"A Characterization of Efficient, Bayesian Incentive Compatible Mechanisms," by S. R. Williams. Economic Theory 14, 155-180 (1999).

A common result in Bayesian mechanism design shows that ex post efficiency in unattainable given that it must be in each agent's self-interest to reveal his private information truthfully. Such impossibility results typically require a budget constraint of some form together with a participation constraint (i.e., individual rationality). The accomplishment of this paper is to reveal a common structure to these impossibility results along with a simpler methodology for completing the proofs. This simpler methodology uses the properties of the VCG mechanism.

We consider a classic problem. An alternative $a \in A$ can be selected that affects the welfare of $n$ agents. We assume that $A$ is a finite set. Each agent $i$ receives utility $v_i(a, t_i)$ directly from the choice of the alternative $a$. Here, $t_i$ denotes agent $i$'s privately known type. For simplicity here, we assume $t_i \in I_i = [\underline{t}_i, \bar{t}_i]$, a closed subinterval of the real line. The beliefs of agents about agent $i$'s type are given by the cumulative distribution $F_i$ whose density is $f_i$. Agent $i$'s utility function is

$$u_i(x_i, a, t_i) = v_i(a, t_i) - x_i$$

where $x_i$ is a monetary transfer from agent $i$.

Let $t = (t_1, ..., t_n)$ denote a vector of types. A *revelation mechanism* specifies an alternative $a(t)$ and a vector of monetary transfers $(x_i(t))$ for each vector of types $t$. The revelation mechanism is *efficient* (EF) if

$$a(t) \in \arg\max_{a \in A} \sum_{i=1}^{n} v_i(a, t_i) \text{ for all } t \in \prod_{i=1}^{n} I_i.$$

We next turn to our constraints. Using $t_i$ to denote a true type and $t_i^*$ a report, let

$$\begin{aligned} U_i(t_i^* | t_i) &= E_{t_{-i}} [u_i(x_i(t_i^*, t_{-i}), a(t_i^*, t_{-i}), t_i)], \\ V_i(t_i^* | t_i) &= E_{t_{-i}} [v_i(a(t_i^*, t_{-i}), t_i)], \\ X_i(t_i^*) &= E_{t_{-i}} [x_i(t_i^*, t_{-i})]. \end{aligned}$$

These are the relevant functions for agent $i$'s interim decision-making assuming that the other agents are reporting honestly. For notational convenience, let

$$U_i(t_i) = U_i(t_i | t_i) \text{ and } V_i(t_i) = V_i(t_i | t_i)$$

so that we don't have to write $t_i$ twice. The constraint of *incentive compatibility* is

$$U_i(t_i) \geq U_i(t_i^* | t_i) \text{ for all } i,\ t_i^*,\ t_i \in I_i. \tag{54}$$

Of course, the revelation principle means that by studying incentive compatible revelation mechanisms we effectively study all possible Bayesian equilibria of all possible mechanisms. The constraint of *interim individual rationality* is

$$U_i(t_i) \geq 0\ i,\ t_i \in I_i. \tag{55}$$

The "0" might be replaced with some other value depending on the outside option of agent $i$. The constraint of *ex ante budget balance* is

$$\sum_{i=1}^{n} E_{t_i} [X_i(t_i)] = 0.$$

Again, the "0" might be replaced with some other value if there is some cost associated with the group choice $a$. Notice also that budget balance as a constraint is imposed in equilibrium (i.e., assuming honest reports).

## The VCG mechanism.

For a vector of reported types $t^*$, the *basic VCG mechanism* selects an efficient alternative $a(t^*)$ and orders the payments

$$x_i(t^*) = -\sum_{j \neq i} v_j(a(t^*), t_j^*).$$

We also consider in this paper individualized taxes on each agent so that his payment is

$$x_i(t^*) = -\sum_{j\neq i} v_j(a(t^*), t_j^*) + k_i$$

for some choice of a constant $k_i$. As you know, we could allow $k_i$ to be any function of $t_{-i}^*$. This particular paper focuses on the interim and ex ante stages, however; the dependence upon $t_{-i}^*$ integrates out in these stages and so no extra generality will be achieved by allowing $k_i$ to depend on $t_{-i}^*$. I'll drop the * because we know that honest reporting is a dominant strategy for each agent in the basic VCG mechanism.

