/\eta$*.*

Suppose that player $i_q \in \pi_k^1$ ($q \neq q^*, 0$) forwards a message $\tilde{y} = (\tilde{y}_i^1, \ldots, \tilde{y}_i^\eta)$ different from the one received $y = (y_i^1, \ldots, y_i^\eta)$, and let $y = (y_i^1, \ldots, y_i^\eta)$ be the message received by the designer from player j_R. The deviation of player i_q is not detected if and only if y differs from $\hat{y}$ in the component η^*. However, the choice of the encryption key $y_i^{\eta^*}$ by player i is a random draw independent of his type and messages received, and the realization is only known to him. It follows that a deviation is not detected with probability at most $1/\eta$. The same argument applies to any player $j_r \in \pi_k^2$, $r > 0$ and when player $k \neq i$ sends two different messages $y = (y_i^1, \ldots, y_i^\eta)$ and $\tilde{y} = (\tilde{y}_i^1, \ldots, \tilde{y}_i^\eta)$

to players i_1 and j_1.

Claim 4. *No active player has an incentive to deviate.*

Let $j \neq i$ be an active player. The expected payoff of j under σ is:

$$\sum_{\theta_{-j}} u_j(f^*(\theta_i), \theta_j, \theta_{-j}) P_j(\theta_{-j} \mid \theta_j) := C.$$

If player j transmits a false message, his expected payoff is at most,

$$\frac{1}{\eta} L + (1 - \frac{1}{\eta}) \sum_{\theta_{-j}} u_j(\underline{a}, \theta_j, \theta_{-j}) P_j(\theta_{-j} \mid \theta_j) := D,$$

where L is an upper bound on player j's payoff. We have

$$C - D = \frac{1}{\eta}(C - L) + (1 - \frac{1}{\eta}) \sum_{\theta_{-j}} (u_j(f^*(\theta_i), \theta_j, \theta_{-j}) - u_j(\underline{a}, \theta_j, \theta_{-j})) P_j(\theta_{-j} \mid \theta_j).$$

Letting $\varepsilon = \min_{a \neq \underline{a}, \theta} \{u_i(a, \theta) - u_i(\underline{a}, \theta)\} > 0$, we find:

$$C - D \geq \frac{1}{\eta}(C - L) + (1 - \frac{1}{\eta})\varepsilon.$$

Thus, for η large enough, the right-hand side is non-negative and player j has no incentive to deviate from σ.

Lastly, player i has clearly no incentive to deviate from σ_i since f satisfies player i's incentive constraints.

Part I with bad outcome. The construction is almost identical to the case with a worst outcome. The difficulty here is to devise a mechanism such that deviations are detected with probability 1. The intuition is as follows. Form the above, for each η, we can devise a test such that any deviation is detected with probability at least $1 - 1/\eta$. We may thus ask the players to pass *all such tests*.[8] There are several ways to construct this, and we provide a relatively simple one. Throughout, assume that $k \neq i$. (A similar construction applies if $k = i$.)

Player $k(i)$. Player $k(i)$ draws two independent infinite sequences $(X_\eta^H, X_\eta^T)_{\eta \geq 1}$ of independently and identically (i.i.d.) distributed random variables, with uniform distribution on $[0, 1]$. This pair of sequences is sent on the two paths $\pi^1_{k(i)}$, $\pi^2_{k(i)}$.

[8] We gratefully thank Sylvain Sorin for suggesting this argument.

Player i. Independently of his type and of the message he receives, player i draws an infinite sequence of i.i.d. fair coins $c_\eta \in \{H, T\}$. Define $(Y_\eta^H, Y_\eta^T)_{\eta \geq 1}$ as $(Y_\eta^H, Y_\eta^T) = ((X_\eta^H + \theta_i) \,\mathsf{mod}_{0,1}, X_\eta^T)$ if $c_\eta = H$, and $(Y_\eta^H, Y_\eta^T) = (X_\eta^H, (X_\eta^T + \theta_i) \,\mathsf{mod}_{0,1})$ if $c_\eta = T$. In words, for each η, player i chooses according to the toss of a fair coin whether to encode his type θ_i with X_η^H or with X_η^T. Player i then sends the pair of sequences $(Y_\eta^H, Y_\eta^T)_{\eta \geq 1}$ to player i_{q^*+1}.

Other players. The other active players should faithfully forward their messages along the path π_k^1, π_k^2.

