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 \( w^i_k < 0 \) (given that any private benefits are sunk). It follows that (since \( w^i_k = B \) if some \( w^i_k \not\in [0,B] \), then \( R_k \) is terminated upon failure, so that the \( w^i_k \) 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 \( w^i_k \in [0,B] \) 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 \( w^i_k \in [0,B] \) 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 \( u^i_k > w^i_k + \alpha_k^i[(1 - \theta)E^i(d_k) + \theta E^i(d_{k-1})] \). 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 \( u^i_k \) and whether \( D_m = Z^i_m \) or not, and \( P_i \) is more likely to quit when \( D_m \neq Z^i_m \) than when \( D_m = Z^i_m \). 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 \( R_{k-1} \).

I now turn to showing that the condition \( b < (1 - p)(1 - 2\theta)(2\nu - 1) - p\theta \nu \) 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 \( \alpha_k^i \), 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 \( v^1_i + v^2_i \geq v^1_i + v^2_i \) and \( v^1_i \geq v^2_i \) (which implies \( v^1_i \geq v^2_i \)). Let also \( \hat{\theta} = \frac{\nu - \nu}{2\nu - 1} \) when \( \nu = (\nu, v, v, v) \), 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 \( S = (Z^1_k, Z^2_k, Z^3_k, Z^4_k, V_1, V_2) \) denote the state in terms of the players’ beliefs and private preferences, let \( \gamma_k(S) \) 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 \( (D_k = Z^i_k) \), and \( \beta_k(S) \) the indicator function that \( D_k \) follows the private preference of \( P(k) \) \( (D_k = V_k) \). The players’
joint utility then equals, after some algebra,

\[
U = \frac{1}{64} \sum_{S, \nu} \gamma_k(S)b + pI_{q,k}(S)u + (1 - pI_{q,k}(S)) \left[ B + \sum_{i} \alpha^i_k \left( \frac{1}{2} + (1 - \theta)(2\beta^i_k(S) - 1)(\nu^i_k - \frac{1}{2}) \right) + \theta(2\beta^i_{-k}(S) - 1)(\nu^i_{-k} - \frac{1}{2}) \right]
\]

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

Consider now the case where one player, say \( P_i \), gets all residual income: \( \alpha^i_1 = \alpha^i_2 = 1 \). In that case, since \( b < (1 - p)(2\nu - 1) \), \( U \) is maximized by also setting \( \beta^i_k(S) = 1, \forall \nu, \forall S \) (i.e., by following \( Z^i_k \), \( \forall \nu \)) and then by setting \( I_{q,k}(S) = 0 \) (i.e., by never quitting) given the assumption \( B + \nu > u \). The joint utility then equals \( U = 2B + v^1 + v^2 + b \). When \( P_i = P_1 \), this is exactly the outcome of \( F_1 \) in the proposition. With \( P_i = P_2 \), this is weakly dominated by the case with \( P_i = P_1 \) and strictly so unless \( v^1 = v^2 \) and \( v^1 \geq v^2 \).

I next want to argue that whenever \( v^1_k \geq v^2_k, \forall \nu \) then the \( F_1 \) outcome (i.e., with \( \alpha^i_k = \beta^i_k(S) = 1, \forall \nu \) and no quitting) is optimal, and strictly so when \( v^1_k > v^2_k \) for some \( k \). (Since players and projects were already named such that \( v^1_k + v^2_k \geq v^2_k + v^2_k \), the extra condition here is that \( P_i \) has more confidence on each project individually.) The above arguments already implied that this \( F_i \) outcome (with \( \alpha^i_k = \beta^i_k(S) = 1, \forall \nu \)) dominates (at least weakly) the case with \( \alpha^i_2 = 1, \forall \nu \), and strictly so when \( v^1_k > v^2_k \) for some \( k \). Consider then the alternative, i.e., any potentially optimal equilibrium that sets \( \alpha^i_m = \alpha^i_{-m} = 1 \) for some \( m \). To see that \( F_1 \) dominates, imagine that (with \( \alpha^i_m = \alpha^i_{-m} = 1 \)) 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 \( \alpha^i_k = 1 \). The expected utility for \( R_k \) is then

\[
U_k = \frac{1}{64} \sum_S \gamma_k(S)b + pI_{q,k}(S)u + (1 - pI_{q,k}(S)) \left[ B + \left( \frac{1}{2} + (1 - \theta)(2\beta^i_k(S) - 1)(\nu^i_k - \frac{1}{2}) \right) + \theta(2\beta^i_{-k}(S) - 1)(\nu^i_{-k} - \frac{1}{2}) \right]
\]

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

Given the renaming of players and projects such that \( v^1_k + v^2_k \geq v^2_k + v^2_k \) and \( v^1_k \geq v^2_k \), the only alternative left is \( (\forall, \nu, \nu, \nu) \), which I now consider. The earlier argument that \( F_i \) is optimal when-
ever $\alpha_i^i = 1, \forall k$ extends, so that we only have to consider potential equilibrium outcomes with $\alpha_m^i = \alpha_m^{-i} = 1$.

