Online Appendix to 'Interpersonal Authority in a Theory of the Firm'

Eric Van den Steen*

April 23, 2009

I. Proofs

Proof of Proposition 2: I will start again with a partial backwards induction. Consider stage 3b. Upon a success, no participant quits in (any) equilibrium (given $\alpha_k^i \geq 0$ and Nash bargaining). Upon a failure, a participant P_i quits R_k iff $w_k^i < 0$ (given that any private benefits are sunk). It follows that (since $w_k^1 + w_k^2 = B$) if some $w_k^i \notin [0, B]$, then R_k is terminated upon failure, so that the w_k^i are then paid only upon success. This case is equivalent to one where the payoff from R_k equals 0 upon failure and $1 + B$ upon success. It thus suffices to study the case where all $w_k^i \in [0, B]$ and no participant quits in 3b, and then to show that getting B upon failure (and $1 + B$ upon success) dominates getting 0 upon failure (and $1 + B$ upon success). At this point, I therefore assume that $w_k^i \in [0, B]$ and no participant quits in 3b.

Second, consider a participant's decision to quit in stage 2c (at which point, again, all private benefits are already sunk). Given the assumption that 'when indifferent the participant stays', participant P_i will quit project R_k iff $\underline{u}_k^i > w_k^i + \alpha_k^i[(1-\theta)E^i(d_k) + \theta E^i(d_{-k})]$. This immediately implies that 1) no participant will quit if the outside values have dropped to zero, and 2) only the participant who owns a_k will ever quit R_k. Furthermore, P_i's decision to quit depends on the game history only through the $\underline{u}_k^i$ and whether $D_m = Z_m^i$ or not, and P_i is more likely to quit when $D_m \neq Z_m^i$ than when $D_m = Z_m^i$. It also follows, for further reference, that a change in one of the actions cannot make P_j simultaneously strictly more likely to quit R_k and strictly less likely to quit R_{-k}.

I now turn to showing that the condition $b < (1-p)(1-2\theta)(2\underline{\nu}-1) - p\theta\underline{\nu}$ is indeed sufficient to ensure that the optimal action is not changed by the presence of the private benefits. In fact, I will show that, in terms of the α_k^i, the decisions D_k, and the in-equilibrium quitting decisions, the subgame equilibrium outcomes identified in Proposition 1 remain optimal (in the same parameter ranges) and are thus again the equilibrium outcomes as long as they are feasible. [I will later show that they are indeed feasible and that the other elements of the proof of Proposition 1 also extend.]

Let, without loss of generality, the players and projects be renamed so that $\nu_1^1 + \nu_2^1 \geq \nu_1^2 + \nu_2^2$ and $\nu_1^1 \geq \nu_2^1$ (which implies $\nu_1^1 \geq \nu_1^2$). Let also $\hat{\theta} = \frac{\overline{\nu} - \underline{\nu}}{2\overline{\nu} - 1}$ when $\boldsymbol{\nu} = (\overline{\nu}, \underline{\nu}, \underline{\nu}, \overline{\nu})$, and $\hat{\theta} = 0$ otherwise. Let $V_k \in \{X_k, Y_k\}$ denote the action that gives the player who executes project R_k a private benefit b, let $S = (Z_1^1, Z_2^1, Z_1^2, Z_2^2, V_1, V_2)$ denote the state in terms of the players' beliefs and private preferences, let $I_{q,k}(S)$ be the indicator function whether one or both players try to quit project k in stage 2c, let $\beta_k^i(S) \in \{0, 1\}$ denote the indicator function that the decision D_k follows P_i's belief ($D_k = Z_k^i$), and $\gamma_k(S)$ the indicator function that D_k follows the private preference of $P(k)$ ($D_k = V_k$). The players'

*Van den Steen: Harvard Business School, 15 Harvard Way, Boston MA 02163, evandensteen@hbs.edu.

joint utility then equals, after some algebra,

$$U = \frac{1}{64}\sum_{S,k}\gamma_k(S)b + pI_{q,k}(S)\underline{u}$$
$$+(1-pI_{q,k}(S))\left[B+\sum_i \alpha_k^i\left(\frac{1}{2}+(1-\theta)(2\beta_k^i(S)-1)(\nu_k^i-\frac{1}{2})+\theta(2\beta_{-k}^i(S)-1)(\nu_{-k}^i-\frac{1}{2})\right)\right]$$

Note now the following for any optimal equilibrium, disregarding (for now) feasibility. First, since $b<(1-p)(2\underline{\nu}-1)$, whenever S is such that $Z_k^i = Z_k^{-i}$, then $D_k = Z_k^i$ in any optimal equilibrium (again, disregarding feasibility). Second, since U is linear in the α_k^i, it suffices to consider the case that $\alpha_k^i \in \{0,1\}$ and later confirm that these are strictly optimal for the derived equilibria.

