Carbon Credit Allocation: A Global Company Distributing Emission Reductions to Departments and Supply Chain Partners

Simone Genetin

Supervisors: Dr./Pr Andreas Heinecke, Dr./Pr Philipp Strack

Abstract

The issue tackled in this capstone thesis is that of a global company allocating carbon emission reductions to its divisions and supply chain partners. This problem is widely faced by multinational firms around the world, as they need to comply with carbon emission caps that either imposed institutionally or as a result of their mandatory participation in *cap and trade* mechanisms. The divisions and partners hold private information often unknown to the company, especially with regard to their cost curves, and exhibit strategic behaviour to the end of minimising the costs of the imposed emission reduction; hence, this problem can be approached as a mechanism design problem, where the social planner is the company and the divisions and partners are strategic emitting agents. This thesis builds on the work of *Lakshimi et. al.* and Bagchi et. al. [9], who proposed respectively a mechanism satisfying Strict Budget Balance (SBB) and Dominant Strategy Incentive Compatibility (DSIC) while minimising the allocative inefficiency of their mechanism, and one satisfying Allocative Efficiency (AE) and DSIC. Motivated by the importance of satisfying SBB in this context, and the clear desire of companies to minimise costs satisfying AE, this thesis proposes a dAVGA mechanism that satisfies AE, SBB, and is Bayesian Incentive Compatible (BIC). After proposing the mechanisms, this project experimentally evaluates the allocative inefficiency of Laskshimi et. al.'s mechanism compared to the proposed one. A reflection is also posed with regard to the risks related to implementing a mechanism that is only BIC, rather than DSIC.

Contents

1 Introduction to the Problem

The increased severity of climate change and its effects on societies around the world is highlighting the importance of reducing carbon emissions at an ever faster rate. Legislation aimed at decreasing carbon dioxide emission is already in place, although changing with high frequency; under current mechanisms adopted world-wide, countries trade emission allowances using *cap and trade* schemes [8]. That is, national governments and the corporations operating under them are free to trade emission allowances under a chosen limit; [8] [7] The limit is often set to decrease over time. The mechanism involves a governing body, which deliberates a strict upper bound (*cap*) on total emissions for a particular country or industry and issues allowances to be traded between *emitting agents*. [8] The initial allocation of emission allowances to emitting agents is chosen with a mixture of modes, though usually involving auctions for at least a portion of the allowances emitted. [7] [11] Extensive research is dedicated to evaluating the effects of different auction formats on the efficiency of initial allocations of carbon allowances to emitting agents. This capstone thesis builds on the work of Lakshimi et. al. [8] and Bagchi et. al. [9], to tackle the issue of a company allocating emission caps to its internal divisions and supply chain partner, once a cap on the company's emissions has been set. To achieve an allocation that minimises the costs of reducing the company's emissions, the company would need to know the cost curves of each division and supply chain partner. These are indeed likely to differ between divisions and partners. However, divisions and partners are often autonomous entities, holding private information and exhibiting strategic behaviour. [8] Hence, the company needs to elicit the cost curves from them prior to allocating allowances. Game theory and mechanism design can then serve to solve this problem, treating the company as a *social planner* and the divisions and partner as strategic *emitting agents*.[5][6] Note that the same problem can be formulated at different scopes; for example, the social planner can be the an institution in charge of establishing the initial allocation of allowances to firms in an industry; the various firms would then be the emitting agents.

The problem will be formulated as that of allocating emission reductions, rather than allowances, for convenience and coherence with existing literature. Figure 1 illustrates the issue at hand:

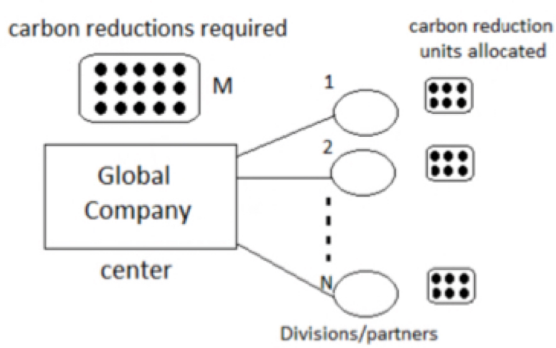

Figure 1: A company allocating a set amount of carbon reductions to emitting agents [8] [9]

2 Game Theory and Mechanism Design: Relevance to the Problem

Allocating emission reductions to strategic agents is merely an optimization problem if the cost curves of each agent are known to the social planner who allocates the reductions. Given that the emitting agents act strategically and hold private information regarding their cost curves, which is unknown to the social planner, this is an *incompletely specified optimization problem, where some of the inputs to the problem are held by [self-interested] agents*. [4] [9] [5] This induces two problems: the *Preference Elicitation* and *Preference Aggregation* problem. To understand them, I will introduce the necessary terminology and definitions.

2.1 Foundational Definitions and Notation

This section takes inspiration from the notation used in [4]. For a rigorous treatment of the definitions and theorems used in this report, as well as foundational terms and theorems of game theory such as the definition of game, strategies, mixed and behavioural strategies, Nash equilibrium, Bayesian Nash equilibrium, dominant Strategy equilibrium etc., Nahari et. al. chapter I to II [4] is suggested. The interested reader might find Fudenberg et. al. [1] chapter I to III to be a more extensive source.

Definition 2.1 (Mechanism)**.** A mechanism is an institution or framework of protocols that prescribes ways of interactions amongst self-interested agents, each with private information, as to ensure an outcome that is socially desirable from the interaction between agents.

Notice that without a mechanism in place, a free interaction between agents might lead to an unwanted outcome. The properties of the outcome are for the social planner (in this case, the global company) to define.

