Handout on Optimal Contracts

Kevin Hasker

There is one problem in Bayesian games that is so general that it can be called "classic." In this problem there is an informed party and an uninformed party that want to interact. The uninformed party and some types of the informed party would like the interaction to be a function of the informed party's *type*, $\theta \in \Theta \subseteq \mathbb{R}_+$. In order to achieve this either the uninformed party may take an action or the informed party has a two dimensional strategy. In either case the equilibrium can be specified as a (reward,cost) vector, implicitly the reward and cost will be a function of the type of the informed party. We have looked at several examples of this type of model. The first one we looked at was second degree price discrimination, where the equilibrium was $(q(\theta), F(\theta))$—$q(\theta)$ is the quantity the person received and $F(\theta)$ was the fee they had to pay. Another one we looked at was the insurance market $(D(\theta), p(\theta))$—$D(\theta)$ is the dividend paid in case of an accident, $p(\theta)$ is the premium that must be paid. The one we have focused on most is $(w(\theta), e(\theta))$—$w(\theta)$ is the wage paid if the person expends the effort or gets the education $e(\theta)$, which has a positive cost for the potential worker.

How do we analyze such models? Frankly the possibilities are enormous. In a general sense a mechanism in such a model could be a $\Gamma = \{A, \Theta, S, g\}$ where A is a (finite) outcome space, Θ is the afore mentioned (finite) space of types, $S = S_1 \times S_2$ is a (finite) strategy space for each party involved, $g : S \to A$ is an outcome function. Since we are free to choose S and g there is literally an infinite number of possibilities regardless of the size of A and Θ. The simplest set of mechanisms is the *direct mechanisms*, $\Gamma_d = \{A, \Theta, \emptyset \times \Theta, F\}$ where $F : \Theta \to A$ is the social choice rule we want to implement, the uninformed party (player 1) has no strategies, and the informed party merely announces their type, $\rho \in \Theta$. The analysis is greatly simplified by the *revelation principle.*

Theorem 1 *Assume that F is implemented by some Γ in (weakly) dominant strategies. Then there is a direct mechanism that implements F in (weakly) dominant strategies.*

To see a proof of this I refer the reader to Proposition 23.C.1 in Mas-Colell, Whinston, and Green.

The benefits of this theorem really can't be understood in this abstract setting. In order to appreciate its power you need to see how it is applied. To do that I need a more precise model, but one which can be reworked to handle any of the above situations. Assume that the two parties are a firm (the uninformed party) and a worker (the informed party). We assume that $w \in \mathbb{R}_+$ is the choice variable of the uninformed party and $e \in \mathbb{R}_+$ is the choice variable of the informed party. The payoffs of the parties are:

$$\begin{aligned} \pi(w, e, \theta) &= R(e, \theta) - w \\ u(w, e, \theta) &= \max\{w - c(e, \theta), r(\theta)\} \end{aligned}$$

where π is the payoff of the firm (uninformed) and u is the payoff of the informed (the worker). $R(\cdot,\cdot)$ is the revenue generated by a worker, $c(\cdot,\cdot)$ is the cost of the signalling variable given the type of the worker, and $r(\cdot)$ is the worker's outside option.

We make the following assumptions about cost $c_e > 0$, $c_{ee} > 0$, $c_\theta < 0$ and $c_{e\theta} < 0$. The last assumption is the *single crossing property* and guarantees that the indifference curves of the workers will cross at most once. The derivatives of $R(\cdot,\cdot)$ are less important, but we will generally assume $R_e \geq 0$, $R_\theta \geq 0$, $\max\{R_e, R_\theta\} > 0$, $R_{ee} \leq 0$, $R_{\theta\theta} \leq 0$, $R_{e\theta} > 0$. The only assumption we will make about $r(\cdot)$ is that it is differentiable, notice we have implicitly assumed that $c(\cdot,\cdot)$ and $R(\cdot,\cdot)$ are twice continuously differentiable. These assumptions are all fine since these are primitives in our model, though one may want to derive them from a more general model. In this general derivation these conditions may not be true.