The designer. The designer receives two pairs of sequences $(x_\eta^H, x_\eta^T)_{\eta \geq 1}$ and $(y_\eta^H, y_\eta^T)_{\eta \geq 1}$. If for each η, it holds true that $x_\eta^H = \eta_n^H$ and $x_\eta^T \neq y_\eta^T$, or $x_\eta^H \neq y_\eta^H$ and $x_\eta^T = y_\eta^T$, the designer concludes that phase 1 of the test succeeds. Then, he computes $\hat{\theta}_i^\eta = (y_\eta^T - x_\eta^T) \,\mathsf{mod}_{0,1}$ if $x_\eta^T \neq y_\eta^T$, and $\hat{\theta}_i^\eta = (y_\eta^H - x_\eta^H) \,\mathsf{mod}_{0,1}$ if $x_\eta^H \neq y_\eta^H$. If all $\hat{\theta}_i^n$ have the same value $\hat{\theta}_i$, the designer concludes that phase 2 of the test succeeds, and implements $f^*(\hat{\theta}_i)$. If the test does not succeed, either in phase 1 or in phase 2, the designer implements the bad outcome.

Under these strategies, the decoded type clearly coincides with the true type. It is also clear that no player gets information about the type of player i. The sequence of coins being privately known to player i, each other active player only observes sequences of i.i.d. uniformly distributed variables. Now, we claim that no profitable deviation exists from these strategies. Indeed, if some active player $j \neq i$ modifies the sequence, to pass the test in phase 2 he must modify an entry of the double sequence for each η. But then, to succeed in phase 1, he should modify only the component selected by player i. Consequently, the probability of passing the test while changing the message is at most the probability of guessing correctly an infinite sequence of fair coins, which is 0. Any deviation will thus bring the bad outcome with probability 1.

Part II. The implementation of any social choice function f follows from Part I and a similar construction as the one adopted in Part II of the proof of Theorem 1.

(only if). The proof is similar to the "only if" part of Theorem 1 and is left to the reader. □

An essential feature of Theorem 2 is the possibility to punish a detected deviation with a bad outcome. It is worth stressing, however, that our definition of a bad outcome is stronger than necessary since it does not depend on the social choice function we aim to implement. It would be enough to find an outcome worse than any outcome in the range of the social choice function. Moreover, in environments with quasi-linear preferences (e.g., team problems), the existence of a bad outcome is natural: the designer can always impose a large fine on the players.

If such a bad outcome does not exist, the main difficulty for the designer is the choice of an appropriate alternative to implement whenever a false report is detected. A characterization of networks that allow to implement all incentive compatible social choice functions in all environments is left as an open problem. Yet, we provide sufficient conditions in the next section.

5 Extensions and Robustness

This section discusses various aspects of our problem and offers some generalizations.

5.1 All environments

We give sufficient conditions on the network for implementing all incentive compatible social choice functions, regardless of the environments.

Theorem 3 *If the communication network $\mathcal{N}$ is 3-connected, then $F_{\mathcal{N}}(\mathcal{E}) = F_{\mathcal{N}^\star}(\mathcal{E})$ for all environments $\mathcal{E}$.*

The intuition is simple. Since the network is 3-connected, for each player $i \in N \setminus D(0)$, there exist three pairs of disjoint paths from player $i \in N \setminus D(0)$ to the designer. On each pair of paths, we can replicate the construction of Theorem 2 so as to detect any false report of messages with probability 1 and to guarantee that no information about player i is revealed. A simple "majority" argument then guarantees that no player has an incentive to lie. More precisely, for any unilateral deviation of player $j \neq i$, there is a pair of path from player i to the designer to which player j does

not belong, and no deviation is detected on that pair of path. The designer can then correctly decode the type of player i according to the messages received from that pair of paths.[9]

Three further remarks are worth making. Firstly, with this construction, we need each player to both draw encoding keys (two infinite sequences) and to encode their type with these encoding keys. Consequently, this technique cannot be used on weakly 2-connected networks as keys might have to come from other players. Secondly, simpler encoding technique with *authentication tests* can be used. In particular, these encoding techniques do not require infinite sequences of encoding keys. These simpler authentication tests originally appeared in the work of Rabin and Ben-Or (1989) and Franklin and Wright (2000). We refer the reader to the working version of this paper. Thirdly, although there are three paths of communication from each player to the designer, a classic majority argument does not work. A player must not truthfully reveal his private information on the three paths. Simply, if a player were to do so, he would change the incentives of other players to truthfully reveal their own private information.