Consider now first the case where one player, say P_i, gets all residual income: $\alpha_1^i = \alpha_2^i = 1$. In that case, since $b<(1-p)(2\underline{\nu}-1)$, U is maximized by also setting $\beta_k^i(S)=1, \forall k, \forall S$ (i.e., by following Z_k^i, $\forall k$) and then by setting $I_{q,k}(S)=0$ (i.e., by never quitting) given the assumption $B+\underline{\nu}>\underline{u}$. The joint utility then equals $U=2B+\nu_1^i+\nu_2^i+b$. When $P_i=P_1$, this is exactly the outcome of F_I in the proposition. With $P_i=P_2$, this is weakly dominated by the case with $P_i=P_1$ and strictly so unless $\nu_1^1=\nu_1^2$ and $\nu_2^1=\nu_2^2$.

I next want to argue that whenever $\nu_k^1 \geq \nu_k^2, \forall k$ then the F_I outcome (i.e., with $\alpha_k^1=\beta_k^1(S)=1\,\forall k$ and no quitting) is optimal, and strictly so when $\nu_k^1>\nu_k^2$ for some k. (Since players and projects were already named such that $\nu_1^1+\nu_2^1 \geq \nu_1^2+\nu_2^2$, the extra condition here is that P_1 has more confidence on *each* project individually.) The above arguments already implied that this F_I outcome (with $\alpha_k^1=\beta_k^1(S)=1\,\forall k$) dominates (at least weakly) the case with $\alpha_k^2=1\,\forall k$, and strictly so when $\nu_k^1>\nu_k^2$ for some k. Consider then the alternative, i.e., any potentially optimal equilibrium that sets $\alpha_m^1=\alpha_{-m}^2=1$ for some m. To see that F_I dominates, imagine that (with $\alpha_m^1=\alpha_{-m}^2=1$) we pick one project R_k and choose all decisions to maximize the expected utility of just that project. I will show that even if you could maximize like this on a project by project basis (i.e., as if D_k can be different for the two projects and without considering feasibility), F_I actually dominates. Let P_i be the player such that $\alpha_k^i=1$. The expected utility for R_k is then

$$U_k = \frac{1}{64}\sum_{S}\gamma_k(S)b + pI_{q,k}(S)\underline{u}$$
$$+(1-pI_{q,k}(S))\left[B+\left(\frac{1}{2}+(1-\theta)(2\beta_k^i(S)-1)(\nu_k^i-\frac{1}{2})+\theta(2\beta_{-k}^i(S)-1)(\nu_{-k}^i-\frac{1}{2})\right)\right]$$

Since D_{-k} only affects $\beta_{-k}^i(S)$, it is clearly optimal to set $D_{-k}=Z_{-k}^i$ and thus $\beta_{-k}^i(S)=1$ (if the sole objective is to maximize U_k). Second, since $b<(1-\theta)(1-p)(2\underline{\nu}-1)$, it is also optimal to always set $D_k=Z_k^i$ and thus $\beta_k^i(S)=1$. Next, since $B+\underline{\nu}>\underline{u}$ it is then also optimal to set $I_{q,k}(S)=0$. This gives $U_k=\frac{b}{2}+B+(1-\theta)\nu_k^i+\theta\nu_{-k}^i$ with overall utility of at most $b+2B+(1-\theta)\nu_k^i+\theta\nu_{-k}^i+(1-\theta)\nu_{-k}^{-i}+\theta\nu_k^{-i}$. But this is at least weakly dominated by F_I when $\nu_k^1\geq\nu_k^2, \forall k$ and strictly so when $P_i=P_2$ and $\nu_k^1>\nu_k^2$ for some k.

Given the renaming of players and projects such that $\nu_1^1+\nu_2^1\geq\nu_1^2+\nu_2^2$ and $\nu_1^1\geq\nu_2^1$, the only alternative left is $(\overline{\nu},\underline{\nu},\underline{\nu},\overline{\nu})$, which I now consider. The earlier argument that F_I is optimal whenever $\alpha_k^i = 1, \forall k$ extends, so that we only have to consider potential equilibrium outcomes with $\alpha_m^i = \alpha_{-m}^{-i} = 1$.