Definition 2.2 (Player Type)**.** The type of a player embodies any private information that is relevant to the player's decision making, including a player's payoff function, their beliefs about other player's pay-off functions, their beliefs about other players' beliefs about their beliefs, recursively to the limit.

Formally, as explained by Strack [12], let A be the set of action profiles for n players in a game, with Ω the set of parameters affecting payoffs. Ω can be interpreted as the set of states of the world. A state $\omega \in \Omega$ specifies the physical universe, past, present, and future state of the world, what players know, what they know about what others know etc. to the limit, what they can do, believe others are doing etc to the limit, their utility functions, their beliefs about other players utility functions etc. to the limit. Let player i's knowledge be modelled by a possibility partition ρ_i of Ω. Note that $P_i(\omega) \in \rho_i(\Omega)$ describes the states that player i thinks are possible when the true state of the world is ω, such that if $\omega, \omega^{'} \in \Omega \in P_i \in \rho_i$, then player i cannot distinguish ω and $\omega^{'}$. Let $u : \Omega \times A \to \mathbb{R}^n$ be the payoff function. Consider now that the tuple (A, Ω, u) does not describe a game, as the players' beliefs about Ω must be specified. Let $\rho_i^1 := \Delta\Omega$ be the set of first orders beliefs of every player. Note that each player's beliefs about other players' beliefs must also be specified; note that we must specify these beliefs to the limit, such that player i's k-th order beliefs p_i^k are an element of $\rho_i^k := \Delta(\Omega \times \times_{j \neq i} \rho_j^{k-1})$, and such that the specification of player's i k-th order beliefs imply the specification over their $(k-1)$th order beliefs when considering the appropriate marginal distribution. Then, it is needed to define $\rho_i^\infty = \times_{k \in \mathbb{N}} \rho_i^k$, the set containing every belief of every player about any belief. Note that with reasonable assumptions, we can define the **universal type space** universal type space for player i: $\rho_i^\infty, i = 1, \ldots, n$, which is well defined and of which every element induces a distribution over $\Delta(\Omega \times \times_{j \neq i} \rho_j^\infty)$. Modelling games and mechanisms over this universal type space is inconvenient, as it involves subjective uncertainty and induces a type space that is too large for any application.

Definition 2.3 (Event)**.** An even is any subset $E \subseteq \Omega$. [12]

Note that the set of all events it then the possibility partition over Ω, $\rho(\Omega)$. [12]

Definition 2.4 (Common Knowledge)**.** The event $E \subseteq \Omega$ is common knowledge among players $i = 1, \ldots, n$ at state $\omega \in \Omega$ if and only if $k \in \mathbb{N}$ and any sequence of players (with possible reptition) $(i_1, \ldots, i_k)$,

$$P_{i_k}(P_{i_{k-1}} \ldots (P_{i_1})) \subset E$$

. [12]

This corresponds to an event being known to every player, and every player knowing that everyone knows about the event, etc. to the limit higher order beliefs. Recall that an even is **self evident** to a player if it is the union of some of their information partition cells. Intuitively, an event E is self evident to a player if the player can be sure that $\backslash E$ is impossible at a state $\omega \in E \subseteq \Omega$. Then, it can be shown that an event $E \subseteq \Omega$ is common knowledge at state $\omega \in \Omega$ if there exists an even F that is **self-evident** to all players (i.e., **public**), such that $\omega \in F \subseteq E$. The reader interested in the proof, as well as a rigorous treatment of a model of knowledge and beliefs, shall refer to [12], or [1] [Definitions and proof to be added to Appendix for final report].

2.2 Bayesian Games

To be able to tackle games of incomplete information, without having to worry about hierarchies of beliefs, an assumption can be made. That is, we can assume that there exists common knowledge amongst agents about an underlying random variable that determines the private information of each player. [12] This way, a game of incomplete information can be reduced to a game of imperfect information. [12] Note that not every element of $(\rho_i^\infty)_{i=1}^n$ can be generated following this assumption, but the infinite hierarchy of beliefs need not be derived in order to analyse a Bayesian game. [12] [1]

Definition 2.5 (Bayesian Game)**.** A Bayesian game is defined by a set of types Θ_i for each player $i = 1, \dots, n$; The types $(\theta_1, \dots, \theta_n)$ are drown from $\Theta := \times\Theta_i$ according to a distribution p which is the **common prior**. Players only observe their own types. The marginal distribution over player i's types, p_i, assigns positive probability to all types of player $i : \forall\theta_i \in \Theta_i, p_i(\theta_i) > 0$. Then, there are specified a set of pure strategies S_i, mixed strategies $\sigma_i \in \Sigma_i$, and payoffs $u_i : S \times \Theta \to \mathbb{R}$.[12] [1]

Note that players can use Bayes rule to determine the probability of any given signal profile of other players, conditional on observing their own profile, which is a very realistic assumption. [12]

Now we define a Bayes Nash equilibrium as per [12], which is the correspondent of the Nash Equilibrium for Bayesian games. Refer to [1] for a definition of Nash equilibrium in games of complete information.

Definition 2.6 (Bayes Nash equilibrium)**.** Given a Bayesian game (Θ, p, S, u), a Bayes Nash equilibrium is a profile $\sigma \in \Sigma$ s.t. $\forall i \in [1, n], \forall\theta_i \in \Theta_i, \forall s_i \in supp\ \sigma_i(\theta_i)$,

$$s_i \in argmax_{s_i' \in S_i} \sum_{\theta_{-i} \in \Theta_{-i}} p(\theta_{-i}|\theta_i) u_i(s_i', s_{-i}(\theta_{-i}), (\theta_i, \theta_{-i}))$$

such that every strategy in the strategy profile maximised the expected value over the probability distribution of types of every other agent, given a player's type.