Consider first the case that $\alpha_i^i = \alpha_i^2 = 1$ (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 $\alpha_i^i = 1$. The expected utility for that project is then

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

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 $U_k = B + (1 - \theta)\nu + \theta\nu + \frac{b}{2}$ for each project. It follows that $U = 2B + 2(1 - \theta)\nu + 2\theta\nu + b$ which is strictly less than the $2B + \nu + \nu + b$ from the $F_i$ equilibrium. It thus follows that $F_i$ always strictly dominates.

Consider finally the case where $\alpha_i^i = \alpha_i^2 = 1$ (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 $\alpha_k^i = 1$):

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

The assumption that $b < (1 - p)(1 - 2\theta)(2\nu - 1) - p\theta\nu$ implies that $(1 - p)(1 - \theta)\nu + (1 - p)\theta(1 - \nu) > b + (1 - p)(1 - \theta)(1 - \nu) + \theta\nu$. It follows that it is always optimal to set $D_k = Z_k^i$ when $i = k$.

When $\theta \leq \hat{\theta}$ then $B + (1 - \theta)\nu + \theta(1 - \nu) > u$ so that it is always optimal to set $I_{q,k} = 0$. This gives the $F_{NI}$ equilibrium, which dominates $F_i$ when $\theta \leq \hat{\theta}$. Then, 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 $\alpha_i^i$, 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 $b < p \min(B + \nu - u, B)$.

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

$$w_1^2 > pu_1^2 + (1 - p)w_1^2 + b$$

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 \((D_1, D_2) = (Z_1, Z_2)\), \(\alpha_1 = \alpha_2 = 1\), and no player quits on the equilibrium path, and that moreover the elements of \(F_1\) 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 \(w_1^1 \in [0, B - b/p] \cap [u - (1 - \theta)v_1^1 - \theta v_2^1, u - (1 - \theta)(1 - v_1^1) - \theta v_2^1]\) and \(w_2^1 \in [0, B] \cap [u - (1 - \theta)\nu_1 - \theta \nu_1, B]\), which are both always non-empty (given that \((1 - \theta)(1 - \nu) + \theta \nu \leq (1 - \theta)(1 - \nu) + \theta \nu\) and that \(u - (1 - \theta)v_1^1 - \theta v_2^1 < B - b/p\) since \(B + \nu - u > b/p\)). In this case, \(P_2\) will never quit (since he owns no assets and \(w_2^1, w_2^2 \geq 0\) while \(P_1\) will not quit when the actions are \((Z_1, Z_2)\) (since \(w_1^1 + (1 - \theta)v_1^1 + \theta v_2^1 \geq u\) and \(w_1^1 + (1 - \theta)v_1^1 + \theta v_2^1 \geq u\)). This also implies that it is, in equilibrium, optimal for \(P_1\) (who executes \(R_2\)) to choose \(D_2 = Z_2\); doing so gives him at least \(w_1^1 + (1 - \theta)v_1^1 + \theta v_2^1 + w_2^1 + (1 - \theta)v_2^1 + \theta v_1^1\) which is larger than \(b + q(l)u + (1 - p(l)(w_1^1 + (1 - \theta)v_1^1 + \theta(1 - v_1^1)) + p(l)(u + (1 - p(l)(w_1^1 + (1 - \theta)(1 - v_1^1) + \theta v_2^1))\). On the other hand, player \(P_1\) will quit at least \(K_1\) when the actions are \((Z_1, Z_2)\) (since \(w_1^1 + (1 - \theta)(1 - v_1^1) + \theta v_2^1 < u\) 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) \(w_2^1 > 0\) from obeying over the (at most) \(b + (1 - p(l))w_2^1 < w_2^2\) 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 \(w_1^1 \notin [0, B]\) 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 \(\alpha_k \leq \theta e_c\) or \(\alpha_k \geq c_e\). Let in what follows \(I_{e,k}\) denote whether \(P(k)\) will exert effort on \(R_k\) (which is thus uniquely determined by \(\alpha_k\)). Consider next stage 3b. Upon a success, no participant quits in (any) equilibrium (given \(\alpha_k \geq 0\) and Nash bargaining). Upon a failure, a participant \(P_i\) quits \(R_i\) iff \(w_i^1 < 0\). It follows that (since \(w_1^1 + w_2^1 = B\)) if some \(w_1^1 \notin [0, B]\), then \(R_k\) is terminated upon failure, so that the \(w_i^1\) 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 \(w_i^1 \in [0, B]\) 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 \(w_i^1 \in [0, B]\) 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 \(w_i^1 > (\alpha_k + c_e)(1 - \theta)E^{d}(d_k) + \theta E^{d}(d_k)\) and a participant \(P_i \neq P(k)\) will quit \(R_k\) iff \(w_i^1 > (\alpha_k + \alpha_k)(1 - \theta)E^{d}(d_k) + \theta E^{d}(d_k)\). 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 \(w_i^1\), \(D_k\) and \(D_{-k}\), and \(P_i\) is more likely to quit when \(D_m \neq Z^1\) than when \(D_m = Z^1\). 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 \(P_{-i}^i\)’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 \(P_i^i\)’s belief about \(P_{-i}^i\)’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