5.2 Pure equilibria

With the notable exception of Serrano and Vohra (2007), the literature on implementation in Bayesian environments has entirely focused on the implementation of social choice functions in *pure* equilibria (see Jackson (2001) for an excellent survey). By contrast, the recourse to equilibria in mixed strategies is essential for our results. In effect, to transmit securely their types to the designer, it is essential for the players to encrypt their types with privately and randomly generated keys (mixing). Although the use of randomly generated keys seems natural in our context and, indeed, used in daily life (internet banking, online shopping, etc.), we might legitimately wonder whether similar results hold in environments where only pure equilibria are considered. The next theorem states that the set of social choice functions partially implementable on $\mathcal{N}$ in pure equilibria coincides with the set of incentive compatible social choice

[9] A complete proof is available upon request.

functions, irrespective of the utility functions, if and only if every player is directly connected to the designer. There is a sharp divide between implementation in pure equilibria and mixed equilibria. Denote $F_{\mathcal{N}}^{pure}(\mathcal{E})$ the set of social choice functions (partially) implementable on $\mathcal{N}$ in pure equilibria when the environment is $\mathcal{E}$.

Theorem 4 *$F_{\mathcal{N}}^{pure}(\mathcal{E}) = F_{\mathcal{N}^\star}^{pure}(\mathcal{E})$ for all environments $\mathcal{E}$ if and only if each player is directly connected to the designer i.e., $D(0) = N$.*

The intuition is simple. If player i is not directly connected to the designer and if the social choice function depends on his type, then he must send an informative message to at least one other player, say player j. Given his updated beliefs, player j might then have no incentive to truthfully report his own private information. This reasoning is valid regardless of how many disjoint paths there are from player i to the designer. The proof is given in Appendix.

While intuitive, Theorem 4 has remarkable implications for the topology of communication networks and implementation in pure equilibria. All but one player, say player 1, might be directly connected to the designer, player 1 might have $n-1$ disjoint paths of communication to the designer and yet, there exist incentive compatible social choice functions, which are not implementable on that network in pure equilibria. While some theorists might feel uncomfortable with equilibria in mixed strategies, the mixing through encoding techniques, as considered in this paper, seems quite natural.

5.3 Continuous type spaces

Many applications of mechanism design theory e.g., contract theory and auction theory, assume a continuous type space. While we have casted our results in environments with finite type spaces, they naturally extend to continuous type spaces.[10]

We now explain how to extend Theorem 1. To ease the exposition, we assume that there are two disjoint paths from player i to the designer, so that $k(i) = i$. If $k(i) \neq i$, the extension is similar. Let each player's type space Θ_i be a subset of $[0, 1]$ and types

[10] Appropriate measurability and integrability assumptions have to be made.

be independently distributed. Let P be the common prior and G_i be the cumulative distribution function of the marginal P^i over Θ_i.

Assume that G_i is continuous. The key observation to make is that $G_i(\theta_i)$ is uniformly distributed on $[0,1]$ and, therefore, might be used as a "pseudo-type." More precisely, to transmit securely his type to the designer, player i draws a random variable Y_i, uniformly distributed on $[0,1]$, and sends the realization y_i on one path. If his type is θ_i, he computes the cypher-type $x_i = (G_i(\theta_i) + y_i) \,\mathsf{mod}_{0,1}$, and sends it on the other path. Players on the two disjoint paths from player i to the designer are required to truthfully forward the message received. The designer thus receives two messages $\hat{x}_i$ and $\hat{y}_i$ and decodes the type of player i to be $G_i^{-1}((\hat{x}_i - \hat{y}_i) \,\mathsf{mod}_{0,1})$.

As in the proof of Theorem 1, it is clear that messages convey no additional information about the type of player i. Furthermore, no active player has an incentive to forward a message other than the one received. Indeed, for any $\hat{x}_i$, the law of $(\hat{x}_i - Y_i) \,\mathsf{mod}_{0,1}$ is the uniform distribution on $[0,1]$ and, therefore, the law of $G_i^{-1}((\hat{x}_i - Y_i) \,\mathsf{mod}_{0,1})$ is P^i. Hence, no active player has an incentive to deviate.

If G_i has atoms, the strategies are about the same. Suppose that θ_i^* is an atom of G_i, i.e., $\lim_{\theta_i \uparrow \theta_i^*} G_i(\theta_i) := G_i^-(\theta_i^*) < G_i^+(\theta_i^*) =: \lim_{\theta_i \downarrow \theta_i^*} G_i(\theta_i)$. Let $\hat{G}_i(\theta_i^*)$ be the realization of a uniform draw on $[G_i^-(\theta_i^*), G_i^+(\theta_i^*)]$. Let $\hat{G}_i(\theta_i) = G_i(\theta_i)$ if θ_i is not an atom. The strategies then require player i to send y_i on one path and $x_i = (\hat{G}(\theta_i) + y_i) \,\mathsf{mod}_{0,1}$ on the other path. All other players should truthfully forward their messages. The designer decodes the type of player i according to $\hat{G}_i^{-1}$ with $\hat{G}_i^{-1}(x_i - y_i) := \inf\{\theta_i : \hat{G}_i(\theta_i) \geq (x_i - y_i) \,\mathsf{mod}_{0,1}\}$. We can readily check that with these strategies, player i securely transmits his type to the designer (provided that the network is 2-connected).