2.3 The Mechanism Design Environment

We can now define the environment and notation to be used to explore the problem at hand. A mechanism design problem is set as follows, as per [4]:

1. There are n agents, with $N = \{1, , n\}$ acting strategically. Agents can be modelled to act rationally according to different axioms (i.e. of risk aversion), though this is usually stated;
2. There exists a well defined set of outcomes, X;
3. Θ_i denotes the set of private values for player $i \in [1, n]$, and $\Theta := \Theta_1 \times \dots \Theta_n$; a profile is represented as a vector, $\theta = (\theta_1, \dots, theta_n)$
4. Prior to making a collective choice from X, each player observes their own preferences over each element of the set, or each *alternative*. This is modelled assuming the observation of each type θ_i for each player $i \in [1, n]$. Every player only observes their type, as in a Bayesian game;
5. We assume the existence of a common distribution $p \in \Delta(\Theta)$, as in a Bayesian game;
6. There exists well defined utility functions over the outcomes, $u_i : X \times \Theta \to \mathbb{R}$, as in a Bayesian game
7. The set X, the set of players N, the type sets Θ_i, the common prior p, and the payoff functions $u_i, i = 1, \dots, n$, are common knowledge between all agents.

Notice indeed that a mechanism is then an institution deliberating the set of strategies and outcomes possible for a Bayesian game. To characterise the choice of a collective alternative given a set of individual types, we define a **social choice function** as per [4]

Definition 2.7 (Social Choice Function). Given a set of agents N, their type sets $\Theta_1, \ldots, \Theta_n$,$\Theta := \Theta_1 \times \cdots \times \Theta_n$ and a set of outcomes X, a social choice function is a mapping $f : \Theta \rightarrow X$ which assigns to each possible type profile a collective choice from X.

Clearly, the social choice function can be chosen by the social planner to select desired outcomes $x \in X$. Note that there are two problems the social planner faces when wanting to assign an element $x \in X$, given a set of agents N: [4]

- **Preference Elicitation Problem**: given a social choice function f, recall that the individual types making up the type profile of the agents are private information, not common knowledge. For the social choice $f(\theta_1, \ldots, \theta_n)$ to be chosen when the type profile is $(\theta_1, \ldots, \theta_n)$, players must disclose their true types. Let $(\hat{\theta_1}, \ldots, \hat{\theta_n})$ be the profile of disclosed types. Recall however that agents act strategically, hence, they might find it in their interest to disclose any type in Θ_i. Then, eliciting true types such that $f(\theta_1, \ldots, \theta_n) = f(\hat{\theta_1}, \ldots, \hat{\theta_n})$ is the first problem to be solved by the social planner.

- **Preference Aggregation Problem**: Once types are disclosed by the agents, the type profile has to be transformed to an outcome, according to the social choice function. The preference aggregation problem is usually an optimization problem.

Definition 2.8 (Direct Mechanism). [4] Given a social choice function $f : \Theta \rightarrow X$, the tuple $(\Theta_1, \ldots, \Theta_n, f(.))$ defines a direct mechanism.

That is, a direct mechanism involves every agent directly reveling their type, rather than taking any other action affected by their type.

Definition 2.9 (Indirect Mechanism). [4] Given a set of strategies (actions) for each player $S_i, i = 1, \ldots, n$. and $g : S_1 \times \cdots \times S_n \rightarrow X$, the tuple $(S_1, \ldots, S_n, g(.))$ denotes an indirect mechanism.

That is, in an indirect mechanism each player has a choice of actions (strategy set), and an outcome is specified for each action profile. This is a more realistic setting, where agents act accordingly to their type in the Bayesian game induced by the mechanism. The figures below help explaining the difference between a direct and indirect mechanism.

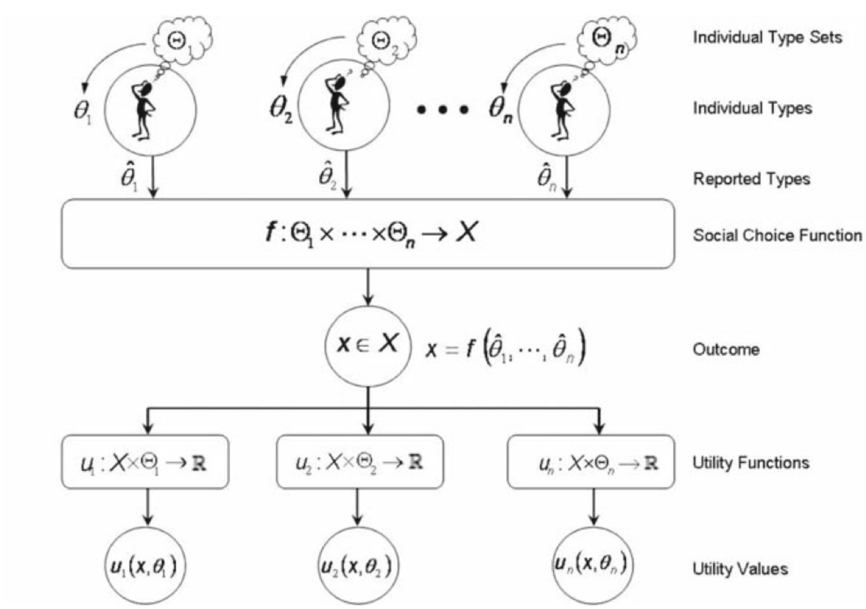

Figure 2: A direct mechanism [4]