Consider first the case that $\alpha_2^1 = \alpha_1^2 = 1$ (i.e., the residual income of a project goes to the player with least confidence on the project's decision). To see that F_I strictly dominates any such equilibrium, I will again show that even if you could maximize on a project by project basis (i.e., as if each D_k can be different by project and without considering feasibility), F_I actually dominates. Let P_i be the player such that $\alpha_k^i = 1$. The expected utility for that project is then

$$U_k = \frac{1}{64}\sum_S \gamma_k(S)b + pI_{q,k}(S)\underline{u}$$
$$+(1-pI_{q,k}(S))\left[B+\left(\frac{1}{2}+(1-\theta)(2\beta_k^i(S)-1)(\underline{\nu}-\frac{1}{2})+\theta(2\beta_{-k}^i(S)-1)(\overline{\nu}-\frac{1}{2})\right)\right]$$

This is again maximized by setting $\beta_{-k}^i(S) = \beta_k^i(S) = 1$ and then by setting $I_{q,k}(S) = 0$ which then gives expected utility $U_k = B + (1-\theta)\underline{\nu} + \theta\overline{\nu} + \frac{b}{2}$ for each project. It follows that $U = 2B + 2(1-\theta)\underline{\nu} + 2\theta\overline{\nu} + b$ which is strictly less than the $2B + \overline{\nu} + \underline{\nu} + b$ from the F_I equilibrium. It thus follows that F_I always strictly dominates.

Consider finally the case where $\alpha_1^1 = \alpha_2^2 = 1$ (i.e., the residual income of a project goes to the player with most confidence on the project's decision). In that case, the part of joint utility that depends *directly* on D_k (with $\alpha_k^i = 1$) :

$$\gamma_k(S)b + (1-pI_{q,k})\left[(1-\theta)(2\beta_k^i-1)(\overline{\nu}-\frac{1}{2})\right] + (1-pI_{q,-k})\left[\theta(2\beta_k^{-i}-1)(\underline{\nu}-\frac{1}{2})\right]$$

The assumption that $b < (1-p)(1-2\theta)(2\underline{\nu}-1) - p\theta\underline{\nu}$ implies that $(1-p)(1-\theta)\overline{\nu} + (1-p)\theta(1-\underline{\nu}) > b + (1-p)(1-\theta)(1-\overline{\nu}) + \theta\underline{\nu}$. It follows that it is always optimal to set $D_k = Z_k^i$ when $i = k$. When $\theta \le \hat{\theta}$ then $B + (1-\theta)\overline{\nu} + \theta(1-\underline{\nu}) > \underline{u}$ so that it is always optimal to set $I_{q,k} = 0$. This gives the F_{NI} equilibrium, which dominates F_I when $\theta \le \hat{\theta}$. When, on the other hand, $\theta > \hat{\theta}$ the F_I equilibrium always dominates this one.

Overall, it follows that only the F_I and F_{NI} equilibrium outcomes can be optimal (when disregarding feasibility) and that F_I dominates F_{NI} when $\theta > \hat{\theta}$. In other words, in terms of the α_k^i, the decisions, and the in-equilibrium quitting decisions, the subgame equilibrium outcomes of Proposition 1 remain the optimal ones (in the same parameter range), when feasible. I will now show that they are indeed feasible under the condition $b < p\min(B + \underline{\nu} - \underline{u}, B)$.

Since the private benefits are immediately sunk upon choosing an action, they only affect the decision in step 2b. Consider the employee's incentive compatibility constraint to always obey the manager in the F_I equilibrium. Assuming, for example, that – when P_2 executes R_1 as in the proof of Proposition 1 – P_1 commits to quitting R_1 upon P_2's disobedience (for which the conditions remain unchanged since b is sunk by the time of the quitting decision), that incentive compatibility constraint becomes

$$w_1^2 > p\underline{u}_1^2 + (1-p)w_1^2 + b$$

or (since P_2 owns no assets)

