Eliciting Decision Advice with Good Incentives and Decision-Maker Freedom

Della Penna and Balduzzi

Thursday 25th August, 2016

Consider a setting with multiple experts providing advice to a subject, who wishes to incentivize them to inform them of which of a set of available actions maximize the reward. One natural way to attempt to do is by applying the machinery of prediction markets based on sequentially shared proper scoring rules, and only settlign the markets where actions are taken. Doing this in an incentive compatible manner for the experts is however incompatible with maintaining the subjects freedom to select the action that appears optimal ex-post ([**??**], that is using the max decision rule. The core of the problem with using the machinery of prediction markets in the decision setting is that at the point where the mechanism has correct belief over the optimal action but not others, a expert can profit by changing expected rewards to make optimal action appear worse than an action for which the expected reward can be corrected.

The guiding design criterion in this work is to restrict as little as possible the decision that the subject must take (maximizing their freedom) while still providing good incentives to the experts. No previous mechanisms in the literature can truthfully aggregate arbitrary signals over many experts in BNE.

We first focus on a direct mechanism. Experts directly report their signals to the mechanism, which then uses them to compute the optimal action given the prior to output to the subject. This requires the experts and the mechanism having access to a common bayesian prior, over the joint probabilities of the rewards, true signals, the output action and the response of the subject to a given output of the mechanism. There is a BNE that aggregates all information over N experts, unlike the case in the previous mechanisms in the literature. We then analyze a simple bid based variation of the mechansim, and provide a sufficient conditions on the signal structures where information can be fully aggregated; we further show that these conditions are sufficient to make truthfulness dominant incentive compatible.

1. MODEL

We consider a single subject with access to K possible choices, their outcomes are sufficiently characterized for the subject by a reward vector R which is unobservable to the subject, where each element in r_k corresponds to the reward the subject receives if he carries out the corresponding action in K . We consider a set of N experts, each of whom have access to some signal s_i which are potentially informative about $E[e|a,c]$, we use S to denote the vector containing all signals. P denotes a prior distribution over $P(R|S)$ which is common knowledge.

To map to the mechanism design literature the P would be over states of the world in Ω and each state of the world consists of a agent type (complier probability and action outcome and the value a utility functin placed on it, which equates to rewards in our notation) and the signals of each expert.

A mechanism anounces a payment rule $p_i(\hat{s}, c, a, r)$, with p_i being the payment to expert i, with $\hat{s}$ being the reported signals, $\hat{c}$, and by c^* the reward maximizing choice given the true signals. Each experts receives their signals s_i, and makes a reports $\hat{s}_i$ to the mechanism, which outputs a chosen action c^*. The subject observes the action chosen by the mechanism $\hat{c}$, and takes a (potentially different) action a. A reward r is observed, and the mechanism makes the corresponding payments.

An experts i report is truthful when $\hat{s}_i = s_i$. A mechanism is incentive compatible when truthful reports maximize the payment experts receive and this holds for all experts. This can be in dominant strategies or in a bayesian nash equilibrium. We seek efficient mechanisms, that is $c^* = \hat{c}$ for all posible signals.

2. A SIMPLE REWARD SHARING MECHANISM

Here we explore a set of mechanisms based around the idea of sharing the rewards from the taken action with the experts.

Let us first consider a very basic direct mechanism, which will help us understand the model and the limitations that subjects freedom imposes. A simple mechanism that is strictly incentive compatible (in equilibrium) for an expert to reveal their signal when subjects are utility maximizers is to give all experts a share of the reward. Equivalently (in expectation) run a lottery after the reports are submitted that assigns a share of the reward to the expert that wins.

In this case the signals may be arbitrary, the payment rule is invariant to them, and defined as $\pi_i = (\alpha/N)r$ with $\alpha < 1$. Note that we are not defining the payment rule to be conditional on the subjects actual action in any way.

2.0.1. Limitations. This mechanism is problematic in a number of ways

- it pays out even when experts signals are useless.
- the entry of useless experts reduces the payment received by valuable ones or the budget needed by the mechanism.

Somewhat more formally we say a expert is useless, if where the expert able to force the action to be his choice, he would not improve the expected reward beyond picking the action that is the max expected utility in the common prior.