As for Theorem 2, it extends to continuous type spaces without difficulty. In sum, all our constructions naturally extend to the continuous case.

5.4 Finite message space

While consistent with mechanism design theory, the use of continuous message spaces seems unappealing from a computer science perspective. Yet, if prior beliefs are rational numbers, then Theorem 1 extends perfectly. It is enough to assume that the (finite)

types spaces are subsets of $\{1, \ldots, n\}$ for n large enough and to draw encoding keys uniformly in $\{1, \ldots, n\}$. Addition is then to be understood modulo n.

The method of Theorem 2 also perfectly extends to finite message spaces without any assumption on priors, provided that there exists a worst outcome. We conjecture that the use of continuous message spaces cannot be dispensed with if there is a bad outcome but no worst outcome.

5.5 Active designer

A salient feature of our model is that the designer cannot communicate with the players. However, in some situations, it is natural to assume that the designer can communicate with the players. For instance, a CEO has the possibility to communicate with his employees either publicly or privately.

If the designer can *directly* communicate with a subset of players $C(0) \subseteq N$, our main results have natural counterparts. Assume that for each player $i \in N \setminus D(0)$, either i is weakly 2-connected to the designer or there exists a path from the designer to player i *and* a path from player i to the designer which have no common vertex (except for 0 and i). See Figure 5 for an example.

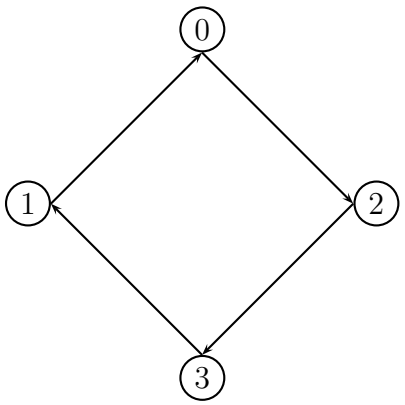

Figure 5: Communication network $\mathcal{N}$

Then, $F_{\mathcal{N}}(\mathcal{E})$ is the set of all incentive compatible social choice functions for all environments with a worst outcome or independent common beliefs and private values. The idea is simply to let the designer play the role of player $k(i)$ in the proof of Theorem

1 or Theorem 2 and to let him draw the encoding keys. To get some intuition, let us consider the secure transmission of player 3's private information in the network of Figure 5, when there is a worst outcome. The designer draws a large number of encoding keys and sends them to player 2. Player 2 forwards the encoding keys to player 3, who selects one key at random and uses it to encode his type. He then sends the unused keys and the encoded type to player 1, who should forward this message to the designer. Finally, the designer compares the vector of keys he sent to player 2 and the vector of keys he receives from 1, and decodes the type of player 3 accordingly. As in the proof of Theorem 2, any deviation by player 1 or player 2 is detected with arbitrarily large probability, no information about player 3's type is revealed and the designer correctly learns the type of player 3. It is an open problem to find necessary and sufficient conditions.

Finally, let us mention that the assumption of an active designer is important in generalized principal-agents models (Myerson (1982)), where players also have to take an action, thus creating a moral hazard problem in addition to the adverse selection problem. In such models, the designer has to "securely recommend" an action to each player. We believe that our results extend to this more general framework. Indeed, if the designer has two disjoint paths of communication to each player, then he can follow our protocols to privately and reliably make a recommendation to each player. A careful analysis of this issue awaits future research.

6 Conclusion

This paper completely characterizes the communication networks for which, in any environments (utilities and beliefs) with either common independent priors and private values, or with a worst outcome, every incentive compatible social choice function is (partially) implementable. We show that any weakly 2-connected communication network can replicate the incentive properties of the direct revelation mechanism. Importantly, our constructions couple encryption techniques together with incentives to secure the transmission of each player's private information to the designer.

To conclude, we believe that this paper delineates promising avenues for future

research. An interesting open problem is the characterization of networks which are "equivalent" to the star network $\mathcal{N}^\star$ for *all* environments. We already know that the 3-connectivity of the network is sufficient, but finding necessary and sufficient conditions remains an open issue. Another interesting open issue is to consider partially known networks e.g., a model where players only known their neighbors. Other open issues include the problem of full implementation or virtual implementation on communication networks.[11]

7 Appendix

7.1 Timing Structure

Lemma 2 *Let $\mathcal{N}$ be a network which has no cycle and such that every player is connected to the designer. The communication rule is such that:* each player sends his messages when he has received a message from all her neighbors. *Then, there exists an integer T and a timing function $t : N \to \{1, \ldots, T\}$ such that $t(i)$ is the stage at which player i sends her messages. Moreover, $ij \in \mathcal{N} \Rightarrow t(i) < t(j)$.*

Proof Let $V_1 = \{i : D(i) = \emptyset\}$. This set is non-empty: choose a path with maximal length in $\mathcal{N}$. Obviously the starting point of this path is in V_1. Note that for every player j, there exists a path from some player in V_1 to j: choose a path with maximal length among the paths with end-point j.