**Example 176** *The Myerson-Satterthwaite bargaining problem fits here as a special case of* $A = \{trade, no\ trade\}$.

**Example 177** *This is an example due to Cramton, Gibbons and Klemperer (1987). The $n$ agents are partners in owning a business. The partners wish to dissolve the partnership and have one of the partners buy out the shares of the others. Each partner privately knows the value $v_i \in [\underline{v}_i, \overline{v}_i]$ that he places upon the entire business. The share $\alpha_i$ of each agent $i$ is public knowledge.*

*The set of alternatives is $A = \{1, ..., n\}$, indicating which agent receives the business. The ex post utility of agent $i$ is*

$$u_i(x_i(t_i^*, t_{-i}), a(t_i^*, t_{-i}), t_i) = \delta_i\left(a(t_i^*, t_{-i})\right) v_i - \alpha_i v_i - x_i(t_i^*, t_{-i}),$$

*where*

$$\delta_i\left(a(t_i^*, t_{-i})\right) = \begin{cases} 1 \text{ if } a(t_i^*, t_{-i}) = i \\ 0 \text{ otherwise} \end{cases}.$$

*We're ignoring the complication of ties or randomization in the allocation. Notice the $-\alpha_i v_i$ term in agent $i$'s ex post utility, which corresponds to his giving up his share of the business. If agent $i$ is to participate in a mechanism to award the business, then he must expect to come out ahead of the status quo, which is the value $\alpha_i v_i$ that he places on his share of the business.*

## A Central Problem of Bayesian Mechanism Design

In a given problem, can a revelation mechanism be efficient, incentive compatible, interim individually rational and ex ante budget balanced? We'll discuss below the "bones" of the standard argument. If we were interested in dominant strategy incentive compatibility, then we could apply the Green and Laffont characterization and restrict attention to the VCG mechanisms. We've relaxed our notion of incentive compatibility to Bayesian-Nash, which should allow us to accomplish more than we can with the more demanding concept of dominant strategies. We'll see, however, that relaxing incentive compatibility in this sense really doesn't help very much.

Our focus here is upon efficiency and so we'll assume this particular choice rule $a(t)$ in all that follows. The first step is to derive a formula for an agent's interim expected payment $X_i(t_i)$ in terms of the choice rule $a(t)$ using the constraint of incentive compatibility. Let's assume differentiability holds as needed and simply apply the envelope theorem: if $(a, x)$ is incentive compatible, then for each agent $i$,

$$\frac{dU_i}{dt_i} = \frac{\partial}{\partial t_i^*} U_i(t_i\,|t_i\,) + \frac{\partial}{\partial t_i} U_i(t_i\,|t_i\,) = \frac{\partial}{\partial t_i} U_i(t_i\,|t_i\,).$$

Now

$$U_i(t_i^*\,|t_i\,) = V_i(t_i^*\,|t_i\,) - X_i(t_i^*)$$

and so

$$\frac{dU_i}{dt_i} = \frac{\partial}{\partial t_i} V_i(t_i^* = t_i\,|t_i\,).$$

Consequently, for any $t_i$, $t_i^{**} \in \left[\underline{t}_i, \overline{t}_i\right]$,

$$U_i(t_i) = U_i(t_i^{**}) + \int_{t_i^{**}}^{t_i} \frac{\partial}{\partial t_i} V_i(y\,|y\,)dy$$

where $y$ is a dummy variable. Alternatively, we have

$$U_i(t_i) = V_i(t_i) - X_i(t_i)$$

and so

$$X_i(t_i) = V_i(t_i) - U_i(t_i^{**}) - \int_{t_i^{**}}^{t_i} \frac{\partial}{\partial t_i} V_i(y\,|y\,)dy,$$

i.e., incentive compatibility determines the expected payment in terms of the choice rule $a(t)$ up to a constant $U_i(t_i^{**})$. Notice that the efficient choice rule is implicit in $V_i$.