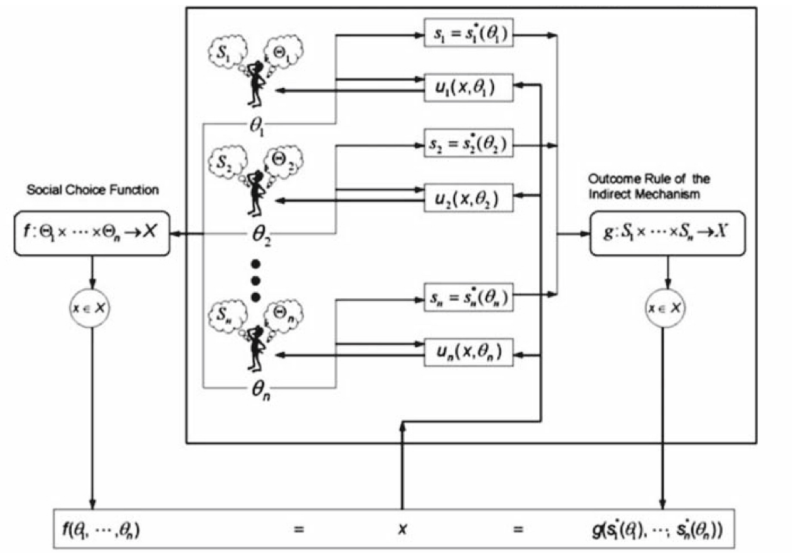

Figure 3: An indirect mechanism [4]

2.4 Implementation of Social Choice Functions

Next, we focus on what it means to implement a social choice function. The inexperienced reader might want to get acquainted with weakly dominant strategy and dominant strategy equilibrium first, either in [4] or [1].

Definition 2.10 (Implementation of Social Choice Function)**.** [4] A mechanism $\mathcal{M} = ((S_i)_{i\in N}, g(.))$ implements a social choice function $f(.)$ if there exists a pure strategy equilibrium profile $s^* = (s_1^*(.), \ldots, s_n^*(.))$ of the induced Bayesian game Γ^b such that $g(s_1^*(\theta_1), \ldots, s_n^*(\theta_n)) = f(\theta_1, \ldots, \theta_n) \forall (\theta_1, \ldots, \theta_n) \in \Theta$.

Depending on the nature of the equilibrium characterising the induced game, a social choice function can be implemented in **Dominant Strategies** or in a **Bayesian Nash Equilibrium**

Definition 2.11 (Implementation in Dominant Strategies)**.** [4] A mechanism $\mathcal{M} = ((S_i)_{i\in N}, g(.))$ implements a social choice function $f(.)$ in a dominant strategy equilibrium if there exists a weakly dominant strategy equilibrium $s^*(.) = (s_1^*(.), \ldots, s_n^*(.))$ of the induced Bayesian game Γ^b such that $g(s_1^*(\theta_1), \ldots, s_n^*(\theta_n)) = f(\theta_1, \ldots, \theta_n) \forall (\theta_1, \ldots, \theta_n) \in \Theta$.

Recall that a strongly dominant strategy equilibrium is a a weakly dominant strategy equilibrium, though not the reverse. Also recall that a strongly dominant strategy equilibrium must be unique, whereas a weakly dominant strategy equilibrium can not be. [4]

Definition 2.12 (Implementation in a Bayesian Nash Equilibrium)**.** [4] A mechanism $\mathcal{M} = ((S_i)_{i\in N}, g(.))$ implements a social choice function $f(.)$ in a dominant strategy equilibrium if there a pure strategy Bayesian Nash equilibrium $s^*(.) = (s_1^*(.), \ldots, s_n^*(.))$ of the induced Bayesian game Γ^b such that $g(s_1^*(\theta_1), \ldots, s_n^*(\theta_n)) = f(\theta_1, \ldots, \theta_n) \forall (\theta_1, \ldots, \theta_n) \in \Theta$.

Recall on this regard that a pure strategy Bayesian Nash Equilibrium might not exist; only mixed strategy Bayesian Nash Equilibrium are guaranteed to exist in every game. [4]

2.4.1 Strong and Weak Implementations

2.5 Incentive Compatibility

Consider any Mechanism Design problem. It is intuitive to think that solving the preference elicitation problem is possibly the most pivotal element in formulating an efficient mechanism. Indeed, no matter the properties we wish our mechanism to satisfy, eliciting truthful types allows the social planner to solve an optimization problem where every input is known. The social planner must offer incentives that assure the agents' best response is to reveal their truthful types. Notice that revelation of types is a concern only in direct mechanism, hence, this section is only relevant when implementing a direct mechanism. Later on, it will be explained why focusing on direct mechanism is desirable and acceptable.

Definition 2.13 (Incentive Compatibility)**.** [4] A social choice function is incentive compatible, or truthfully implementable, if the Bayesian game induced by the direct mechanism $\mathcal{D} = ((\Theta_i)_{i\in N}, f(.))$has a pure strategy equilibrium $s^*(.) = (s_1^*(.), \dots, s_n^*(.))$ in which $s^*(\theta_i) = \theta_i \forall \theta_i \in \Theta_i, \forall i \in N$.