If $V_1 = N$, then $\mathcal{N} = \mathcal{N}^*$ and the proof is complete. Otherwise, let $V_2 = \{i : i \notin V_1 \text{ and } D(i) \subseteq V_1\}$.

Claim 1 *If $V_1 \neq N$, V_2 is non-empty.*

Proof. Define $W_1 = \cup_{i \in V_1} C(i)$. By construction, if j is in W_1, $D(j)$ is non-empty and therefore $j \notin V_1$. Consider then a path π of maximal length among the paths from a player in W_1 to the designer. Let j be the starting point of this path, we claim that j is in V_2. Otherwise, there exists $k \in D(j)$, $k \notin V_1$. There exists then a path from

[11] Renou (2008) is a first attempt at characterizing the social choice correspondences fully implementable in Nash equilibria on communication networks.

some point m in V_1 to k, denoted $\tau = m \to l \to \cdots k \to j$. Then l is in W_1 and $\tau\pi$ contradicts the maximality of π. •

If $V_1 \cup V_2 = N$, the construction ends. Otherwise, we let

$$V_3 = \{i : i \notin V_1 \cup V_2 \text{ and } D(i) \subseteq V_1 \cup V_2\}.$$

We continue this construction by induction. Assume that for some $k \geq 2$, the set V_s has been defined, $s \leq k$. If $\cup_{s\leq k} V_s = N$, the construction ends. Otherwise we let,

$$V_{k+1} = \{i : i \notin \cup_{s\leq k} V_s \text{ and } D(i) \subseteq \cup_{s\leq k} V_s\}.$$

Claim 2 *If $\cup_{s\leq k} V_s \neq N$, V_{k+1} is non-empty.*

Proof. Let $W_{k+1} = \{j \notin \cup_{s\leq k} V_s : \exists i \in \cup_{s\leq k} V_s,\, j \in C(i)\}$. Since $\cup_{s\leq k} V_s \neq N$, W_{k+1} is non-empty. Consider then a path π of maximal length among the paths from a player in W_{k+1} to the designer. The starting point j of this path is in V_{k+1}. Otherwise, there exists $k \in D(j)$, $k \notin \cup_{s\leq k} V_s$. There exists then a path from some point m in $\cup_{s\leq k} V_s$ to k. The follower of m on this path is in W_{k+1} and this contradicts the maximality of π. •

The sequence $(\cup_{s\leq k} V_s)_k$ is a weakly increasing sequence of sets, and is strictly increasing as long as $\cup_{s\leq k} V_s \neq N$. Since N is finite, there exists k such that $\cup_{s\leq k} V_s = N$. The timing function is then defined as $t(i) = s$ if $i \in V_s$. □

7.2 Proof of Theorem 4

The "if" part being clear, we prove the "only if" part. Assume that there exists a player $i \notin D(0)$. We construct a profile of utility and an incentive-compatible social choice function $f : \Theta \to X$, which is not implementable on $\mathcal{N}$. The main feature of our construction is that when player j on a path from player i to the designer learns the type of player i, he has an incentive to misreport her own type.

Up to a relabeling of players, assume that player $1 \notin D(0)$ and $D(1) = \emptyset$, i.e., player 1 receives no messages and thus sends his messages at time 1, $t(1) = 1$.

Fix two alternatives a and b, and consider a social choice function with range $\{a, b\}$: $f : \times_{i=1}^{n} \Theta_i \to \{a, b\}$. With each type θ_i of player i, we associate a number in $\{0, 1\}$,

i.e., we fix an onto mapping $\varphi_i : \Theta_i \to \{0, 1\}$. The social choice function we construct depends on types through these numbers only. Moreover, for each player i, there exists a unique $\theta_i^0 \in \Theta_i$ such that $\varphi_i(\theta_i^0) = 0$. Given a type profile $(\theta_1, \theta_2, \ldots, \theta_n)$, we denote $S = \sum_{j=2}^{n} \varphi_j(\theta_j)$ and $S_{-i} = S - \varphi_i(\theta_i)$. For convenience of language, we call $\varphi_i(\theta_i)$ the pseudo-type of player i.