An equilibrium is then a pair $(w(\theta), e(\theta))$, and the primary impact of the revelation principle is that we can write down the characterization of this equilibrium as three constraints:

$$\begin{aligned}
\text{Incentive Compatibility (IC)—}\theta &\in \arg\max_{\rho\in\Theta} w(\rho) - c(e(\rho),\theta) \\
\text{Individual Rationality (IR)—}r(\theta) &\leq w(\theta) - c(e(\theta),\theta) \\
\text{Profitability—}0 &\leq E_\theta\left[R(e(\theta),\theta) - w(\theta) \mid (w(\theta), e(\theta)) = \left(w\left(\theta'\right), e\left(\theta'\right)\right)\right] .
\end{aligned}$$

We won't be paying attention to the last constraint for a while. Right now we will be focusing on the first two, the IC (incentive compatibility) and IR (individual rationality) constraints. First, however, we want to establish a basic property of $w(e)$. This is the implicit function determined by $(w(\theta), e(\theta))$, in other words we want to know how w must change with respect to e, can we have, for example, one of them increasing and the other decreasing?

Lemma 1 *Since $c_e > 0$ we can assume $\frac{\partial w}{\partial e} > 0$.*

Proof. Assume not, but then

$$\frac{\partial w}{\partial e} - c_e(e,\theta) < 0$$

and no one will choose that particular $w(e)$. Thus we can ignore any part of the function where this is true. ■

Now we have established that $w(e)$ is monotonic, can we further establish that $e(\theta)$ is monotonically increasing in θ?

Lemma 2 *Since $c_{e\theta} < 0$ $e_\theta \geq 0$ without loss of generality.*

Proof. Consider θ and θ' such that $\theta > \theta'$ but $e(\theta) = e < e\left(\theta'\right) = e'$ then we know that $w(\theta) = w < w\left(\theta'\right) = w'$ by the last lemma, and by incentive compatibility we know that:

$$\begin{aligned}
w' - c\left(e', \theta'\right) &\geq w - c\left(e, \theta'\right) \\
w' - w &\geq c\left(e', \theta'\right) - c\left(e, \theta'\right)
\end{aligned}$$

$$
\begin{aligned}
c\left(e',\theta'\right)-c\left(e,\theta'\right) &= \left(e'-e\right)\frac{c\left(e',\theta'\right)-c\left(e,\theta'\right)}{e'-e} \qquad (1)\\
c\left(e',\theta\right)-c\left(e,\theta\right) &= \left(e'-e\right)\frac{c\left(e',\theta\right)-c\left(e,\theta\right)}{e'-e}
\end{aligned}
$$

$$
\begin{aligned}
& \left(c\left(e',\theta'\right)-c\left(e,\theta'\right)\right)-\left(c\left(e',\theta\right)-c\left(e,\theta\right)\right)\\
= & \left(e'-e\right)\left[\frac{c\left(e',\theta'\right)-c\left(e,\theta'\right)}{e'-e}-\frac{c\left(e',\theta\right)-c\left(e,\theta\right)}{e'-e}\right]\\
= & \left(e'-e\right)\left[\frac{c\left(e',\theta'\right)-c\left(e,\theta'\right)}{\Delta e}-\frac{c\left(e',\theta\right)-c\left(e,\theta\right)}{\Delta e}\right]\\
= & \left(e'-e\right)\left(\theta'-\theta\right)\frac{\frac{c\left(e',\theta'\right)-c\left(e,\theta'\right)}{\Delta e}-\frac{c\left(e',\theta\right)-c\left(e,\theta\right)}{\Delta e}}{\Delta\theta}
\end{aligned}
$$

$$
sign\left(\left(e'-e\right)\left(\theta'-\theta\right)\frac{\frac{c\left(e',\theta'\right)-c\left(e,\theta'\right)}{\Delta e}-\frac{c\left(e',\theta\right)-c\left(e,\theta\right)}{\Delta e}}{\Delta\theta}\right)=sign\left(\left(e'-e\right)\left(\theta'-\theta\right)c_{e\theta}\right)=+ \qquad (2)
$$

therefore

$$
\begin{aligned}
\left(c\left(e',\theta'\right)-c\left(e,\theta'\right)\right)-\left(c\left(e',\theta\right)-c\left(e,\theta\right)\right) &> 0\\
c\left(e',\theta'\right)-c\left(e,\theta'\right) &> c\left(e',\theta\right)-c\left(e,\theta\right)\\
w'-w &> c\left(e',\theta\right)-c\left(e,\theta\right)\\
w'-c\left(e',\theta\right) &> w-c\left(e,\theta\right)
\end{aligned}
$$