Definition 2.14 (Dominant Strategy Incentive Compatibility (DSIC))**.** [4] A social choice function is dominant strategy incentive compatible if the Bayesian game induced by the direct mechanism $\mathcal{D} = ((\Theta_i)_{i\in N}, f(.))$ has a weakly dominant pure strategy equilibrium $s^*(.) = (s_1^*(.), \dots .s, s_n^*(.))$ in which $s^*(\theta_i) = \theta_i \forall \theta_i \in \Theta_i, \forall i \in N$

Definition 2.15 (Bayesian Incentive Compatibility (BIC))**.** [4] A social choice function is Bayesian incentive compatible if the Bayesian game induced by the direct mechanism $\mathcal{D} = ((\Theta_i)_{i\in N}, f(.))$ has a pure strategy Bayesian Nash equilibrium $s^*(.) = (s_1^*(.), \dots, s_n^*(.))$ in which $s^*(\theta_i) = \theta_i \forall \theta_i \in \Theta_i, \forall i \in N$.

Theorem 2.1. Any social choice function $f(.)$ that is dominant strategy incentive compatible is Bayesian incentive compatible

Proof. Recall that a weakly dominant strategy equilibrium is a Bayesian Nash equilibrium. [4][1][12]. It follows that if the Bayesian game induced by the direct mechanism $\mathcal{D} = ((\Theta_i)_{i\in N}, f(.))$ has a weakly dominant pure strategy equilibrium $s^*(.) = (s_1^*(.), \dots .s, s_n^*(.))$ in which $s^*(\theta_i) = \theta_i \forall \theta_i \in \Theta_i, \forall i \in N$, then it has a pure strategy Bayesian Nash equilibrium where the same is true. □

2.6 The Revelation Principle

The revelation principle for dominant strategy equilibrium and Bayesian Nash equilibrium allows us to focus on direct mechanisms only, which makes the analysis of implementable social choice functions significantly less complex. For a proof, refer to Narahari et. al., page 75. [4]

Theorem 2.2. Assume there exists an indirect mechanism $\mathcal{M} = ((S_i)_{i\in N}, g(.))$ which implements a social choice function $f(.)$ in a dominant strategy equilibrium. Then, $f(.)$ is dominant incentive strategy compatible.

Consider also that the set of social choice functions which are DSIC is a subset of the social choice functions which are implementable in dominant strategies; indeed, consider that direct mechanism are indirect mechanisms where the set of strategies available to each player is limited to revealing a type. Then clearly, if a social choice function is DSIC, it must be implementable in dominant strategies as well. Figure 4 shows that the set difference between the set of social choice functions which are implementable in dominant strategies and DSIC is the empty set.

Theorem 2.3. Assume there exists an indirect mechanism $\mathcal{M} = ((S_i)_{i\in N}, g(.))$ which implements a social choice function $f(.)$ in a Bayesian Nash equilibrium. Then, $f(.)$ is Bayesian Nash incentive compatible.

For a proof, refer to Narahari et. al., page 76. [4]. Consider also that the set of social choice functions which are BIC is a subset of the social choice functions which are implementable in a Bayesian Nash equilibrium. Then clearly, if a social choice function is BIC, it must be implementable in a Bayesian Nash equilibrium. Figure 5 shows that the set difference between the set of social choice functions which are implementable in a Bayesian Nash equilibrium and BIC is the empty set.

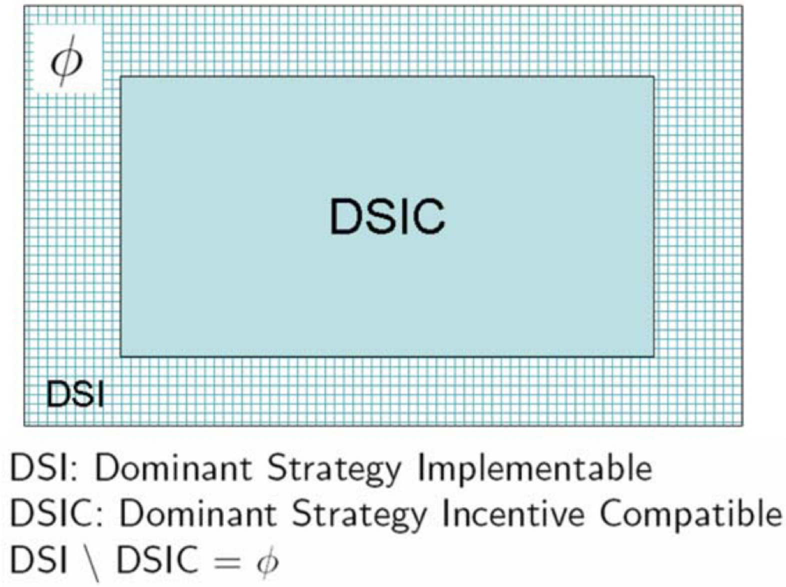

Figure 4: The set difference between dominant strategy implementable and DSIC social choice functions [4]

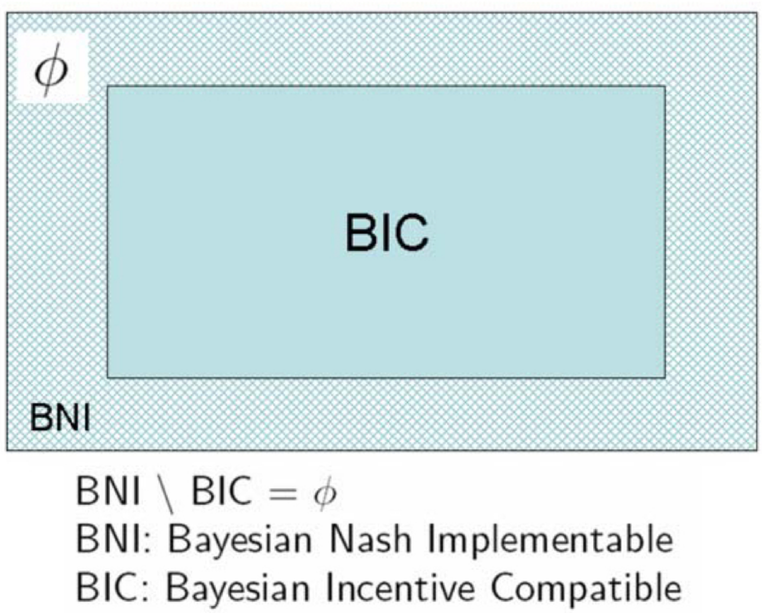

Figure 5: The set difference between Bayesian Nash equilibrium implementable and BIC social choice functions [4]

3 Social Choice Function Properties

Next, we define some properties that will let us understand the behaviour of the social choice function to be implemented, or that are implemented in related works.