The social choice function. We define the following social choice function:

$$f(\theta_1, \theta_{-1}) = \mathbf{1}_{\{\varphi_1(\theta_1)=0\}} \left[a\mathbf{1}_{\{S\leq\alpha\}} + b\mathbf{1}_{\{S>\alpha\}}\right] + \mathbf{1}_{\{\varphi_1(\theta_1)=1\}} \left[b\mathbf{1}_{\{S\leq\alpha\}} + a\mathbf{1}_{\{S>\alpha\}}\right],$$

where α is a fixed integer and $\mathbf{1}_E$ is the indicator function on the event E. In words, when $\varphi_1(\theta_1) = 0$, f chooses a if a large proportion of players $i = 2, \ldots, n$ are of pseudo-type 0. When $\varphi_1(\theta_1) = 1$, f chooses a if a small proportion of players are pseudo-type 0.

Next, we show that for a suitable choice of α, and for a class of utility functions, f is incentive compatible.

The utility functions. The utilities are as follows. Regardless of his type, player 1 is indifferent between a and b.

Player $i = 2, \ldots, n$ prefers a when he is of pseudo-type 0 and b when he is of pseudo-type 1. Further, his utility depends on his type and on the pseudo-type of player 1. The utility function is represented below:

	a	b
$\theta_i : \varphi_i(\theta_i) = 0$	$t_i(\theta_i)$	0
$\theta_i : \varphi_i(\theta_i) = 1$	0	$u_i(\theta_i)$

$\theta_1 : \varphi_1(\theta_1) = 0$

	a	b
$\theta_i : \varphi_i(\theta_i) = 0$	$v_i(\theta_i)$	0
$\theta_i : \varphi_i(\theta_i) = 1$	0	$w_i(\theta_i)$

$\theta_1 : \varphi_1(\theta_1) = 1$

where for each θ_i, $t_i(\theta_i), u_i(\theta_i), v_i(\theta_i), w_i(\theta_i)$ are positive numbers. We first show that f is incentive compatible and, therefore, implementable on $\mathcal{N}^\star$ in pure strategies.

Claim 1. *For $\alpha = n - 2$ and suitable choices of $(t_i(\theta_i), u_i(\theta_i), v_i(\theta_i), w_i(\theta_i))_{i=2}^n$, f is incentive compatible.*

Consider the incentive constraints of player $i = 2, \ldots, n$. Since f depends on pseudo-types only, the incentive constraints reduce to the incentive constraints over

pseudo-types. If player i is of type θ_i such that $\varphi_i(\theta_i) = 0$, his expected payoff of announcing 0 is

$$t_i(\theta_i)P_i(S_{-i} \leq \alpha, \varphi_1(\theta_1) = 0 \mid \theta_i) + v_i(\theta_i)P_i(S_{-i} > \alpha, \varphi_1(\theta_1) = 1 \mid \theta_i)$$

If he announces 1, his expected utility is:

$$t_i(\theta_i)P_i(S_{-i} + 1 \leq \alpha, \varphi_1(\theta_1) = 0 \mid \theta_i) + v_i(\theta_i)P_i(S_{-i} + 1 > \alpha, \varphi_1(\theta_1) = 1 \mid \theta_i)$$

The associated incentive constraint says that the former is no less than the latter. This amounts to:

$$t_i(\theta_i)P_i(S_{-i} = \alpha, \varphi_1(\theta_1) = 0 \mid \theta_i) \geq v_i(\theta_i)P_i(S_{-i} = \alpha, \varphi_1(\theta_1) = 1 \mid \theta_i). \tag{1}$$

Because of the full-support assumption, both sides are positive and for each θ_i such that $\varphi_i(\theta_i) = 0$, one can find $(t_i(\theta_i), v_i(\theta_i))$ such that (1) holds.

Similarly, if player i is of type θ_i such that $\varphi_i(\theta_i) = 1$, his expected payoff of announcing 1 is:

$$u_i(\theta_i)P_i(S_{-i} + 1 > \alpha, \varphi_1(\theta_1) = 0 \mid \theta_i) + w_i(\theta_i)P_i(S_{-i} + 1 \leq \alpha, , \varphi_1(\theta_1) = 1 \mid \theta_i)$$

If he announces 0, his expected utility is:

$$u_i(\theta_i)P_i(S_{-i} > \alpha, \varphi_1(\theta_1) = 0 \mid \theta_i) + w_i(\theta_i)P_i(S_{-i} + 1 \leq \alpha, \varphi_1(\theta_1) = 1 \mid \theta_i)$$

The associated incentive constraint amounts to:

$$u_i(\theta_i)P_i(S_{-i} = \alpha, \varphi_1(\theta_1) = 0 \mid \theta_i) \geq w_i(\theta_i)P_i(S_{-i} = \alpha, \varphi_1(\theta_1) = 1 \mid \theta_i). \tag{2}$$

Both sides are positive and for each θ_i such that $\varphi_i(\theta_i) = 1$, one can find $(u_i(\theta_i), w_i(\theta_i))$ such that (2) holds.