And we have established that θ strictly prefers the outcome (w',e') to (w,e). Thus we can ignore any situation where $e_\theta<0$ since no one will choose these outcomes. ■

Algebraically this is a complicated proof. What we have to do is show a relationship between $c\left(e',\theta'\right)-c\left(e,\theta'\right)$ and $c\left(e',\theta\right)-c\left(e,\theta\right)$ based on $c_{e\theta}$. Since $c_{e\theta}$ is constant this is easy (help!) to do, but we must transform the difference into a delta format to get it right. The really complicated fraction in equation 2 is exactly the delta format we need.

Now we need to start getting rid of constraints. The way I have written it down incentive compatibility seems insanely easy to check, but for each type we have $|\Theta|-1$ conditions to check, so in total we have $|\Theta|\left(|\Theta|-1\right)$ conditions to check. If we can not simplify things more then we are basically stuck.

Lemma 3 *Since $c_{e\theta}<0$ if for $\theta>\theta'$ $u\left(w\left(\theta\right),e\left(\theta\right),\theta\right)\geq u\left(w\left(\theta'\right),e\left(\theta'\right),\theta\right)$ then it is also true for $\tilde{\theta}\geq\theta$.*

Proof. Like before let $w=w\left(\theta\right),e=e\left(\theta\right),w'=w\left(\theta'\right),e'=e\left(\theta'\right)$ then we have:

$$
\begin{aligned}
w-c\left(e,\theta\right) &\geq w'-c\left(e',\theta\right)\\
w-w' &\geq c\left(e,\theta\right)-c\left(e',\theta\right)
\end{aligned}
$$

The difference of methodology in this proof is that I am going to write w as a function of e, and drive $e' \to e$. Notice that I am assuming differentiability of $w(e)$, but if you don't want to make this assumption you can work through the last proof again, this proof is no different. Given this we can see that the above inequality is that:

$$w_e \geq c_e(e, \theta)$$

and now look at the function:

$$\begin{aligned} sign\left(\left(\theta - \tilde{\theta}\right)\frac{c_e(e,\theta) - c_e\left(e,\tilde{\theta}\right)}{\theta - \tilde{\theta}}\right) &= sign\left(\left(\theta - \tilde{\theta}\right)\lim_{\tilde{\theta}\to\theta}\frac{c_e(e,\theta) - c_e\left(e,\tilde{\theta}\right)}{\theta - \tilde{\theta}}\right) \\ &= sign\left(\left(\theta - \tilde{\theta}\right)c_{e\theta}\right) < 0 \end{aligned}$$

where this is all true because $c_{e\theta}$ has a constant sign. Thus

$$c_e(e,\theta) \geq c_e\left(e,\tilde{\theta}\right)$$

and we are done. ■

This has allowed us to get rid of the lower triangle of constraints, to make it concrete how many it is easier if we assume that the space of types is finite. Throughout these lemmas it is easier to treat it as infinite, but let us assume that there are only K types to clarify how much of an improvement this is. Previously we has $K(K-1)$ constraints. Now for each type we only need to check whether k prefers his allocation to $k-1$. Thus for the highest type we only have one constraint, for the next highest only two (one down and one up) for the next three (one down, two up) and so on. So we have $\sum_{j=1}^{K} j = \frac{1}{2}K(K+1)$ constraints left. Or we have dropped $\frac{1}{2}K(K-3)$ constraints. Quite a dramatic improvement, but still not enough. Fortunately we can use the exact same proof one more time, and we will get rid of the rest of the constraints.