This motivates considering the next mechanism.

3. FREEDOM AND LIMITS TO DEFIANCE

If we consider subjects with potentially unrestricted freedom, i.e. subjects are not necesarily expected reward maximizers, then the mechanism is not necesarily truthful, even for a single expert.

EXAMPLE 1 (TOTAL FREEDOM YIELDS BAD INCENTIVES FOR A SINGLE EXPERT). *For example, consider a subject that has a prejudice in favour of treatment A, and so will take it unless the expected reward of treatment B is more than 10 utils highers. Consider a expert with access to a signal wich is an unbiased estimate of the rewards over both actions, receives a signal that B reward is higher by 9 utils. The expert is strictly better off reporting that his signal has B as more than 10 better.*

EXAMPLE 2 (SUBJECTS WHO DEFY EXPERTS YIELD PERVERSE INCENTIVES). *More generically we can consider of an agent who does the oposite of what maximizes reward (the min decision rule), this creates incentives for experts to flip the ordering that their signals imply.*

Thus, it can be seen that providing unrestricted freedom in the broadest sense to subjects makes the setting hopeless.

One restriction on subjects freedom that is sufficient to allow (in the single expert setting) dominant strategy truthful mehcanism is *no defiers*: assuming that all subjects are either utility maximizers (compliers) or have some action they will take irrespective of the reports of the experts (always takers or never takers in the binary setting).

This preserves substantial freedom, in that it is not required the subject known a priori that they will compliers, their realization of the always taker type could come after the mechanism has been run and before the action is taken.

This can be somewhat relaxed, since the needed property it implies is that the mechanism choosing an action with higher expected reward if actually carried out leads to the expected reward conditional on the chosen action to be higher. Formally the condition that is sufficient (**[NDP**: necesary? **]**)

$$E[^*|c_i] \geq E[r^*|c_j] \text{ if } E[^*|a_i] \geq E[r^*|a_j] \forall j \neq i \in K$$

Note that under this assumption existing mechanisms for the many expert setting are not incentive compatible unless subjects can credibly commit to using a randomized strategy which places positive mass on dominated action (i.e. there are no ex-post incntive compatible mechanisms based on proper scoring rules and voiding given a no defiers assumption):

There is some extra flexibility more than no defiers which we should use, for example differential compliance rates for different actions, the mechanisms selects the action that most raises the reward not the one who has the highest reward (for example an action that has the second highest reward but 1% noncompliance might be chosen instead of one that has 99% noncompliance and the highest reward. No defiers buys us the extra part that chosen is the highest reward action, so makes life easier for the subject (if the noncompliance depended on the signals the expert receives a utility maximizing decision maker would have to consider other decision markets noncomplaince characteristics in the prior before deciding how to interpret the chosen action; this removes such ambiguities and doesn't seem to loose too much generality)

4. SCORING RULE AND VOID DECISION MARKET MECHANISMS

An attempt to adapt Market Scoring Rules prediction markets to the decision setting (see background chapter for details) which might be termed Scoring-rule and Void Decision Market Mechanisms (we will abreviate to SVDM). The basic idea is to use a sequentially shared proper socring rule to run a prediction market for the value of the reward conditional each possible action, and to void the markets for the non-chosen actions.

EXAMPLE 3 (WITH A UTITLITY MAXIMIZING SUBJECT AN EXPERT CAN BENEFIT FROM EXAGGERATING TH
*Example from [**?**]*

[NDP: Having the first theorem in a paper not be one of the main theorems of the paper seems less than ideal, maybe we can instead state the following corollary which reexpresses it in our terms? **] [DBA**: I don't understand the words before "less than ideal". **] [NDP**: Sorry, id acidentally cutted off some text there. Now is fixed. **]**

Further, an immediate corollary to a theorem X of [**?**] which states that Randomized decision rules with full support are necesary for incentive compatible SVDM mechanisms:

If a subject is free to either maximize their expected utility at decision time, or can have probability zero of taking some actions, then a SRDM is not IC for them.

5. EFFICIENT OPTIMAL DECISION ELICITATION

We now consider a VCG style mechanism for the optimal decision elicitation problem. The sense in which it is VCG is that it is efficient (always chooses the action whose choice maximizes rewards, note this can be different form the actual action which if carried out with perfect compliance would do so).