To complete the proof of Theorem 4, we now show that the social choice function f is not partially implementable on $\mathcal{N}$.

Assume by contradiction that there exists a mechanism (M, g) on $\mathcal{N}$ and a pure strategy Bayesian-Nash equilibrium s of the induced game such that $f = g \circ s$.

Since player 1 receives no messages ($D(1) = \emptyset$), the strategy of player 1 of type θ_1 is a tuple of messages $s_1(\theta_1) = (m_{1j}(\theta_1))_{j \in C(1)}$, where $m_{1j}(\theta_1)$ is the message that player 1 sends to player $j \in C(1)$ when he is of type θ_1.

Note that, for every type profile of the other players, the pseudo-type of player 1 determines the alternative chosen by f: $\forall \theta_{-1}$,

$$f(0, (\varphi_j(\theta_j))_{j \neq 1}) \neq f(1, (\varphi_j(\theta_j))_{j \neq 1}).$$

It follows that the tuple of messages sent by player 1 of type θ_1 s.t. $\varphi_1(\theta_1) = 0$ is different from the tuple of messages sent by player 1 of type θ_1 s.t. $\varphi_1(\theta_1) = 1$. Recall that there is only one type θ_1^0 of player 1 such that $\varphi_1(\theta_1) = 0$. We thus have,

$$\forall \theta_1 \text{ s.t. } \varphi_1(\theta_1) = 1, (m_{1j}(\theta_1^0))_{j \in C(1)} \neq (m_{1j}(\theta_1))_{j \in C(1)}$$

and therefore, for each θ_1 such that $\varphi_1(\theta_1) = 1$, there exists a player $i \in C(1)$ for whom $m_{1i}(\theta_1^0) \neq m_{1i}(\theta_1)$. We conclude that, for each $\theta_1 \neq \theta_1^0$, when player 1 is of type θ_1, there exists a player $i \in C(1)$ who learns from the messages that player 1's type is not θ_1^0. In particular, player i learns that the pseudo-type of player 1 is 1. We claim that this player has an incentive to deviate after receiving a message different from $m_{1i}(\theta_1^0)$.

Claim 2. *Player $i \in C(1)$, of type θ_i^0, receiving a message $m_{1i} \neq m_{1i}(\theta_1^0)$ from player 1, has an incentive to deviate from s.*

Proof. Let us fix θ_1^* such that $\varphi_1(\theta_1^*) = 1$ and a player $i \in C(1)$ of type θ_i^0, such that $m_{1i}(\theta_1^*) \neq m_{1i}(\theta_1^0)$. Consider also a profile of types $(\theta_k^*)_{k \neq 1, k \neq i}$ such that for each k, $\varphi_k(\theta_k^*) = 1$. For this type profile, $S_{-i} = n - 2$. Since $f = g \circ s$, if player i announces 0, i.e. plays what s recommends for type θ_i^0, then $S = n - 2$ and y is chosen. If player i announces 1, i.e. plays what s recommends for a type $\theta_i \neq \theta_i^0$, then $S = n - 1$ and x is chosen. Thus, if player i *knew* that the pseudo-type is 1 for every other player, he would have a clear incentive to play as if he were *not* of type θ_i^0.

Let m_i^* be the tuple of messages received by player i (under s) when the types of the other players are $(\theta_1^*, (\theta_k^*)_{k \neq 1, k \neq i})$. This tuple of messages occurs with positive probability. From the above discussion, player i deduces the pseudo-type of player 1

from m_i^*:

$$P_i(\varphi_1(\theta_1) = 1 \mid m_i^*, \theta_i^0) = 1.$$

Further, player i attributes a positive posterior probability to $(\theta_1^*, (\theta_k^*)_{k\neq 1, k\neq i})$:

$$P_i(\theta_1^*, (\theta_k^*)_{k\neq 1,k\neq i} \mid m_i^*, \theta_i^0) > 0.$$

We have thus exhibited a situation (i.e. messages, or an information set in the extensive game) where player i of type θ_i^0 knows that $\varphi_1(\theta_1) = 1$ and influences the selected alternative with positive probability. He faces thus the same kind of incentive problem as in the direct mechanism, except for the beliefs (priors are replaced by posteriors).