Lemma 4 *Since $c_{e\theta} < 0$ if for $\theta < \theta'$ $u(w(\theta), e(\theta), \theta) \geq u(w(\theta'), e(\theta'), \theta)$ then it is also true for $\tilde{\theta} \leq \theta$.*

Notice that we have essentially reversed the order of θ and θ', so we also reverse the order of θ and $\tilde{\theta}$. This guarantees that $(e - e')\left(\theta - \tilde{\theta}\right) \leq 0$, which is what we need to get the inequalities right.

Proof. We have:

$$w - w' \geq c(e,\theta) - c(e',\theta)$$

now the right hand side is a statement about the marginal impact on costs of e. Since $e' \geq e$ it is a negative value. We know that

$$c(e,\theta) - c(e',\theta) \geq c\left(e,\tilde{\theta}\right) - c\left(e',\tilde{\theta}\right)$$

because $\theta \geq \tilde{\theta}$ and $c_{e\theta} < 0$. In other words the right hand side difference is larger in absolute value than the left hand side difference because θ has decreased, since both sides are negative we get the order above. This concludes the proof. ■

So now we have gotten rid of as many incentive compatibility constraints as we can. To be precise the highest and lowest type only have one constraint, the rest have two, so we have $2(|\Theta| - 1)$ constraints. Comparing this to before we have managed to get rid of $\frac{2(|\Theta|-1)}{|\Theta|(|\Theta|-1)} = \frac{2}{|\Theta|}$ of our constraints, or we have an order of magnitude improvement. This is incredible, and absolutely necessary to make the problem tractable. Now, can we get rid of any of our individual rationality constraints? The following Lemma tells us that in general we can.

Lemma 5 *Let $u^*(\theta) = u(w(\theta), e(\theta), \theta)$, then since $c_\theta < 0$, $u^*_\theta > 0$.*

Proof. Assume $\theta > \theta'$, then

$$w(\theta) - c(e(\theta), \theta) \geq w(\theta') - c(e(\theta'), \theta) > w(\theta') - c(e(\theta'), \theta')$$

since $c_\theta > 0$. ■

How does the revelation principle work? By "paying" people to announce they are a higher type. Since high θ is better higher θ must get a strictly better payoff. This is essential to how the mechanism works. Notice that the derivative, u^*_θ, may be positive infinity.

Now this can tell us exactly when we can ignore individual rationality constraints, and if we can not what we know then.

Corollary 1 *If for given $(w(\theta), e(\theta))$, $u^*_\theta > r_\theta$ then we can ignore all individual rationality constraints except for the lowest type that participates. Specifically if $r(\theta) = r$ then there is a lowest type $\underline{\theta}$ for whom $u^*(\theta) = r$ and no individual with $\theta < \underline{\theta}$ participates.*