All agents know a common prior that gives a proper prior probability distribution of rewards over all posible signals, chosen actions and actual actions.

MECHANISM 1 (EFFICIENT VCG STYLE MECHANISM). *For each expert i, compute the optimal action to choose without using that experts report (i.e. marginalizing over the prior on signal i signal), denote this action $\hat{c}_{-i}$ and it's expected reward $\hat{S}_{-i} = \bigcup \hat{s}_j \forall j \neq i \in N$ then the expected reward given the others reports is:*
$E[\hat{r}_{-i}] = E[r|\hat{c}_{-i}, \hat{S}_{-i}]$
The payment rule is announced as follows:

$$\pi_i = \begin{cases} r - E[\hat{r}_{-i}], & \textit{if } \hat{c}_{-i} \neq c \\ 0, & \textit{otherwise} \end{cases}$$

Each expert reports their signal, the mechanism calculates the action that maximizes the reward c, and chooses it. After observing the actual action taken a and its reward r, the corresponding payments are made, this concludes the mechansim.
Note that conditional on the action taken and the payments do not depend on the reported signal of agent i if we hold fixed the chosen action.

LEMMA 2. *the only way a experts reports affects their payments is via the chosen action.*

PROOF. **[NDP**: can we formalize the above lemma argument into a proof? It is that the derivative of π_i wrt to the singal s_i is zero if we hold c fixed, but ofcourse signals are discrete so derivatives are not hte right thing. **]** □

EXAMPLE 4 (REDUNDANT EXPERTS). *If you have two experts and they both observe and report the same signal, they both get outcome zero.*

LEMMA 3. *Useless experts cannot profit from participation*

PROOF. This is inmidiate from how the payment rule and useless experts are defined, since by definiion there is no information that can be reported to raise r and that is all that can lead to higher rewards. □

5.1. Incentive Compatibility and Efficient

DEFINITION 1 (BAYES-NASH INCENTIVE COMPATIBILITY). *A BNE where all experts truthfully report their signals.*

There is a equilibrium where all experts maximize their payment by maximizing the probability that the action with the highest reward is taken; if all other reports are truthful and under a no defiers assumption then this is optimized by being truthful.

PROOF. For a given expert i let all other experts report truthfully, then substituing in the true signals into the reports the expert faces

$$\pi_i = \begin{cases} r(a) - E[r|c_{-i}, \bigcup s_j \forall j \neq i \in N], & \text{if } \hat{c}_{-i} \neq c \\ 0, & \text{otherwise} \end{cases}$$

thus the only thing his report will affect in his payment is $r(a)$ via the choice of a, given all other experts are truthful then by construction if he is truthful this will select the a that maximizes r. Thus truthfulness is a BNE.

□

5.2. Potential Negative Expetation for Subject

Payments can be more than the improvement in the reward the mechanism creates **[NDP**: This is standard of VCG style stuff; we are getting efficiency and truthfulness, so the payments dont work out or the mechanism isnt voluentary **]**.

EXAMPLE 5 (EXPENSIVE EXPERTS). *Let there be two experts, two actions α, β and four equal probability states of the world, A, B, C, D, with payoffs to the actions correspondingly as $(0,1), (0,-1), (-1,0), (1,0)$. Let one expert oberve that they are in either A, B or C, D and the other B, C or A, D. A uniform random strategy achieves an expected value of 0. An optimal strategy with access to a single expert (either one) achieves a value of 0. Using the signals of both we can achieve reward 1. Thus the VCG payments are 1 to each, and the mechanism improves the subjects welfare by 1 while making 2 in payments, thus -1.*

6. RATIONAL ENTRY

As presented the VCG style direct mechanism might have higher payments than the benefits it provides in terms of higher rewards (if there are states of the world with sufficiently high reward differentials and a sufficiently large number of experts with complementary signals). Thus the subject might be better off not participating a priori. Here we introduce a variation on the mechanism that re-scales payments using the prior so a subject a priori wishes to parcitipate in the mechanism.