**Example 178** *Consider the Myerson-Satterthwaite bargaining model in which* $[\underline{v}, \overline{v}] = [\underline{c}, \overline{c}] = [0, 1]$. *The efficient choice rule is*

$$a(v, c) = \begin{cases} \textit{trade if } v \geq c \\ \textit{no trade if } v < c \end{cases}$$

*Therefore, letting the buyer be agent* $i$*,*

$$\begin{aligned} V_i(t_i^*\,|t_i\,) &= E_{t_{-i}}\left[v_i(a(t_i^*, t_{-i}), t_i)\right] \\ &= \int_0^{v^*} v f(c) dc \\ &= vF(v^*) \end{aligned}$$

*where* $v^*$ *denotes his report. This implies*

$$\frac{\partial}{\partial t_i} V_i(y\,|y\,) = F(y).$$

*For the seller, we have*

$$\begin{aligned} V_i(t_i^*\,|t_i\,) &= E_{t_{-i}}\left[v_i(a(t_i^*, t_{-i}), t_i)\right] \\ &= \int_{c^*}^{1} -cg(v) dv \\ &= -c\,(1 - G(c^*)) \end{aligned}$$

*where* $v^*$ *denotes his report. This implies*

$$\frac{\partial}{\partial t_i} V_i(y\,|y\,) = -\left(1 - G(y)\right).$$

Let's return to the central problem. For the efficient choice rule $a(t)$, does there exist a payment rule $x(t)$ such that $(a, x)$ satisfies IC, interim individual rationality and ex ante budget balance? We know that $x(t)$ must satisfy for each $i$ the equation

$$X_i(t_i) = V_i(t_i) - U_i(t_i^{**}) - \int_{t_i^{**}}^{t_i} \frac{\partial}{\partial t_i} V_i(y\,|y\,)dy.$$

Also,

$$U_i(t_i) = U_i(t_i^{**}) + \int_{t_i^{**}}^{t_i} \frac{\partial}{\partial t_i} V_i(y\,|y\,)dy.$$

This states that at the interim stage, the payment of each agent and his utility is determined up to a constant by the constraint of incentive compatibility.

We now make the following observation: the VCG mechanisms are incentive compatible and efficient and so they satisfy the two formulas above. By varying the individualized taxes in the VCG mechanisms, we can make the constants $U_i(t_i^{**})$ assume whatever values we want. Consequently, at the interim stage, any IC and efficient mechanism "looks" to each agent just like a VCG mechanism. By "looks like", this means that the interim expected utilities and the interim expected payments are exactly the same as in some VCG mechanism.

Let's go back to the constraints of interim individual rationality and ex ante budget balance. These are interim and ex ante constraints; therefore, they can be expressed in terms of $U_i(t_i)$ and $X_i(t_i)$. If an incentive compatible mechanism exists that is efficient, ex ante budget balanced and interim individually rational, then some VCG mechanism is necessarily ex ante budget balanced and interim individually rational, because the VCG mechanisms span the entire set of interim utility and payment functions that satisfy efficiency and incentive compatibility. Consequently, if we want to determine whether or not an incentive compatible, efficient, interim individually rational and ex ante budget balanced mechanism exists, we can restrict our attention to the family of VCG mechanisms.

This is surprising because it states that for this particular problem there is nothing gained from relaxing the solution concept from dominant strategies to Bayesian Nash equilibrium. Bayesian Nash equilibrium is far less plausible as predictor of behavior than dominant strategies; economic theory moved toward Bayesian Nash equilibrium, however, because results such as the Gibbard-Satterthwaite Theorem showed how restrictive dominance could be. The sense was, "Dominant strategies are nice but we can not accomplish much using them...let's relax our notion of incentive compatibility." In this particular problem, however, this relaxation does not gain us anything.