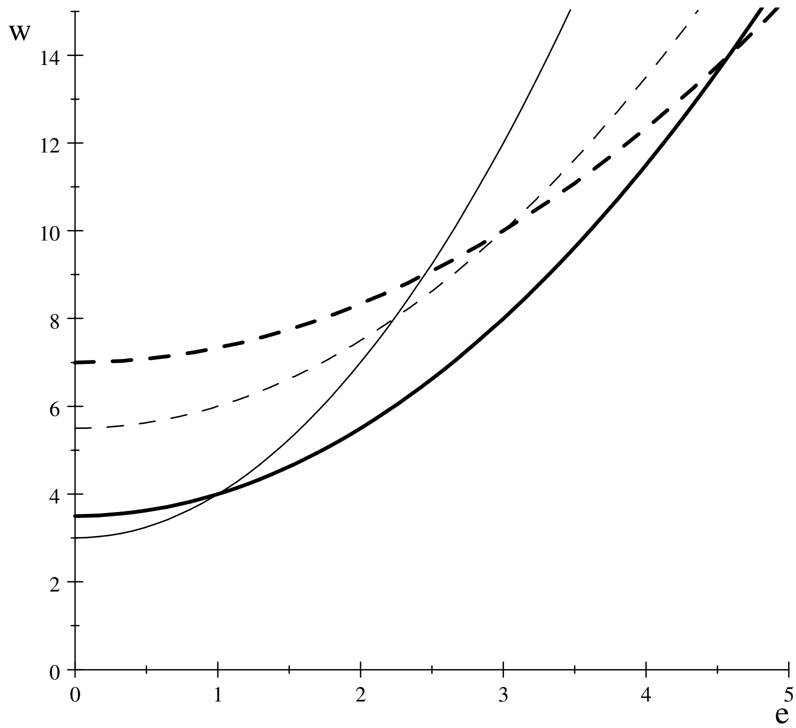
In general there will be a set of types $\underline{\Theta} \subseteq \Theta$ and a second set $\Theta_p \subseteq \Theta$ such that for $\underline{\theta} \in \underline{\Theta}$ $u^*(\underline{\theta}) = r(\underline{\theta})$, if $\theta \in \Theta/\Theta_p$ then θ does not participate, if $\theta \in \Theta_p$ then the only relative incentive compatibility constraints are with regard to $\underline{\theta}(\theta) = \arg\max_{\tilde{\theta} \in \underline{\Theta} | \tilde{\theta} \leq \theta} \tilde{\theta}$ and $\bar{\theta}(\theta) = \arg\inf_{\tilde{\theta} \in \Theta/\Theta_p | \tilde{\theta} \geq \theta} \tilde{\theta}$.

The most important part of the corollary is when $r(\theta) = r$, which is the standard case. I include the second part merely to indicate what happens when this is not true. It is quite interesting to note that it is basically the standard problem except there are now a series of intervals. In other words if there are M such intervals we can analyze each problem independently. Of course this depends on $(w(\theta), e(\theta))$, so there may be other solutions with more or less intervals. This is why the standard assumption is that $r_\theta = 0$, and of course we can easily generalize this to $r_\theta \leq 0$.

Notice that up to now it has often been more convenient to simply let Θ be an interval from $\mathbb{R}_+$, many of the proofs above would be more complicated if we had to consider only discrete intervals for θ. Now, however, it won't be convenient to continue with this assumption so we will change it. Thus assume the set of types is finite, or $\Theta = \{\theta_1, \theta_2, \theta_3, ...\theta_K\}$ and for $k < K$ $\theta_k < \theta_{k+1}$.

An equilibrium can then be written as $(w_k, e_k)_{k=1}^{K}$, where $w_k = w(\theta_k)$ and $e_k = e(\theta_k)$. Further assume from now on that $r(\theta) = r$, it has been fun talking about the more general case, but as the corollary shows we can almost think of it as this case and just move on. Given the last lemma this means we can ignore all of the individual rationality constraints except for the lowest type, and without loss of generality we can assume that θ_1 participates.

Now we can begin to analyze this problem graphically. First we choose a level for (w_1, e_1), given this choice we have an upper and lower bound for (w_2, e_2) given by the indifference curves of type 1 and 2 at (w_1, e_1). Given these constraints we can choose a (w_2, e_2), and then we have two constraints for (w_3, e_3) and so on. In the graph below, $u = w - \frac{e^2}{\theta}$,$\theta_1 = 1$, $\theta_2 = 2$, $\theta_3 = 3$ and $\{w_1, e_1\} = (4, 1)$, $(w_2, e_2) = (10, 3)$. (w_2, e_2) had to be between the two solid lines ($u_1 = u_1(4, 1)$ and $u_2 = u_2(4, 1)$). (w_3, e_3) has to be between the two dashed lines ($u_2 = u_2(10, 3)$ and $u_3 = u_3(10, 3)$). In each case the indifference curve of the higher type is a heavier line.



In other words for each utility level for one type we get a cone of possible utility levels for the next higher type, and so on. So what is the space of viable contracts? Well first of all we will look at *completely separating contracts*, or contracts where for every θ, $(w(\theta), e(\theta))$ is different.

In this case the contracts basically depend on:

1. The utility of the lowest type to participate (θ_1 without loss of generality).
2. Whether the upper or lower constraint is binding.