MECHANISM 4 (EFFICIENCY AND RATIONAL ENTRY).
Let $E[r|c(s)]$ denote the expected reward obtained from choosing the optimal action using all signals if they where reported truthfully, and $E[r|c(P)]$ the expetec treward from picking the action using only the prior and no singals. Thus $E[r|c(s)] - E[r|c(p)]$ is the expected surplus from using the information in a truthful mechanism relative to using only the priors. For a efficient mechanism to be rational apriori for the subject to implement it must be that $E[r|c(s)] - E[r|c(P)] > E[\sum \pi]$ where π is either the result of a nash equilibrium or in dominant strategy (i.e. we set all elements of $\hat{s}$ jointly to maximize $\sum \pi$, even ifthis does not result in a equilirium for the experts).
Note that a constant afine transformation of the rewards preserves the incentives. Denote by $alpha$ and $beta$ two constants that recale the reward part of the payment.

$$\rho_i = \begin{cases} \alpha + \beta r(a) - E[r|c_{-i}, \bigcup s_j \forall j \neq i \in N], & \textit{if } \hat{c}_{-i} \neq c \\ 0, & \textit{otherwise} \end{cases}$$

Before the payment rule is anounced the two constants can be set given the prior such that
$0 =< E[\sum \rho_i] <$ this is an expectation over the true

LEMMA 5. *the rescaling component of the mechanism cannot be gamed*

PROOF. As they are set apriori based on expectations only. both alpha and beta are invariant to any reports the experts make. Since the transformation is afine, whatever report maximizes payoffs remains unchanged. □

the experts have to have the potential for negative payoffs (even through the expected payoff in equilibrium is positive). Since the mchanism stoping the payments around zero would change the expectation.

DEFINITION 2 (DOMINANT STRATEGY TRUTHFULNESS). *ftw*

The incentive compatibility here is only in one nash equilibrium, and not in dominant strategies. In particular, if another expert does not report their signal truthfully, an expert might maximize the reward by making a countervailing report that is not truthful, to cancel out the other untruthful expert.

EXAMPLE 6 (MECHANISM IS NOT DOMINANT-STRATEGY INCENTIVE COMPATIBLE). *Consider a setting with a reward maximizing subject with probability 1, two possible actions, where two experts know with certainty that the rewards are* $1, 2$ *and one of them reports* $1, 0$*. A truthful report would lead to choosing the first action (the average of the two reports) with rewards of* 1 *while a report of* $0, 2$ *would cause the max to select the seccond action and obtain a reward of* 2*.*

Also note that the payout to the in this example to the expert is positive it can triviall be made negative, so the expert does not make the countervailing report as it wouldnt maximize their payment; for example if the untruthful agent reports was $3, 0$.

CONJECTURE 1. *No mechanism can get around this (intuition; we are tyring to get as close to vcg as possible, and i dont see a way to get closer).*

[DBA: proving a negative is hard; e.g. this paper is circumventing Chen's negative result **]**

6.1. A sufficient condition for dominat truthfulness in reports

in general the mechanism is not dominant truthful, since a trickster agent who inverts their reporrts would make the other agents try to compensate if possible, as seen above. If the action chosen depends only on the signal of the highest type, (the type whose signal without using he other signals implies the highest reward for the optimaly chosen action)

To see this, note that bidding the valuation conditional on their signal and winning is domminant strategy in the

6.2. Noncompliance and Incentive Compatibility

For the same reassonsas before there is no dominant strategy incentive compatibility.

A interesting question that arrises is how to handle the case where the action that is optimal given the reports is not the actual chosen action. Three natural strategies are:

- make payments irrespective. **[DBA**: I realized here that you're using (1) received payoff instead of (3) expected payoff. This seems problematic. E.g. if noncompliance leads to lower reward (which is presumably to be expected if experts know more than subject) then experts' payouts take a hit based on subject's capriciousness. This is especially problematic if payouts are negative due to subject's noncompliance. **] [NDP**: I initially tried setting to void market if action nto followed, but this creates perverse incentives, where the subject might not follow the advice and receive a lower payoff to avoid payment. While your statements abotu the experts are true they (A) the prior taking into account noncompliance makes these expectations workable, you take into account the fact that some advice might not be followed with higher probability. (B) the *no defiers* (terrible name, sinc eit is even stronger than that, uniform noncompliance might be a better term), is a sufficient condition to rule this out. **]**
- make the payment for the βr_a but do not require the payment of b_{i-1}.
- make no payment, and require no payment.

