Understanding mechanism design – Part 2 of 3

The Vickrey-Clarke-Groves Mechanism

Michael Wooldridge
Department of Computer Science
University of Oxford, UK
mjw@cs.ox.ac.uk

Jeffrey S. Rosenschein
Department of Computer Science
Hebrew University of Jerusalem
jeff@cs.huji.ac.uk

As we saw in the first part of this short series, a mechanism design problem involves engineering the rules of a game so that, if participants then behave rationally in the game, (by choosing strategies that maximize their expected utility, for example) then the result will satisfy some desired property. So far, however, we have said nothing about what these desirable properties might be, or what mechanisms might achieve them. Here, we will dig into these two issues in a little more detail.

First let's consider the question of what desirable properties we might want a mechanism to achieve. One very natural idea is that you might want to choose a mechanism that results in the best overall outcome for the group of players as a whole. This then leads to the idea of *maximising social welfare,* a concept that we have discussed previously in this column. The simplest way to measure how good a particular outcome is from the point of view of a society is to add together the utilities that the players would obtain if this outcome were to come about: this is *utilitarian social welfare.* The problem is, for this to work, we actually have to know for each agent the utilities that they would actually obtain from the outcomes. And we can't simply ask them – they have no incentive to tell us the truth. So what we need is a mechanism that will *incentivise agents to tell the truth* about their utilities, so that we can confidently compute social welfare. For this purpose, an ingenious class of mechanisms exist, known as the Vickrey-Clarke-Groves (VCG) mechanisms, after the three economists that contributed to them: William Vickrey (1914-96), Edward H. Clarke (1939-2013), and Theodore Groves (1941-).

VCG mechanisms constitute a family of related formal approaches that 1) enable us to select an outcome out of a set of possible candidate outcomes, 2) implement the utilitarian social welfare function, and 3) incentivise agents to tell the truth about their utilities. As one example of a VCG mechanism, consider the Clarke Pivot Rule, or Clarke

Tax mechanism, which works as follows. Suppose the players are trying to choose an outcome from a set $O = \{o_1, \ldots, o_n\}$ of possible outcomes:

1. Each agent *i* declares the utility that they would obtain from each outcome o_j.
2. The mechanism computes the outcome that maximises utilitarian social welfare, according to the declared utilities: this outcome is the one selected by the mechanism.
3. The mechanism then computes for each player *i* the *total loss in utility* that player *i*'s presence caused to the other players. It does this by computing what would have happened had player *i* not participated.
4. Each player whose presence made a difference to the outcome then pays a tax corresponding to the loss of utility to other players caused by their presence.

Thus, the crux of the mechanism is that a player is *taxed* according to the loss of utility that their presence causes to others. This simple idea turns out to be remarkably powerful, for it causes the individual to declare its own utility *truthfully*, which is also what we want for the group's maximisation of true social welfare.

Let's see an example. Suppose five agents *A1,…, A5* desire to select one of three outcomes $o_1, \ldots, o_3$, and suppose their *actual* utilities were as described in the following table:

	utility to player of o_1	utility to player of o_2	utility to player of o_3
A1	27	-33	6
A2	-36	12	24
A3	-9	24	-15
A4	-18	-15	33
A5	17	2	-19
SUM	-19	-10	29

Now, if the players declare their utilities truthfully (a point we'll return to in a moment), then clearly o_3 will be the one that maximises social welfare (since 29 > -10 > -19). The next step is to compute the taxes that each player will incur. To do this, we consider for each player *i* the social welfare of each outcome assuming that player *i* had not participated. The following table summarises these hypothetical situations.

	Social welfare of o_1 if *i* had not participated	Social welfare of o_2 if *i* had not participated	Social welfare of o_3 if *i* had not participated	Tax for player *i*
A1	-46	23	23	0
A2	17	-22	5	12
A3	-10	-34	44	0
A4	-1	5	-4	9
A5	-36	-12	48	0

To see how we compute these values, consider the value -46 in the top left of the table: this is the social welfare that would be obtained from all other players had player A1 not participated. We obtain this value from the first table, by adding together the utilities of the players other than A1: -36 + -9 + -18 + 17 + -19 = -46.

Now, it turns out that the presence of players A1, A3, and A5 makes no difference in this case: outcome o_3 would have been selected even if they had not been present, and so their participation made no difference to the outcome. They are therefore taxed nothing. However, the presence of both A2 and A4 makes a difference: if A2 had not been present, then o_1 would have been selected (since 17 > 5 > -22). If A4 had not been present, then o_2 would have been selected (since 5 > -1 > -4). So both of these players are taxed, and the relevant taxes are given in the rightmost column.

To see how we compute the taxes, consider player A2. The presence of this player caused outcome o_3 to be selected, resulting in a social welfare to other players of 5. Now, if A2 had not participated, then o_1 would have been selected, yielding the other players a social welfare of 17. So, we tax A2 this loss in social welfare: 17 – 5 = 12.

Now, all of the calculations above assume that the utilities declared in the first table are truthful (otherwise, we can't have any confidence that the values really do represent social welfare). But here is the remarkable thing: given this mechanism, *a player can do no better than truthfully declare their utilities.* To see this, consider the two possibilities:

- Suppose an agent *overbids* (declares their utility from an outcome to be more than is actually the case) – but in this case, if an overbid sways the outcome where a truthful bid would not, the agent will have to pay more than the outcome is worth – so overbidding can never make sense.

- Suppose an agent *underbids* – then they might save some tax, but that will not compensate them for the loss in utility from the outcome that they would have gotten had they bid truthfully. Also, an underbid risks not getting the desired outcome, while not changing the tax they ultimately pay.

In short, this simple mechanism – which taxes players according to the loss of utility that their presence causes to others – incentivises players to truthfully declare their preferences. In game theoretic terms, we say that truthfully declaring utilities over outcomes is a *dominant strategy*. And, because rational players will truthfully declare their preferences, we can have confidence that the outcome that is selected by the mechanism will indeed be the one that maximises social welfare.

VCG mechanisms have been very widely studied in the AI community, and indeed by the wider computer science community; in the third part of this series, we will look at some of these in AI.

Further reading

Two articles that introduced mechanism design to artificial intelligence were:

- Gilad Zlotkin and Jeffrey S. Rosenschein. Negotiation and Task Sharing Among Autonomous Agents in Cooperative Domains. IJCAI 1989: 912-917.
- Eithan Ephrati and Jeffrey S. Rosenschein. The Clarke Tax as a Consensus Mechanism Among Automated Agents. AAAI 1991: 173-178.

For a comprehensive collection of articles on mechanism design in computer science and AI, see:

- N. Nisan *et al*, editors. *Algorithmic Game Theory.* Cambridge University Press, 2007.