$$w_1^2 > \frac{b}{p}$$

Given this, I will now again first argue that there always exists a subgame equilibrium with $(D_1,D_2)=(Z_1^1,Z_2^1)$, $\alpha_1^1=\alpha_2^1=1$, and no player quits on the equilibrium path, and that moreover the elements of F_I in the proposition are necessary to optimally implement this equilibrium. In particular, this subgame equilibrium can always be implemented as follows: P_1 owns both assets, P_1 executes R_2, P_2 executes R_1, the contract sets $w_1^1 \in [0,B-b/p)\cap[\underline{u}-(1-\theta)\nu_1^1-\theta\nu_2^1,\underline{u}-(1-\theta)(1-\nu_1^1)-\theta\nu_2^1)$ and $w_2^1\in[0,B]\cap[\underline{u}-(1-\theta)\nu_2^1-\theta\nu_1^1,B]$, which are both always non-empty (given that $(1-\theta)(1-\overline{\nu})+\theta\overline{\nu}\le(1-\theta)(1-\underline{\nu})+\theta\underline{\nu}$ and that $\underline{u}-(1-\theta)\nu_1^1-\theta\nu_2^1<B-b/p$ since $B+\underline{\nu}-\underline{u}>b/p$). In this case, P_2 will never quit (since he owns no assets and $w_1^2,w_2^2\ge 0$) while P_1 will not quit when the actions are (Z_1^1,Z_2^1) (since $w_1^1+(1-\theta)\nu_1^1+\theta\nu_2^1\ge\underline{u}$ and $w_2^1+(1-\theta)\nu_2^1+\theta\nu_1^1\ge\underline{u}$). This also implies that it is, in equilibrium, optimal for P_1 (who executes R_2) to choose $D_2=Z_2^1$: doing so gives him at least $w_1^1+(1-\theta)\nu_1^1+\theta\nu_2^1+w_2^1+(1-\theta)\nu_2^1+\theta\nu_1^1$ which is larger than $b+pI_{q,1}\underline{u}+(1-pI_{q,1})(w_1^1+(1-\theta)\nu_1^1+\theta(1-\nu_2^1))+pI_{q,2}\underline{u}+(1-pI_{q,2})(w_2^1+(1-\theta)(1-\nu_2^1)+\theta\nu_1^1)$. On the other hand, player P_1 *will* quit at least R_1 when the actions are $(\overline{Z}_1^1,Z_2^1)$ (since $w_1^1+(1-\theta)(1-\nu_1^1)+\theta\nu_2^1<\underline{u}$ and b is sunk by the time he decides on quitting). Since P_1 will thus quit iff P_2 disobeys, P_2 prefers the (at least) $w_1^2>0$ from obeying over the (at most) $b+(1-p)w_1^2<w_1^2$ from disobeying, and thus obeys. And given all that, the equilibrium in the message game is indeed that P_1 (and only P_1) tells P_2 what to do. The proof of the necessity of the elements in Proposition 1 remains unchanged. The same is true for the parts of the proof of Proposition 1 that relate to the F_{NI} partition.

It is now, finally, again straightforward to see that any equilibrium with $w_k^i\notin[0,B]$ is indeed strictly dominated. This completes the proof of the Proposition.

Proof of Proposition 3: I will start again with a partial backwards induction. First, a player will exert effort in stage 3c if and only if the project was not yet terminated, the decision was correct, and $\alpha_k^i\theta_e\ge\theta_e c_e$ or $\alpha_k^i\ge c_e$. Let in what follows $I_{e,k}$ denote whether $P(k)$ will exert effort on R_k (which is thus uniquely determined by α_k^i). Consider next stage 3b. Upon a success, no participant quits in (any) equilibrium (given $\alpha_k^i\ge 0$ and Nash bargaining). Upon a failure, a participant P_i quits R_k iff $w_k^i<0$. It follows that (since $w_k^1+w_k^2=B$) if some $w_k^i\notin[0,B]$, then R_k is terminated upon failure, so that the w_k^i are then paid only upon success. This case is equivalent to one where the payoff from R_k equals 0 upon failure and $1+B$ upon success. It thus suffices to study the case where all $w_k^i\in[0,B]$ and no participant quits in 3b, and then to show that getting B upon failure (and $1+B$ upon success) dominates getting 0 upon failure (and $1+B$ upon success). At this point, I therefore assume $w_k^i\in[0,B]$ and no participant quits in 3b. Third, consider a participant's decision to quit in stage 2c. Given the assumption that 'when indifferent the participant stays', a participant $P_i=P(k)$ will quit project R_k iff $\underline{u}_k^i>w_k^i+(\alpha_k^i-C_eI_{e,k})((1-\theta)E^i(d_k)+\theta E^i(d_{-k}))(1-\theta_e+\theta_eI_{e,k})$ and a participant $P_i\ne P(k)$ will quit R_k iff $\underline{u}_k^i>w_k^i+\alpha_k^i((1-\theta)E^i(d_k)+\theta E^i(d_{-k}))(1-\theta_e+\theta_eI_{e,k})$. This immediately implies two things: 1) no participant will quit if the outside values have dropped to zero, and 2) only the participant who owns a_k will ever quit R_k. Furthermore, P_i's decision to quit depends on the game history only through the $\underline{u}_k^i$, D_k and D_{-k}, and P_i is more likely to quit when $D_m\ne Z^i$ than when $D_m=Z^i$. It also follows, for further reference, that a change in one of the actions cannot make P_j simultaneously strictly *more* likely to quit R_m and strictly *less* likely to quit R_{-m}.

Fourth, before continuing the backwards induction, I want to argue that the message subgame is such that at the start of stage 2b of any (pure strategy) equilibrium, player P_i either always knows P_{-i}'s belief or never knows it (i.e., it is never the case that he knows it in some states but not in others). The reason is that in a pure-strategy equilibrium where a player's only private information at the time of sending a message is his own belief (and he can have only two possible beliefs), there are only two possibilities: either the player acts the same independent of his belief (and then the other player never knows his belief) or he takes a different action dependent on his belief (and then the other player always knows his belief). (Note though that P_i's belief about P_{-i}'s *action* may depend on seemingly irrelevant events, i.e., they could use such events to coordinate in the presence of multiple equilibria.)