What is the possible range for the lowest participating type? The lowest is obviously r, the highest is when they capture all of the revenue that they will produce and have the optimal level for e given this constraint, or:

$$\bar{u}(\theta_1) = \max_e R(e, \theta_1) - c(e, \theta_1)$$

so $u^*(\theta_1) \in [r, \bar{u}(\theta_1)]$. Then the contract where all the lower constraints are binding is:

$$\begin{aligned}
u_1^*(\theta) &= w_1 - c(e_1, \theta_1) \\
w_1 &= u_1^*(\theta) + c(e_1, \theta_1) \\
w_2 - c(e_2, \theta_2) &= w_1 - c(e_1, \theta_2) \\
w_2 &= c(e_2, \theta_2) + [c(e_1, \theta_1) - c(e_1, \theta_2)] + u_1^*(\theta) \\
w_{k+1} &= c(e_{k+1}, \theta_{k+1}) + \sum_{j=1}^{k} [c(e_j, \theta_j) - c(e_j, \theta_{j+1})] + u^*(\theta_1) \quad (3) \\
w_2 &= [c(e_2, \theta_2) - c(e_1, \theta_2)] + u_1^*(\theta) + c(e_1, \theta_1) \\
w_{k+1} &= \sum_{j=1}^{k} [c(e_{j+1}, \theta_{j+1}) - c(e_j, \theta_{j+1})] + u^*(\theta_1) + c(e_1, \theta_1) \quad (4)
\end{aligned}$$

Which way you present this equilibrium depends on what you want to analyze. Representation 3 clearly tells us the *revelation cost*, for type θ_k this is $\sum_{j=1}^{k-1} [c(e_j, \theta_j) - c(e_j, \theta_{j+1})] + u^*(\theta_1)$. This is essentially how much they must be paid to truthfully reveal their type. Representation 4 allows us to immediately see how the equilibrium would change with *partial pooling*. In this case for some j $e_{j+1} = e_j$ and these elements would disappear from the sum, immediately implying that all pooling parties get the same wage. (I know, that is a rather obvious result, but just sit back and go "wow" for my ego, will you?)

If we assume the upper constraints are binding the solution looks very similar, except for the key difference that now it is type θ_k who is indifferent between $(w(\theta_k), e(\theta_k))$ and $(w(\theta_{k+1}), e(\theta_{k+1}))$.

$$\begin{aligned}
w_2 - c(e_2, \theta_1) &= w_1 - c(e_1, \theta_1) = u^*(\theta_1) \\
w_2 &= c(e_2, \theta_1) + u^*(\theta_1) \\
w_3 - c(e_3, \theta_2) &= w_2 - c(e_2, \theta_2) \\
w_3 &= c(e_3, \theta_2) + [c(e_2, \theta_1) - c(e_2, \theta_2)] + u^*(\theta_1) \\
w_{k+1} &= c(e_{k+1}, \theta_k) + \sum_{j=2}^{k} [c(e_j, \theta_{j-1}) - c(e_j, \theta_j)] + u^*(\theta_1) \ .
\end{aligned}$$

Now in both of these boundary contracts we still need to solve for K levels of e, but for appropriate values of e this is the space of contracts.

1 Finding the equilibria in Signalling Games and the Principle agent model.

In this section I will focus only on equilibria where no two types choose the same w and e. In signalling games this is called a *pure separating equilibrium.* In the principle agent model this is an assumption that will only be met sometimes. It very well might be that allowing for partial pooling increases the principle's profit. However that case can be solved fairly easily from the pure separating analysis.