Only the first is incentive compatible. TODO: formalize. Sketch: a crucial property is that w separate bids and reports, this lets the reports be truthful while the bids are shaded (as is standard in common value seccond price auction). Further, it means that noncompliance (given the no defiers assumption) **[DBA**: maybe I'm missing something. But if the subject has a strong tendency to pick low-reward actions, then the experts are disincentivized from providing any advice at all, since they will likely receive negative payouts **]** **[NDP**: The prior is over the noncompliance (what you are terming it the tendency to pick low actions here) **]**

If we do not require the payment, it can incentivize experts to report in such a way as to encourage non-compliance, since this would increase their payout. **[DBA**: example **]** If we void the payment completely this can have the oposite effect, biasing experts away from reporting the **[DBA**: Is this paragraph dependent on agent's having a model of the subject's noncompliance? E.g. If subject's noncompliance is random (ignore advice with probability p and then pick action uniformly), are the second and third options incentive compatible? **]**

7. THE SIGNAL REPORTING VCG MECHANISM CAN WORK EVEN WHEN UTILITIES ARE NOT PART OF THE COMMON PRIOR

the mechanism, but not necesarily the experts, needs access to the utility function of the decision maker and the prior over which expectations are taken

[DBA: does the utility function of the decision maker (along with prior) "explain" its noncompliance? This is somehow a conceptual question, not sure if it makes a difference to the meat of the paper. **]**

for a cancer example; the chemiotherapy specialist might explain the likely side effects conditional on their knowledge of the specifics of the patient (the rapidly dropping cost of sequencing a genome means this) without needing to understand the relative value the patient places on discomfort relative to life lengthening, while the surgeon might explain the (again subject tot he specific characteristics of the subject) risks he foresees in the surgury (in contrast to the baseline risks of the populatin that gets the procude which is what cna be observed, a reasoable prior) for that patient, without needing to udnerstand the risk aversion of the patient which would be necesary for calculating their utility.

the subject (oracle style) can then be queried for their chosen action and expected utility for different subsets of the signals (for every expert we need to query what the action (and the expected payoff) would be without that experts signal having been reported)

contrast the information structure in the cancer example and its non separability

8. A BIDDING MEHCNAISM

In many practical applications of interest, it might be that the signals can't be practically reported. For example, because while the expert knows something when they see it, they do not have a vocabulary to unambigously express it.

If we replace the reports in the direct mechnaism by bids, and we allow the winner of the auction to pick the action, we have a mechanism that doesnt make any strong assumptions beyond that the agents understand the rules of the game.

For some information structures this still aggregates information well, for example:

LEMMA 6 (SPECIALIST EXPERTS). *if each expert knows the exact outcome conditonal on one action and knows he doesnt know what happens under other actions, and there is at least one expert that is informaed about each action.*

For other information structures it does not work, for exmaple

EXAMPLE 7. *the classic example without separability that fails in the bayesian games so broadly (TODO Add cite) also fails here.*

8.1. Interpreting bids

the bids reflect the expected value of the mechanism choosing a particular action, not of that action being taken. **[DBA**: I'm confused. In previous section, "each expert knows the exact outcome conditonal on one action", and now its conditional on action being chosen, not taken. These sentences sentences seem to be at odds. **]** Example: if half are allways takers, the bid effectively waters down the true effect by half. **[DBA**: So the bids made by experts depend on the expert's model of the subject's behavior? **]**

further, due to the winners curse, it is the expected effect conditional on having won the auction. **[DBA**: can you write down how expected conditional on having won is computed? **]**

[DBA: At this point, the interpretation of the bid has been qualified twice, and I'm lost **]**

8.2. A sufficient separability condition for efficiency

if each expert knows the value of a action the MSR and Void mechanism in the style of Hanson fail to be myopically incentive compatible when agents always follow the action whose price is maximal in the market ([**?**])

9. TRADEOFF BETWEEN DOMINANT STRATEGY INCENTIVE COMPATIBLE AND EXPRESSIVITY

9.1. A sufficient condition

when the expert who won the auction selects and action, particularly in the early rounds she might be doing it becuase he wishes to explore, and thus the incentives

10. TWO-SIDES

A sequence of subjects choosing each choosing from the same fixed set of action which only affect them directly, as in our frst model, while eliciting advice for those subjects from a fixed set of experts, as in our seccond model. At each step a subject arrives, experts recieve signals conditional on the subject, reports of experts are submited to the mechanism, the subject observes the action chosen by the mechanisms, picks an action and receives a reward, the mechanism then uses the observed reports, action and reward to assign payments to the experts.