Consider next the decision in period 2b and let $I_{q,k}$ denote whether one or both players try to quit R_k in period 2c (which may depend on the decisions D_1 and D_2). Player P_i's expected utility, assuming that P_i executes R_k, can be written

$$\begin{aligned} &pE^i[I_{q,k}]\underline{u}^i_k + (1-pE^i[I_{q,k}])[w^i_k + (\alpha^i_k - C_e I_{e,k})((1-\theta)E^i(d_k) + \theta E^i(d_{-k}))(1-\theta_e+\theta_e I_{e,k})] \\ &\quad + pE^i[I_{q,-k}]\underline{u}^i_{-k} + (1-pE^i[I_{q,-k}])[w^i_{-k} + \alpha^i_{-k}((1-\theta)E^i(d_{-k}) + \theta E^i(d_k))(1-\theta_e+\theta_e I_{e,-k})] \end{aligned}$$

Since this is a simultaneous-move game, P_{-i}'s decision on D_{-k} is fixed from P_i's perspective, while $I_{e,k}$ depends uniquely on the α^i_k. It follows that only $E^i[I_{q,m}]$ and $E^i(d_k)$ can depend on P_i's decision for D_k.

Consider now first the case where, given his information at the start of 2b, P_i believes that the likelihood that P_{-i} quits is the same (for both projects) or smaller (for one or both projects) when he chooses Z^i than when he chooses the other action, which I will denote as $\overline{Z}^i$. If he believes that it is the same, then P_i should only consider $E^i(d_k)$ when choosing D_k, so that it is optimal for P_i to choose Z^i. The same holds true when it is smaller (on one or both projects) since P_i can always quit himself where P_{-i} used to quit, which brings this back to the earlier case. In these cases, it is thus optimal for P_i to choose $D_k = Z^i$. Note that this includes any case where P_i does not know P_{-i}'s belief Z^{-i} or where P_i knows that $Z^i = Z^{-i}$. Note also that this implies that to get P_i to choose anything other than Z^i, it must be that P_{-i} is strictly *more* likely to quit one or both projects when P_i chooses Z^i than when he chooses $\overline{Z}^i$.

This concludes the partial backwards induction. I now turn to determining the equilibria.

By exactly the same arguments as in the proof of Proposition 1, the F_{NI} and $F_{I,pe}$ equilibria are always feasible (under the asset allocations, residual income allocations, and wage allocations there specified, after taking into account the cost of effort) and their outcomes require the contract values and actions specified in the Proposition, as well as the specified asset ownership when $\underline{u} > B$. (The $\underline{u} > B$ condition is the reason why the specified asset ownership structures are 'the only ownership structure that is part of an equilibrium for all parameter values,' rather than necessary for all parameter values as in Proposition 1.)

Furthermore, the contracts (and implied subgame equilibria) of F_{NI} and $F_{I,pe}$ jointly strictly dominate any other contracts (and implied subgame equilibria) where either no player exerts effort or one player exerts effort. To see this, consider what contracts (and equilibria) would maximize joint utility if effort (and correspondingly whether one or both players incur C_e) was exogenously given. The arguments of Proposition 1 imply that the maximizing contracts and equilibria would be exactly these of F_{NI} and $F_{I,pe}$. Imposing extra restrictions to force one or both players *not* to exert any effort can only reduce joint utility (since effort is efficient), which proves this step.

Let in what follows the projects be renamed so that P_1 executes R_1 and P_2 executes R_2. I can thus limit attention now to the case where both players exert effort (i.e., $\alpha_{i,i} \geq c_e$ for both). I will first argue that it suffices to look at equilibria where each player P_i either always chooses Z^i or always chooses Z^{-i}. Note first that the arguments above imply that when $Z^i = Z^{-i}$, then $D_k = Z^i$ in any subgame perfect equilibrium. So I only have to consider the case that $Z^i \neq Z^{-i}$. The result is then trivial when neither player knows the other's belief: each player P_i will always choose Z^i. Consider next the case where one player, say P_i, (always) knows P_{-i}'s belief but P_{-i} doesn't (ever) know P_i's belief. In that case, P_{-i} will always choose Z^{-i}. (Moreover, for any set of parameters, either P_{-i} will always quit if P_i chooses $\overline{Z}^{-i}$ or P_{-i} will never quit.) Furthermore, given the indifference assumptions, P_i will only choose $\overline{Z}^i = Z^{-i}$ if it gives him strictly higher expected continuation utility than choosing Z^i. But if so, then that must be the case in all $Z^i \neq Z^{-i}$ states and thus P_i will always choose Z^{-i}. Consider, finally, the case where each player always knows the other's belief. This gives four independent complete-information subgames. Any action combination can be described as one of the following four: (Z^1,Z^1), (Z^2,Z^2), (Z^1,Z^2), or (Z^2,Z^1). I will now argue that there can be no equilibrium in which two states (in terms of beliefs and with $Z^i \neq Z^{-i}$) have two different outcomes from among these four. For example, there can be no equilibrium such that the outcome is (Z^1,Z^2) when $Z^1 \neq Z^2 = X$ and (Z^2,Z^1) when $Z^1 \neq Z^2 = Y$. To see this, note that if any of these four is the outcome in one state, then it must be feasible in all states. So, given the assumption of Pareto optimality and the fact that these are four independent subgames, any particular equilibrium can only have different outcomes in different states if they give the same joint utility. But the indifference assumptions imply that even when financially indifferent, (Z^1,Z^2) is preferred over both (Z^1,Z^1) and (Z^2,Z^2), which are preferred over (Z^2,Z^1). Moreover, (Z^1,Z^1) and (Z^2,Z^2) can only both appear in the same equilibrium if both players communicate their beliefs while 'always (Z^1,Z^1)' (and analogously 'always (Z^2,Z^2)') require only one player to communicate. This implies that any equilibrium must have the same one of these four outcomes for all states, so that in any equilibrium each player P_i either always chooses Z^i or always chooses Z^{-i}.