3.1 Dictatorship and Efficiency

Definition 3.1 (Dictatorship)**.** [4] A social choice function $f : \Theta \to X$ is *dictatorial* if there exists an agent d satifying the following property:

$$\forall \theta \in \Theta, f(\theta) \text{ is s.t. } u_d(f(\theta), \theta_d) \geq u_d(x, \theta_d) \forall x \in X \ .$$

That is, every outcome picked by the social choice function is the most favoured by at least one agent, the dictator.

Definition 3.2 (Ex-post efficiency)**.** [4] For any given set of social choice functions F, and any member $f(.) \in F$, $f(.)$ is ex-post efficient in F if and only if there exists no other $\hat{f}(.) \in F$ such that:

$$\begin{aligned} &u_i(\hat{f}(\theta), \theta_i) \geq u_i(f(\theta), \theta_i) \ \forall i = 1, \dots, n, \forall \theta \in \Theta \\ &u_i(\hat{f}(\theta), \theta_i) > u_i(f(\theta), \theta_i) \text{ for some } i = 1, \dots, \text{ and for some } \theta \in \Theta \end{aligned}$$

Note that a social choice function is then ex post efficiency as per definition 3.32 if and only if for every profile of agents' types $\theta \in \Theta$, the outcome $f(\theta)$ is a Pareto optimal outcome, when $F = \{f : f : \Theta \to X\}$. [4]

3.2 The Gibbard—Satterthwaite Impossibility Theorem

A brief mention of the G—S Impossibility Theorem is mentioned below. This will give a justification to our choice of focusing on quasi linear environments, and on BIC social choice functions, for the development of the mechanism at hand. For a more rigorous treatment of the subject, refer to [4] [10]. Knowledge is assumed about the definition of rational preference relations and strict-total preference relations on a set, but the sources cited also treat this topic in depth if the reader wishes to delve more rigorously into the topic.

Let the set fo all ational preference relations and strict-total preference relations on the set X be defined as $\mathcal{R}$ and $\mathcal{P}$, as per [4]. Recall also that an agent's utility function, defined over every element of X, induces a rational preference relation over X. In that sense, the set of ordinal preferences for agent i is defined as such: set $\mathcal{R}_i = \{\succsim : \succsim = \succsim_i (\theta_i) \text{ for some } \theta_i \in \theta_i\}$, and clearly $\mathcal{R}_i \in \mathcal{R}$.

Theorem 3.1. [4] For a social choice function $f : \Theta \to X$, assume that

1. X is finite, and contains at least three elements;
2. $f(.)$ is a surjective function;
3. $\mathcal{R}_i = \mathcal{P} \ \forall i \in N$;

then, $f(.)$ is DSIC if and only if it is dictatorial.

For a proof, consult Proposition 23.C.3 from [2]. Consider clearly that in most contexts, having a dictatorial social choice function is heavily undesirable. For sure it is in our context, as it limits the allocations to all those preferred by a single division. Also, note that the last means that all agents have an extremely rich set of preferences, respectively, their preferences are exactly the set of strict total preferences relations on X. Restricting the set of preferences is a trick heavily used i the literature to get around the G—S theorem. Another route clearly is to focus on BIC mechanism, as this project will do later.

3.3 The Quasi-linear Environment

The following characteristics define a class of environments widely adopted to approach mechanism designs, where the G—S theorem does not hold. let X be a set of vectors in the form $x = (k, t_1, \ldots, t_n)$, where $k \in K$, the set of allocations, assumed to be finite for simplicity and realism. $t_i \in \mathbb{R}$ represents the monetary transfer of agent i, where $t_i > 0$ implies receiving money, $t_i < 0$ paying money. Then,

$$X = \{(k, t_1, \ldots, t_n) : k \in K, t_i \in \mathbb{R} \forall i \in N\} \ .$$

Clearly, a social choice function in this environment takes the form of

$$f(\theta) = (k(\theta), t_1(\theta), \ldots, t_n(\theta)),\ k(\theta) \in \Theta\ \forall \theta \in \Theta$$

. Then, for any direct mechanism in this environment, the agents' utility functions take the form

$$u_i(x, \theta_i) = u_i((k, t_1, \ldots, t_n), \theta_i) = v_i(k, \theta_i) + m_i + t_i$$

such that m_i is the initial endowment of a player, and the function $v_i(.)$ is their evaluation function. It is possible that in a direct mechanism, the set of possible valuation functions might coincide with Θ. Then, the utility functions would not be common knowledge as discussed in 2.3. [4] Note an interesting result:

Lemma 3.1.1. [4] All social choice functions in a quasi-linear environment are non-dictatorial

For a proof, refer to Narahari et. al., page 97. [4]

In the quasi-linear environment, two more properties need to be defined which are pivotal to this project.