The expected utility of player i of type θ_i^0, conditional on m_i^* and if plays according to $s_i(\theta_0^i)$ is:

$$v_i(\theta_0^i)P(S_{-i} > n-2 \mid m_i^*, \theta_i^0) := \underline{v}.$$

If he "announces" 1, that is if he plays according to $s_i(\theta_i)$ with $\varphi_i(\theta_i) = 1$, his expected utility is:

$$v_i(\theta_0^i)P(S_{-i} + 1 > n-2 \mid m_i^*, \theta_i^0) := \overline{v}.$$

One has,

$$\overline{v} - \underline{v} = v_i(\theta_0^i)P(S_{-i} = n-2 \mid m_i^*, \theta_i^0) \geq v_i(\theta_0^i)P_i(\theta_1^*, (\theta_k^*)_{k\neq 1,k\neq i} \mid m_i^*, \theta_i^0) > 0.$$

This gives the desired contradiction. □

References

[1] Partha Dasgupta, Peter Hammond and Eric Maskin, The Implementation of Social Choice Rules: Some General Results on Incentive Compatibility, The Review of Economic Studies, 1979, 46, pp. 185-216

[2] Yvo Desmedt and Yongge Wang, Perfectly Secure Message Transmission Revisited, Lecture Notes in Computer Science, Advances in Cryptology EUROCRYPT 2002, 2002, Volume 2332/2002, pp. 502-517

[3] Danny Dolev, Cynthia Dwork, Orli Waarts and Moti Yung, Perfectly Secure Message Transmission, Journal of the ACM, 1993, 40, pp. 17–47

[4] Matthew K. Franklin and Rebecca N. Wright, Secure Communication in Minimal Connectivity Models, Journal of Cryptology, 2000, 13, pp. 9-30

[5] Allan Gibbard, Manipulation of Voting Schemes: A General Result, Econometrica, 1973, 41, pp. 587-601

[6] Milton Harris and Robert M. Townsend, Resource Allocation Under Asymmetric Information, Econometrica, 1981, 49, pp. 33-64

[7] Matthew O. Jackson, Bayesian Implementation, Econometrica, 1991, 59, pp. 461-477

[8] Matthew O. Jackson, A Crash Course in Implementation Theory, Social Choice and Welfare, 2001, 18, pp. 655-708

[9] Vijay Krishna, Auction Theory, Academic Press, 2002

[10] Eric Maskin and Tomas Sjöström, Implementation Theory, Chapter 5 in Handbook of Social Choice and Welfare, Volume 1. Eds. K.J Arrow, A.K. Sen, and K. Suzumura, Elsevier 2002

[11] Dov Monderer and Moshe Tennenholtz, Distributed Games: From Mechanisms to Protocols, Sixteenth National Conference on Artificial Intelligence, 1999, pp. 32–37

[12] Dilip Mookherjee, Decentralization, Hierarchies and Incentives: A Mechanism Design Perspective, Journal of Economic Literature, 2006, XLIV, pp. 367-390

[13] Roger B. Myerson, Incentive Compatibility and the Bargaining Problem, Econometrica, 1979, 47, pp. 61-73

[14] Roger B. Myerson, Optimal Coordination Mechanisms in Generalized Principal-Agent Problems, Journal of Mathematical Economics, 1982, 10, pp. 67-81

[15] Noam Nisan and Ilya Segal, The Communication Complexity of Efficient Allocation and Supporting Prices, Journal of Economic Theory, 2006, 129, pp. 192-224

[16] Noam Nisan, Tim Roughgarden, Eva Tardos, and Vijay V. Vazirani (Eds.), Algorithmic Game Theory, 2007, Cambridge University Press

[17] T. Rabin and Michael Ben-Or, Verifiable Secret Sharing and Multiparty Protocols with Honest Majority, Proceedings of the 21st Symposium on the Theory of Computing, 1989, pp 73–85

[18] Roy Radner, The Organization of Decentralized Information Processing, Econometrica, 1993, 61, pp. 1109-1146

[19] Jérôme Renault and Tristan Tomala, Probabilistic Reliability and Privacy of Communication Using Multicast in General Neighbor Networks, Journal of Cryptology, 2008, 21, pp. 250-279

[20] Ludovic Renou, Nash Implementation and Communication Networks, 2008, Mimeo, University of Leicester.

[21] Bernard Salanie, The Economics of Contracts - A Primer. Cambridge University Press, 2000

[22] Roberto Serrano and Rajiv Vohra, Multiplicity of Mixed Equilibria in Mechanisms: a Unified Approach to Exact and Approximate Implementation, mimeo, 2007

[23] Timothy Van Zandt, Communication Complexity and Mechanism Design, Journal of European Economic Association, 2007, 5, pp. 543-553