Consider now first any (Z^1,Z^2) equilibrium. Disregarding feasibility for a moment (and taking into account that the players always have the same confidence ν), total utility is maximized by setting $\alpha_1^1 = \alpha_2^2 = 1$ and then by never quitting on the equilibrium path. This is the F_{NI} equilibrium, and that is indeed feasible by the earlier argument.

Consider next any (Z^2,Z^1) equilibrium. Disregarding again feasibility for a moment, total utility is now maximized by shifting income from R_1 as much as possible to P_2 and income from R_2 to P_1. This is constrained, however, by $\alpha_1^1, \alpha_2^2 \geq c_e$. But that means that joint utility is lower than in the (Z^1,Z^2) case, which thus dominates (and which was always feasible).

I can thus limit attention now to the case where both players exert effort (i.e., $\alpha_{i,i} \geq c_e$ for both) and, for some $i \in \{1,2\}$, the decisions always follow P_i's beliefs: $(D_1,D_2) = (Z^i,Z^i)$. Since players and projects are symmetric, it suffices to consider the case with $\boldsymbol{D} = (Z^1,Z^1)$ and P_1 executes R_1.

Consider first the case that P_1 commits to quitting both projects iff P_2 disobeys (which requires that P_1 never quits on the equilibrium path). Note that this requires that P_1 owns both assets since he otherwise cannot commit to quitting, which further implies that P_2 will never quit in 2c since his outside option is zero. The subgame perfection constraints are then that $\alpha_{i,i} \geq c_e \, \forall i$,

$$w_1^1 \in [0,B] \cap [\underline{u} - \nu[\alpha_1^1 - \theta_e c_e], \underline{u} - [(1-\theta)\nu + \theta(1-\nu)][\alpha_1^1 - \theta_e c_e]), \tag{1}$$

$$w_2^1 \in [0,B] \cap [\underline{u} - \alpha_2^1 \nu, \underline{u} - \alpha_2^1[(1-\theta)(1-\nu) + \theta\nu]) \tag{2}$$

(which imply that all $w_k^2 \in [0,B]$), and

$$w_1^2 + w_2^2 > \alpha_1^2 \left[\frac{(1-p)}{p} \theta(2\nu - 1) - (1-\nu) \right] + (\alpha_2^2 - \theta_e c_e) \left[\frac{(1-p)}{p} (1-\theta)(2\nu - 1) - (1-\nu) \right] \tag{3}$$

while joint utility equals

$$U = B + (\alpha_1^1 - \theta_e c_e)\nu + \alpha_1^2 \frac{1}{2} + B + \alpha_2^1 \nu + (\alpha_2^2 - \theta_e c_e) \frac{1}{2}$$

Joint utility is always maximized by shifting residual income to P_1. I will now argue that the subgame perfection constraints are also maximally relaxed by shifting residual income to P_1.