Definition 3.3 (Allocative Efficiency (AE))**.** [4] A social choice function $f(.) = (k(.), t_1(.), \ldots, t_n(.)$ is allocatively efficient if, $\forall \theta \in \Theta$,

$$k(\theta) \in \operatorname*{argmax}_{k \in K} \sum_{i=1}^{n} v_i(k, \theta_i)$$

which corresponds to

$$\sum_{i=1}^{n} v_i(k(\theta), \theta_i) = \operatorname*{argmax}_{k \in K} \sum_{i=1}^{n} v_i(k, \theta_i) \ ,$$

which implies that the allocation $k(\theta)$ maximises the sum of values of all players for every $\theta \in \Theta$. The other property has to do with payments:

Definition 3.4 (Strict Budget Balance)**.** [4] A social choice function $f(.) = (k(.), t_1(.), \ldots, t_n(.)$ is allocatively efficient if, $\forall \theta \in \Theta$,

$$\sum_{i=1}^{n} t_i = 0 \ .$$

This implies that the system is closed, with no leakage or deficit of transfers.

Lemma 3.1.2. [4] A social choice function $f(.) = (k(.), t_1(.), \ldots, t_n(.)$ is ex-post efficient in a quasi linear environment if and only if it is allocatively efficient and strongly budget balanced.

For a proof, refer to Narahari et. al., page 96. [4]. We are now ready to introduce the results obtained by Lakshimi et. al. and Bagchi et. al. . [8] [9]

3.4 Vickrey—Clarke—Groves Mechanisms

In this subsection, I show that there exists a class of social functions in the quasi-linear environment which are AE and DSIC, however, cannot be SBB. This impossibility theorem will set the context for the works of Laskhimi et. al. and Bagchi et. al. [8] [9]

Theorem 3.2. Groves Theorem [4] Let the social choice function $f(.) = (k(.), t_1(.), \ldots, t_n(.)$ be allocatively efficient. Then, $f(.)$ is DSIC if it and only if satisfies the following condition satisfies the following payment structure:

$$t_i(\theta) = \left[\sum_{j \neq i} v_j(k^*(\theta), \theta_j)\right] + h_i(\theta_{-i}),\ \forall i = 1, \ldots, n$$

where $h_i : \Theta_{-i} \to \mathbb{R}$ is any arbitrary function such that $\sum_i t_i(\theta) \leq 0\ \forall \theta \in \Theta$.

For a proof, the reader shall consult Narahari et. al. page 99 (2.14.2) [4].

Definition 3.5 (Groves Mechanisms (VCG))**.** Any direct mechanism where the social choice function is AE and satisfies 3.2 is called a Groves mechanism, or more popularly a VCG mechanism.

Now, let the $\mathcal{F} = \{f : K \to \mathbb{R}\}$, that is, $\mathcal{F}$ contains all the valuation functions of the agents. We can now express the impossibility theorem regarding the budget imbalance of any Groves mechanism.

Theorem 3.3. Green—Laffont Impossibility Theorem [4] Suppose that for every agent $i = 1, \ldots, n$, $\mathcal{F} = \{v_i(k(\theta), \theta_i) : \theta_i \in \Theta_i\}$; that is, every possible valuation function arises for some $\theta_i \in \Theta_i$. Then, no social choice function can be ex-post efficient and DSIC.

A proof is available for the reader at Narahari et. al., page page 102 (2.14.3.1). [4] Note that in the above example v_i does not depend on everyone's type, but only agent i's. This is a common assumption in the literature, though it is opportune to ask whether this should be assumed, or $v(.)$ should depend on everyone's types. In our context, it is reasonable to believe that different divisions and partners have close to no interest in the cost curves of other divisions and partners. [4].

Recall that in the quasi-linear environment, any ex-post efficient social choice function is both AE and SBB 3.1.2. Then, if the set of possible agent types is rich enough, no social choice function exists that is AE, SBB and DSIC. Recall that we have already shown that there exists a class of social choice functions that is AE and DSIC instead 3.2. For completeness, a possibility theorem is also reported that explains in which circumstances a social choice function in the quasi-linear environment is AE, SBB and DSIC.

Theorem 3.4. Strict Budget Balanced Groves Mechanisms [4] If the preferences over K are known to the social planner for at least one agent, then $h(.)$ can be chosen such that $\sum_{i=1}^{n} t_i(\theta) = 0$.

Again, a proof is available to the reader at Narahari et. al., page 102 (2.14.3.1) [4]. Note that it is very unlikely that the type set for any agent in the mechanism is a singleton, and this possibility theorem has potentially very few applications. [4]

4 Related Works

4.1 Mechanism Design for Allocation of Carbon Emission Reduction Units: A Study of Global Companies with Strategic Divisions and Partners [9]

4.1.1 Set Up

This paper sets up the context and mechanism environment for Lakshimi et. al.'s one too — indeed, Dr. Lakshimi is its second author. In this work, the cost curves of each agent are their private information. Agents are assumed to be intelligent, with capability to compute their own emission levels, and with accurate knowledge of their cost curves as defined by the cost needed to reduce some amount of carbon emissions. Cost curves are assumed to be marginally increasing, piecewise constant. That is, a cost curve for agent p can be represented as a sequence of tuples $< p, u, c >$, indicating the number of emissions u that can be reduced at cost c. For example, $< i, u_{i1}, c_{i2} >, < i, u_{i2}, c_{i2} >, < i, u_{i3}, c_{i3} >, < i, u_{it}, \infty >$. Note that t is the number of tuples in agent's i type. [9] Note that it must be that $u_{i1} < u_{i2} < \cdots < u_{it}$ and $c_{i1} < c_{i2} < \cdots < c_{it}$. A cost curve for an arbitrary agent is represented in figure 6. Given a sequence of tuples for agent i, the cost