1.0.1 Signalling Games

In signalling games we have the constraint that given the beliefs of the firms, $w(e) = E_\theta [R(e,\theta)|e]$, or each type must get it's expected benefit to the firm. This is justified by assuming that the firms are perfect competitors for the (apparently monopolistic) worker. Given this constraint $u^*(\theta_1) = \bar{u}(\theta_1)$ and we have the following solutions for the upper and lower bounds:

$$R(e_{k+1},\theta_{k+1}) = c(e_{k+1},\theta_{k+1}) + \sum_{j=1}^{k} [c(e_j,\theta_j) - c(e_j,\theta_{j+1})] + \bar{u}(\theta_1)$$

$$R(e_{k+1},\theta_{k+1}) = c(e_{k+1},\theta_k) + \sum_{j=2}^{k} [c(e_j,\theta_{j-1}) - c(e_j,\theta_j)] + \bar{u}(\theta_1)$$

we solve these constraints sequentially, starting with the lowest education level yet to be determined. For the lower bound the first solution is characterized by:

$$\min_{e_2} (R(e_2,\theta_2) - c(e_2,\theta_2)) \text{ s.t. } R(e_2,\theta_2) - c(e_2,\theta_2) \geq u_1^*(\theta) + [c(e_1,\theta_1) - c(e_1,\theta_2)]$$

For the upper bound we solve the problem:

$$\max_{e_2} (R(e_2,\theta_2) - c(e_2,\theta_1)) \text{ s.t. } (R(e_2,\theta_2) - c(e_2,\theta_1)) \leq u^*(\theta_1)$$

but of course there are many contracts in between each of these boundary contracts, in signalling games there are always a continuum of viable contracts, and I am not even going specify the contracts with partial pooling. It is not difficult and can be generated from this analysis.

1.1 Principle-Agent Models

Here there is a monopolistic firm and competitive workers, so because of this the lower bound contract is always the one that matters. The monopolist is not going to pay more than she absolutely has to, and this is specified by the lower bound contract. Furthermore $u^*(\theta_1) = r$ by the same reasoning.

$$w_{k+1} = c(e_{k+1},\theta_{k+1}) + \sum_{j=1}^{k} [c(e_j,\theta_j) - c(e_j,\theta_{j+1})] + r$$

And we need to set up the monopolist's objective function, letting p_k be the probability of a worker of type k, this is:

$$\max_{\{e_k\}_{k=1}^K} \sum_{k=1}^{K} p_k \left(R\left(e_k, \theta_k\right) - w_k\right)$$
$$\max_{\{e_k\}_{k=1}^K} \sum_{k=1}^{K} p_k \left(R\left(e_k, \theta_k\right) - c\left(e_k, \theta_k\right) - r - \sum_{j=2}^{k} \left(c\left(e_{j-1}, \theta_{j-1}\right) - c\left(e_{j-1}, \theta_j\right)\right)\right)$$

where I define that for $k = 1$ $\sum_{j=2}^{k} \left(c\left(e_{j-1}, \theta_{j-1}\right) - c\left(e_{j-1}, \theta_j\right)\right) = 0$. By expanding this out we can get:

$$\max_{\{e_k\}_{k=1}^K} \sum_{k=1}^{K} p_k \left(R\left(e_k, \theta_k\right) - c\left(e_k, \theta_k\right)\right) - r - \sum_{k=1}^{K} p_k \sum_{j=2}^{k} \left(c\left(e_{j-1}, \theta_{j-1}\right) - c\left(e_{j-1}, \theta_j\right)\right)$$

now we want to simplify this last double summation. This can be quite simply done by counting the number of times a particular term is in the summation. For example, how many times does

$$c\left(e_1, \theta_1\right) - c\left(e_1, \theta_2\right)$$

appear? It appears in all terms except for the first one since the summation always starts at $j = 2$. If $P_k = \Sigma_{j=1}^{k} p_k$ then the probability of this event is

$$1 - P_1 = 1 - p_1 = \Sigma_{j=2}^{K} p_k$$

. How many times does $c\left(e_2, \theta_2\right) - c\left(e_2, \theta_3\right)$ appear? Well this appears every time except for when $k \in \{1, 2\}$, and the probability of this event is $1 - P_3$. And so on. So:

$$\begin{aligned} \sum_{k=1}^{K} p_k \sum_{j=2}^{k} \left(c\left(e_{j-1}, \theta_{j-1}\right) - c\left(e_{j-1}, \theta_j\right)\right) &= \sum_{j=2}^{K} \left(1 - P_{j-1}\right) \left(c\left(e_{j-1}, \theta_{j-1}\right) - c\left(e_{j-1}, \theta_j\right)\right) \\ &= \sum_{j=1}^{K-1} \left(1 - P_j\right) \left(c\left(e_j, \theta_j\right) - c\left(e_j, \theta_{j+1}\right)\right) \end{aligned}$$