Since the w_k^i do not figure in U, their choice simply has to satisfy the subgame perfection conditions (if possible). It is easy to verify that such selection exists if and only if the w_k^2 that correspond to the lower bounds on w_k^1 (and thus the upper bound on w_k^2) as defined by equations (1) and (2) satisfy equation (3) strictly. Since the assumption that $\underline{u} > (1 - \theta_e c_e)\nu$ (combined, for w_2^1, with $\alpha_2^2 \geq c_e$ and thus $\alpha_2^1 \leq 1 - c_e$) implies that these lower bounds are $w_1^1 = \underline{u} - \nu[\alpha_1^1 - \theta_e c_e]$ and $w_2^1 = \underline{u} - \alpha_2^1 \nu$, such selection exists iff

$$\begin{aligned} &B - \underline{u} + \nu[(1-\alpha_1^2) - \theta_e c_e] + B - \underline{u} + (1-\alpha_2^2)\nu \\ &\quad > \alpha_1^2 \left[\frac{(1-p)}{p} \theta(2\nu - 1) - (1-\nu) \right] + (\alpha_2^2 - \theta_e c_e) \left[\frac{(1-p)}{p} (1-\theta)(2\nu - 1) - (1-\nu) \right] \end{aligned}$$

This condition is indeed most relaxed by minimizing α_2^2 and α_1^2.

A straightforward but tedious analysis shows that the conditions are even harder to satisfy when P_1 commits to quitting only one of the projects iff P_2 disobeys. It thus follows that the only potentially optimal equilibrium with both players exerting effort and $\boldsymbol{D} = (Z^i, Z^i)$ has P_i executing R_k, $\alpha_k^i = 1$, $\alpha_{-k}^i = 1 - c_e$. It is again straightforward that these equilibrium outcomes require the contract values and actions specified in the proposition, as well as the asset ownership when $\underline{u} > B + c_e(1 - \theta_e)(1-\nu)$. This concludes the proposition.

Proof of Proposition 4: The proof of Proposition 3 goes through with very few changes. In particular, making the actions contractible removes any constraint on the outcomes that follow from either the need to commit a player to quit upon disobedience by the other or from the need to make a player obey. This does not affect the backwards induction arguments or anything related to the F_{NI} and $F_{I,pe}$ equilibria in the proof of Proposition 3. (While some arguments in the proof have become superfluous, removing them does not affect the ultimate conclusions. Examples are some steps in the feasibility proof and some arguments in the asset ownership proof.) On the other hand, it does relax the feasibility conditions of the $F_{I,fe}$ equilibrium. In particular, conditions (1) and (2) become

$$w_1^1 \in [0,B] \cap [\underline{u} - \nu[\alpha_1^1 - \theta_e c_e], B], \tag{4}$$

$$w_2^1 \in [0,B] \cap [\underline{u} - \alpha_2^1 \nu, B] \tag{5}$$

while condition (3) is not necessary any more. This is always feasible by setting $w_1^1 = w_2^1 = B$. It follows that $F_{I,fe}$ becomes feasible over a larger parameter range (although its expected joint utility does not change). The rest of the proof is again not affected (at least not in terms of conclusions). This concludes the proof.

II. Relationship to Other Theories of Firm Boundaries

As mentioned in the paper I will discuss here in more detail how the theory in this paper complements the theories highlighted by Gibbons (2005). The key point of the discussion is to show that this theory is highly complementary to theories based on rent-seeking, adaptation, or incentives, while it is compatible – in a more orthogonal sense – with the property rights theory.

Consider first the property rights theory of Grossman and Hart (1986), Hart and Moore (1990), and Hart (1995). For the case of an entrepreneur-owner, the theory in this paper integrates quite easily – though without much interaction – with the property rights theory. In particular, if – in the model of this paper – the player can make investments prior to the *contracting* stage, then asset ownership will give investment incentives very similar to these in the GHM property rights model. Things get more complex when considering a larger firm with multiple shareholders, however. In particular – as pointed out first by Holmstrom and Roberts (1998) and later by Hart and Holmstrom (2002) – the property rights theory is essentially about individuals – rather than firms – owning assets. This is an issue for further research.

Consider next the rent-seeking and hold-up models in the style of Klein, Crawford and Alchian (1978) or Williamson (1985). Combining the main model with a 'costly rent-seeking' model in the style of Masten (1986) suggests that such rent-seeking models integrate well with the current theory. More importantly, it seems that the two theories enrich each other. In one direction, the current theory may provide formal answers to Hart's (1995) criticism of the rent-seeking models. In particular, it suggests a formal answer how firm boundaries affect authority and rent-seeking and it can simultaneously endogenize the cost of integration in the form of a suboptimal allocation of individual projects (from a standalone perspective). In the other direction, the rent-seeking models provide complementary predictions on firm boundaries for this 'authority theory of the firm'.

A somewhat similar relationship exists with the adaptation theory (Simon 1951, Williamson 1975). Consider the main model, but the exact magnitude of the externality will become clear only later in the game. Prior to the game, the parties can write a contract on that decision, but contractibility vanishes once the game starts (due, for example, to time pressure). The decision on firm boundaries will now depend on the need for adaptation. Also in this case, the current paper suggests formal micro-foundations for the adaptation story, while the adaptation theory provides complementary predictions. The same is true for theories, such as Hart and Holmstrom (2002), that assume that ownership of physical assets somehow conveys control over the projects that use these assets. While the Hart-Holmstrom assumption is uncontroversial for single-person projects, it is more problematic for large projects that require the non-trivial participation of a group of people. In that case, control over the project requires interpersonal authority over people. The current paper provides micro-foundations for why ownership of assets would indeed convey interpersonal authority over people, and thus allows to apply these theories to larger projects.