**Example 179** *Let's go back and redo our proof of the Myerson-Satterthwaite impossibility result. We assume for simplicity that* $[\underline{v}, \overline{v}] = [\underline{c}, \overline{c}] = [0, 1]$. *Can a mechanism exist that is efficient, incentive compatible, ex ante budget balanced and interim individually rational? If so, then some VCG mechanism exists with this property. We apply the constraints to the family of VCG mechanisms and show that no member of this family can satisfy these constraints. Therefore, no Bayesian Nash incentive compatible mechanism can satisfy them either.*

*The family of VCG mechanisms in this problem are as follows: given the truthful reports* $v$, $c$, *trade occurs iff* $v \geq c$. *The buyer pays* $k_b$ *whether or not trade occurs; in addition, he pays* $c$ *when trade occurs. The seller pays* $k_s$ *whether or not trade occurs; in addition, he pays* $-v$ *when trade occurs.*

*The only variables we have are the constants* $k_b$ *and* $k_s$. *Can such constants be chosen so that ex ante budget balance and interim individual rationality are both satisfied? For notational simplicity, let*

$$\Gamma = E\left[v - c\,|v \geq c\right] \cdot \Pr\left(v \geq c\right),$$

*i.e., the expected gains from trading in efficient allocation rule. In the basic VCG mechanism, each trader's ex ante expected utility equals* $\Gamma$. *Ex ante budget balance therefore means that the basic VCG mechanism runs a deficit equal to* $\Gamma$. *If some other VCG mechanism is to satisfy ex ante budget balance, the constants* $k_b$ *and* $k_s$ *must cover this deficit,*

$$k_b + k_s = \Gamma. \tag{56}$$

*The interim expected payoff of a buyer with value* $0$ *or a seller with cost* $1$ *is zero in the basic VCG mechanism because these types of these traders trade with probability zero. When we consider more general VCG mechanisms, each of these traders pays* $k_b$ *or* $k_s$ *regardless of the fact that they do not trade. Consequently, interim individual rationality implies*

$$k_b, k_s \leq 0. \tag{57}$$

*The two equations (56),(57) contradict each other. No VCG mechanism can satisfy both interim individual rationality and ex ante budget balance, and therefore no Bayesian Nash incentive compatible mechanism can satisfy these properties and also be efficient.*

# Matching

This discussion is drawn from:

- *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*, by Alvin Roth and Marilda Sotomayor (Cambridge, 1990).
- Roth, Alvin E., "Deferred Acceptance Algorithms: History, Theory, Practice, and Open Questions," *International Journal of Game Theory*, Special Issue in Honor of David Gale on his 85th Birthday, 36, March 2008, 537-569.
- Gale, David, and Lloyd Shapley, "College Admissions and the Stability of Marriage," *American Mathematical Monthly*, 69, 1962, 9-15.

Matching as a topic dates from the early 1960's. It continues to grow in importance because of its usefulness in applications. This work is being led by Al Roth of Harvard University. The following are among the applications of matching theory:

- The assignment of new medical school graduates to hospital residency programs. This is the problem that initially interested Roth in matching. The hospitals have preferences over doctors based upon their specialties and academic records. The doctors have preferences over hospitals based upon the qualities of their programs and the locations of the hospitals. Salaries are not a big factor in the preferences of doctors; rather, the major issue is the quality of the hospital residency program and the skills and reputation that one will obtain by participating in it. A similar matching problem occurs in the matching of new law school graduates to clerkships with judges.