Where the last step is just a change of variables. To make this clearer let's write each term out explicitly for $K = 4$.

$$\begin{aligned}
& p_1 [(0)] \\
& +p_2 [(c(e_1,\theta_1) - c(e_1,\theta_2))] \\
& +p_3 [(c(e_1,\theta_1) - c(e_1,\theta_2)) + (c(e_2,\theta_2) - c(e_2,\theta_3))] \\
& +p_4 [(c(e_1,\theta_1) - c(e_1,\theta_2)) + (c(e_2,\theta_2) - c(e_2,\theta_3)) + (c(e_3,\theta_3) - c(e_3,\theta_4))] \\
= \; & (p_2 + p_3 + p_4)(c(e_1,\theta_1) - c(e_1,\theta_2)) \\
& + (p_3 + p_4)(c(e_2,\theta_2) - c(e_2,\theta_3)) \\
& + p_4 (c(e_3,\theta_3) - c(e_3,\theta_4)) \\
= \; & (1 - P_1)(c(e_1,\theta_1) - c(e_1,\theta_2)) \\
& + (1 - P_2)(c(e_2,\theta_2) - c(e_2,\theta_3)) \\
& + (1 - P_3)(c(e_3,\theta_3) - c(e_3,\theta_4))
\end{aligned}$$

So going back to the objective function:

$$\begin{aligned}
& \max_{\{e_k\}_{k=1}^K} \sum_{k=1}^K p_k (R(e_k,\theta_k) - w_k) \\
= \; & \max_{\{e_k\}_{k=1}^K} \sum_{k=1}^K p_k (R(e_k,\theta_k) - c(e_k,\theta_k)) - \sum_{j=1}^{K-1} (1 - P_j)(c(e_j,\theta_j) - c(e_j,\theta_{j+1}))
\end{aligned}$$

Now if we define $(c(e_K,\theta_K) - c(e_K,\theta_{K+1})) = \Delta$ (the exact value does not matter) then we can also rewrite this as:

$$\max_{\{e_k\}_{k=1}^K} \sum_{k=1}^K p_k \left(R(e_k,\theta_k) - c(e_k,\theta_k) - \frac{(1 - P_k)}{p_k} (c(e_k,\theta_k) - c(e_k,\theta_{k+1})) \right)$$

and $c(e_k,\theta_k) + \frac{(1-P_k)}{p_k} (c(e_k,\theta_k) - c(e_k,\theta_{k+1}))$ is sometimes referred to as the "virtual cost" of type $k \in \{1, 2, 3, ...K\}$. Analyzing this objective function gives us the following proposition:

Proposition 1 *Only the highest type gets the optimal amount of e, all lower types get too little e. Only the lowest type gets paid the competitive wage, r.*

Proof. The first order condition for type k is:

$$R_e(e_k,\theta_k) - c_e(e_k,\theta_k) - \frac{(1 - P_k)}{p_k} (c_e(e_k,\theta_k) - c_e(e_k,\theta_{k+1})) = 0$$

if $k = K$ then $(1 - P_K) = 0$ and this is the unconstrained optimum for this type. Otherwise $\frac{(1-P_k)}{p_k} (c_e(e_k,\theta_k) - c_e(e_k,\theta_{k+1})) > 0$ and

$$R_e(e_k,\theta_k) - c_e(e_k,\theta_k) > R_e(e_k,\theta_k) - c_e(e_k,\theta_k) - \frac{(1 - P_k)}{p_k} (c_e(e_k,\theta_k) - c_e(e_k,\theta_{k+1}))$$

thus the level of e_k is below the optimal level. The second statement is obvious from the wage function in equation 3. ■

The former result is known as "no distortion at the top." Notice the dramatic difference between the two models that in a signalling model it is the worst workers who get the optimal amount of e, in the Principle-Agent model it is the best workers.