The relationship to the incentive theory of firm boundaries (Holmstrom and Milgrom 1994, Holmstrom 1999) is in some sense even tighter. In particular, imagine that some decisions affect, among other things, the value of an asset. Asset ownership by an employee is then similar to high-powered incentives and will thus lead to disobedience. This shows that the essential idea of the asset-incentives theory translate nearly literally to the current context.

Overall, apart from delivering a self-contained theory of the firm that is fully consistent with a differing priors interpretation of Knight's view, the theory in this paper is thus also a strong complement to some of the major perspectives on firm boundaries outlined in Gibbons (2005).

Another point that deserved more discussion was the relationship with Holmstrom and Milgrom (1994) who simultaneously explain – by combining monotone comparative statics (Milgrom and Roberts 1990) with multi-tasking (Holmstrom and Milgrom 1991) – the following triple: the firm can exclude employees from certain returns (such as the ability to take outside jobs), employees do not own assets, and employees have low-powered incentives. While excluding employees from certain returns can be done contractually, it can also be done by using authority. When taking that perspective, their paper also deals with the triple asset ownership, low-powered incentives, and authority. One key difference is what is meant with authority: their paper deals with the *use* of authority to forbid employees to receive income from certain activities, while this paper deals with the *origin* of interpersonal authority that is used to *directly* tell employees what to do and what not to do in a fairly general sense. Another important difference is in the role of assets. In Holmstrom and Milgrom (1994), it does not matter who owns the assets, as long as they are not owned by the employee, so that shifting assets from one firm to another does not matter, in contrast to the current paper. This latter distinction is important when it comes to discussing firm boundaries. A key insight of Holmstrom and Milgrom (1994) that has influenced this paper considerably is the observation that low-powered incentives in firms may not simply be an unfortunate consequence but the explicit purpose of transacting through a firm. Their paper is also the first to think about the firm in terms of a set of complementary practices.[1]

[1] Williamson (1991) also argues that firm and market governance each correspond to a set of internally consistent practices in the sense of Milgrom and Roberts (1990) or Holmstrom and Milgrom (1994).

References

Gibbons, Robert, "Four Formal(izable) Theories of the Firm?," *Journal of Economic Behavior and Organization*, 2005, *58*, 200– 245.

Grossman, Sanford J. and Oliver D. Hart, "The Costs and Benefits of Ownership : A Theory of Vertical and Lateral integration," *Journal of Political Economy*, August 1986, *94* (4), 691–719.

Hart, Oliver, *Firms, Contracts, and Financial Structure*, Oxford: Clarendon Press, 1995.

____ **and Bengt Holmstrom**, "A Theory of Firm Scope," 2002. MIT Department of Economics Working Paper No. 02-42.

____ **and John Moore**, "Property Rights and the Nature of the Firm," *Journal of Political Economy*, December 1990, *98* (6), 1119– 1158.

Holmstrom, Bengt, "The Firm as a Subeconomy," *Journal of Law, Economics, and Organization*, 1999, *15* (1), 74– 102.

____ **and John Roberts**, "The Boundaries of the Firm Revisited," *Journal of Economic Perspectives*, 1998, *12* (4), 73– 94.

____ **and Paul Milgrom**, "Multi-Task Principal Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics, and Organization*, 1991, *7*, 24– 52.

____ **and** ____ , "The Firm as an Incentive System," *American Economic Review*, September 1994, *84* (4), 972– 991.

Klein, Benjamin, Robert Crawford, and Armen Alchian, "Vertical Integration, Appropriable Rents, and the Competitive Contracting Process," *The Journal of Law and Economics*, 1978, *21*, 297– 326.

Masten, Scott E., "Institutional choice and the organization of production: the make-or-buy decision," *Journal of Institutional and Theoretical Economics*, 1986, *142*, 493– 509.

Milgrom, Paul and John Roberts, "The economics of modern manufacturing : Technology, strategy, and organization," *American Economic Review*, June 1990, *80* (3), 511– 528.

Simon, Herbert, "A Formal Theory of the Employment Relationship," *Econometrica*, 1951, *19*, 293– 305.

Williamson, Oliver, *Markets and Hierarchies : Analysis and Antitrust Implications*, New York: Free Press, 1975.

____ , *The Economic Institutions of Capitalism*, New York: Free Press, 1985.

Williamson, Oliver E., "Comparative Economic Organization: The Analysis of Discrete Structural Alternatives," *Administrative Science Quarterly*, June 1991, *36*, 269– 296.