Bounded regret algorithms from the compliance awareness can be seen as addressing the special case where the experts signals are known apriori to not be informative, and thus only the experience can be learned from. Our one-decision mechanism is the special case for T=1, thus there is no role for exploration, since there are no future decisions in the game. The compliance-aware algorithm can be used for initial rounds where a reserve price has not been met.

The mechanism consists of a seccond price auction of the right to select the sequence of chosen actions for all remaning periods, and a share of the resulting rewards.

For the special case where knowledge is fully substitutable accross at least two experts informed per actions, it incentivizes the experts to carry out the optimal amount of exploration.

While existing results for bayesian incentive compatible exploration and its limits still apply on the subjects part if they are all rational and this is common knowledge. For medical applications this is unlikely to be the case, with a diverse set of motivations for agents making them some far from rational, and that this will be common knowledge ammong the rational agents. Somewhat paradoxically

11. A SEQUENCE OF ONE SHOT MECHANISMS DOES NOT EXPLORE

Attempting to run the single action elicitation mechanism at the start of each period does not preserve its incentive compatibility or its efficiency. The single shot mechanism creates incentives that are equivalent to a greedy policy that only exploits and does not exploit. If we reveal the outcome of the round to all experts, no agent can internalize the information revelation that comes from the outcome of a action. On the other hand, if we only allow the winner of the bidding on a given round to have access to the subject's outcome, then the mechanism is extremely data-inefficient.

12. A DIRECT MECHANISM WITH STRONG COMMON KNOWLEDGE

We use the bidding rules of the direct mechanism from chapter 2 but we add a reserve price and instead of the rights to a share of the periods payoff, you are buying the right to a share of all future periods payoffs (when there is only one period left the two coincide). The rights are chained so the reserve price is set by the winner of previous period, and if no sale then the

13. A PRACTICAL BID BASED MECHANISM

instead of reporting signals the experts send bids and actions they would pick if the winning bid. The actions are now plans for all posible contigent periods of chosen or actual action.

As the mechanism (representing the collective of subjects ideally? what about the experts welfare footnote)

14. EFFICIENCY OF THE MECHANISM

Bayesian arguments (?) bound the best rate at which we could learn if we had accss to the experts signals (i.e. given a correct prior, how fast do we identify the optimal action? how does this vary with the concentration of the prior? , the direct mechanism is equivalent to learning it, for example under Optimally Confident UCB can we give numbers? There is the 79 bounds on the minimax rate in bnadits of gittins, does that work? note dificulty of simulations, since it involves solvng for strategies that might exploit the specific algo we propose as the initiator.

14.1. One Action per Expert is not enough for incentive compatibility

even when each expert observes the signal that is informative of just one action, versus when the expert observes a statistic about an action (say the mean of a given arm) which can still be informative of the mean of other arms. Note in the one shot setting this doesnt hold, and ne qctin per expert is sufficient, this is becuase the seccond price auction is effective in cuting off manipulation there, but here the effect on uture beliefs and thus bids in later rounds is not cut (while in one shot it is inexistent)

14.2. A sufficient condition; Expert signals only informative about one action

Both the subjects rewards are idependent conditional on the treatment, and the treatments are conditionally indepedent.

if they are not independent one can consider a situation in which if you take my bid as truthful I can gain by lying to confuse your bid in the next period, (and i buy it in the period before or during tricking you)

15. INCENTIVE COMPATIBILITY FOR SUBJECTS

One natural question given the bayesian incentive compatible bandit exploration literautre, is wether these mechanisms can work when all subjects are expected utility maximizers. If the experts bring enough information to bear, the answer is yes, and it can be so without hidding past subjects outcomes, if the experts bring sufficient information. Note however,that there are intermediate situations