  The history is as follows: pre-1945, the hospitals competed by offering positions to students earlier and earlier in their medical training. By 1945, some students were being offered positions when they had as much of two years of further study yet to complete. The medical schools tried to put a stop to this by refusing to release student records until a date appropriately chosen late in a student's term of study. The hospitals responded by making offers with extremely tight deadlines following the date in question; some offers were even "exploding", i.e., the student needed to respond immediately upon receiving the offer. Hospitals sometimes found that their next choices were already off the market once their initial offers had been rejected. This has been replaced by the National Residency Matching Program. It is not exactly the same as the deferred acceptance algorithm in which hospitals make the proposals, but it produces the same stable outcome as this algorithm. This last point will bcome clear in the discussion that follows.

- The matching of students to public schools. It is increasingly common for public school systems to allow students a measure of choice in selecting which schools to attend. The schools have capacity constraints, however, and they may also have preferences over students based upon diversity of skills, excellence, etc.. Roth and his colleagues have been consulted by school systems for advice in the design of the algorithm that matches students to schools based upon the reported preferences of the students. Transfers of money cannot be involved in the assignment process.

- The matching of organ donors to transplant recipients. In particular, consider the case of kidney transplants. A person who needs a kidney transplant has preferences over potential donors based upon the likelihood of rejection of the transplant, which is determined by a number of biological markers (e.g. , blood type). There are "cadaver donors", but of particular interest is living donors who choose to donate one of their two kidneys. There are some "good Samaritan" donors who are motivated to help someone they may or may not know. More typically, however, a donor is motivated to assist a spouse or close relative who needs a transplant. The interesting aspect of this problem is that while a donor may not be a good match for the person he wishes to help, he may be a good match for some other person in need of a transplant. That person may know a perspective donor. Gains from trading may therefore exist among a number of donor-recipient pairs. Roth has been involved in setting up a matching process to arrange transactions of this kind. It is illegal in the U.S. for money to change hands as part of the transplant process.

We can model these problems as two distinct sets such that we wish to match elements of one set to elements of the other. The elements of the two sets have preferences and so the matching creates welfare. It is clear that some matchings may be better from a welfare perspective than others. Another feature of these problems is money is either insignificant or can play no role in allocation of welfare.

- Roth and Niederle have identified three types of market failure in which a matching algorithm or clearinghouse may be an appropriate remedy:

  1. *Unraveling* so that offers were being made at earlier and at dispersed times. This is a form of "thinness" in the market.
  2. *Congestion* so that employers found that they did not have the time to make all of the offers that they wished to make.
  3. *Strategic behavior* in the sense that participants are concerned that they cannot act straighforwardly based upon their true preferences.

The use of a matching algorithm is a form of market intervention. Economists are generally wary of intervening in markets, though the problems that Roth and Niederle cite motivate intervention. What are the attributes of a market in which a matching algorithm may be helpful? I can list four possible attributes:

1. The goods that are traded are extremely **heterogeneous**. For instance, doctors vary by both speciality and their intellects.
2. There is extreme **excess demand** for some goods, which is what causes the market unraveling.
3. **Money is "second-order"** in preferences. Lawyers who seek judicial clerkships, for instance, bear the opportunity cost of extremely large salaries in corporate law. The point of a clerkship is both the high-level experience and its signaling value on one's resumé throughout the remainder of one's career. Perhaps the new lawyers even have lexicographic preferences where the quality of the clerkship counts first and salary counts second (i.e., there is no trade-off between the quality of the clerkship and salary). It is difficult for the price mechanism to successfully allocate goods in a situation in which the participants are not particularly concerned about the prices.
4. Roth emphasizes that **traders on both sides of many of these markets have strong preferences** over who they are matched with. Buyers typically care only about the good and sellers only about the price. A person selling a house, for instance, is concerned mainly about the price and not who buys the house.

We understand how externalities cause market failure and thus motivate market intervention, and we understand how the attributes of a public good cause underprovision of that good. The four attributes above are perhaps a step towards similarly understanding in a formal theoretical sense when a matching algorithm may be needed to improve upon a market's allocation.