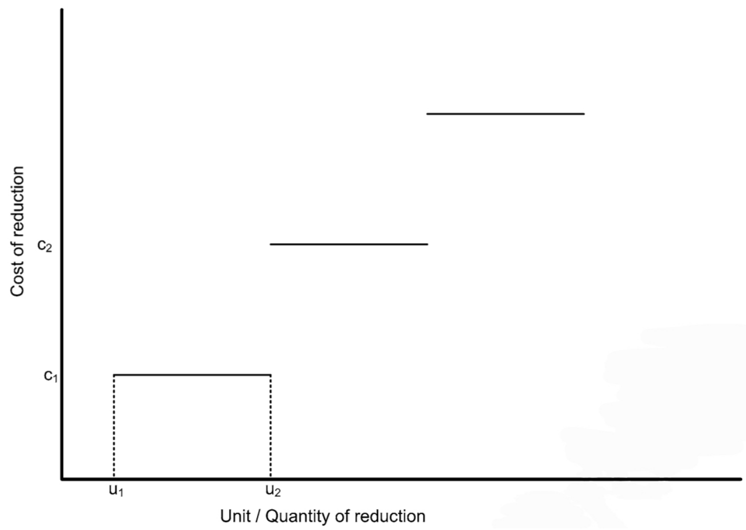

Figure 6: A marginally increasing, piecewise continuous cost curve [9]

for x units of reductions for agent i can be computed as follows:

$$cost_i(x) = \begin{cases} ICost + SCost & if\ x > o \\ 0 & otherwise \end{cases}$$

$$ICost = \begin{cases} u_{i1} \times c_{i1} & if\ x \geq u_{i1} \\ x \times c_{i1} & otherwise \end{cases}$$

$$SCost = \begin{cases} \left[x - \left[\sum_{k=1}^{n} (u_{i(k+1)} - u_{ik})\right]\right] \times c_{i(n+1)} + \sum_{k=1}^{n} \left[(u_{i(k+1)} - u_{ik} \times c_{i(k+1)})\right] & if\ x \geq u_{i1} \\ 0 & otherwise \end{cases}$$

4.1.2 Contribution to the Literature

This paper is possibly the first in the literature to address this problem disregarding the naïve assumption that divisions and partners are honest agents and agree to disclose their cost curves without indulging in strategic interactions. The paper places utmost importance on satisfying AE and DSIC for any mechanism developed. Recall that we have already shown that a VCG mechanism would satisfy these properties; clearly, the cost curves at hand are quasi linear. However, the paper wishes to minimise the budget imbalance created by the mechanism. This is pivotal, as we cannot assume that a company would have the liquidity to subsidise the allocation of carbon reductions, or the willingness to withhold any positive monetary transfer which would arise from any agent payment; it is indeed likely the company would want it to be redistributed in some way to the divisions or partner of the company.

4.1.3 Mechanisms Proposed

This section will briefly mention the mechanism proposed by Bagchi et. al. .[9] For a rigorous treatment of the matter, an experimental analysis of the budget imbalance created in different stylised examples, and a consideration regarding their feasibility, the reader is suggested to consult the cited source.

VCG Mechanism with Redistribution Mechanism The first contribution of the paper is to propose a redistribution mechanism as per Cavallo et. al . [3] The mechanism proposed by Cavallo et. al. is proven to be

DSIC, ex-post individually rational, weakly budget balanced (the sum of payment is greater or equal to 0, i.e. 'no-deficit' or sellr ex-poste individually rational); out of all mechanisms with these qualities, Cavallo et. al.'s yields the greatest payoff to the agents in the mechanism.

5 Bibliography

References

[1] Drew Fudenberg and Jean Tirole. *Game Theory*. MIT Press, 1991.

[2] Andreu Mas-Colell, Michael D. Whinston, and Jerry R. Green. *Microeconomic Theory*. New York: Oxford University Press, 1995.

[3] Ruggiero Cavallo. "Optimal Decision-Making with Minimal Waste: Strategyproof Redistribution of VCG Payments". In: AAMAS '06 (2006), 882–889. DOI: 10.1145/1160633.1160790. URL: https://doi.org/10.1145/1160633.1160790.

[4] *Game Theoretic Problems in Network Economics and Mechanism Design Solutions*. 2009. DOI: 10.1007/978-1-84800-938-7.

[5] Radhika Arava et al. *Mechanism design problems in carbon economics*. 2010.

[6] Radhika Arava et al. "Optimal allocation of carbon credits to emitting agents in a carbon economy". In: 2010. DOI: 10.1109/COASE.2010.5584129.

[7] Giuseppe Lopomo et al. "Carbon allowance auction design: An assessment of options for the United States". In: *Review of Environmental Economics and Policy* 5 (1 2011). ISSN: 17506824. DOI: 10.1093/reep/req024.

[8] L. Udaya Lakshmi et al. "A strategy-proof and budget balanced mechanism for carbon footprint reduction by global companies". In: 2012. DOI: 10.1109/CoASE.2012.6386456.

[9] Deepak Bagchi et al. *Mechanism design for allocation of carbon emission reduction units: A study of global companies with strategic divisions and partners*. 2013. DOI: 10.1007/978-94-007-5995-4_3.

[10] Tilman Börgers, Daniel Krähmer, and Roland Strausz. *An Introduction to the Theory of Mechanism Design*. 2015. DOI: 10.1093/acprof:oso/9780199734023.001.0001.

[11] Yana Jin et al. "Allowance allocation matters in China's carbon emissions trading system". In: *Energy Economics* 92 (2020). ISSN: 01409883. DOI: 10.1016/j.eneco.2020.105012.

[12] Philipp Strack. *Game Theory Lecture Notes*. 2021.