\[
pE^i[I_{q,k}]\mathcal{U}_k^i + (1 - pE^i[I_{q,k}])\left(w_i^k + (\alpha_k^i - C_e I_{e,k})((1 - \theta)E^i(d_k) + \theta E^i(d_{-k}))(1 - \theta_e + \theta_e I_{e,k})\right)
\]

\[
+ pE^i[I_{q,-k}]\mathcal{U}_{-k}^i + (1 - pE^i[I_{q,-k}])\left[w_i^i + \alpha_i^i((1 - \theta)E^i(d_{-k}) + \theta E^i(d_k))(1 - \theta_e + \theta_e I_{e,-k})\right]
\]

Since this is a simultaneous-move game, \(P_{-i}^i\)’s decision on \(D_{-k}\) is fixed from \(P_i^i\)’s perspective, while \(I_{e,k}\) depends uniquely on the \(\alpha_k^i\). It follows that only \(E^i[I_{q,m}]\) and \(E^i(d_k)\) can depend on \(P_i^i\)’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 \(P_{-i}^i\) quits is the same (for both projects) or smaller (for one or both projects) when he chooses \(Z_i^i\) 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 \(Z_i^i\). The same holds true when it is smaller (on one or both projects) since \(P_i\) can always quit himself where \(P_{-i}^i\) used to quit, which brings this back to the earlier case. In these cases, it is thus optimal for \(P_i\) to choose \(D_k = \overline{Z}_i\). Note that this includes any case where \(P_i\) does not know \(P_{-i}^i\)’s belief \(Z_{-i}^i\) or where \(P_i\) knows that \(Z_i^i = Z_{-i}^i\). Note also that this implies that to get \(P_i\) to choose anything other than \(Z_i^i\), it must be that \(P_{-i}^i\) is strictly more likely to quit one or both projects when \(P_i\) chooses \(Z_i^i\) than when he chooses \(\overline{Z}_i^i\).

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

By exactly the same arguments as in the proof of Proposition \([\square]\), 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 \(u > B\). (The \(u > B\) 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 \([\square]\).)

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 \([\square]\) 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., \( \alpha_{i,j} \geq c_e \) 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 \( Z^{-i} \) or \( P_{-i} \) will never quit.) Furthermore, given the indifference assumptions, \( P_i \) will only choose \( Z^i = Z^{-i} \) 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 \( Z^1 \)’ (and analogously ‘always \( Z^2 \)’) 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 \( \nu \)), total utility is maximized by setting \( \alpha^1 \nu = \alpha^2 \nu = 1 \) 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., \( \alpha_{i,j} \geq c_e \) 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 \( D = (Z^1,Z^1) \) 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 $\alpha_{i,t} \geq c_e \forall i$,

$$w^1_i \in [0, B] \cap [u - v(\alpha^1_i - \theta_e c_e), u - [(1 - \theta) v + \theta(1 - v)][\alpha^1_i - \theta_e c_e]], \quad (1)$$

$$w^2_i \in [0, B] \cap [u - \alpha^2_i v, u - \alpha^2_i [(1 - \theta)(1 - v) + \theta v)] \quad (2)$$

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

$$w^1_i + w^2_i > \alpha^1_i \left[\frac{(1-p)}{p} \theta(2v - 1) - (1 - v)\right] + (\alpha^2_i - \theta_e c_e) \left[\frac{(1-p)}{p} (1 - \theta)(2v - 1) - (1 - v)\right]$$

while joint utility equals

$$U = B + (\alpha^1_i - \theta_e c_e) v + \alpha^2_i \frac{1}{2} + B + \alpha^1_i v + (\alpha^2_i - \theta_e c_e) \frac{1}{2}$$

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 $w^i_k$ 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 $w^2_k$ that correspond to the lower bounds on $w^1_k$ (and thus the upper bound on $w^2_k$) as defined by equations (1) and (2) satisfy equation (3) strictly. Since the assumption that $u > (1 - \theta_e c_e) v$ (combined, for $w^2_i$, with $\alpha^2_i \geq c_e$ and thus $\alpha^2_i \leq 1 - c_e$) implies that these lower bounds are $w^1_i = u - v(\alpha^1_i - \theta_e c_e)$ and $w^2_i = u - \alpha^2_i v$, such selection exists iff

$$B - u + v[(1 - \alpha^1_i - \theta_e c_e) + B - u + (1 - \alpha^2_i)]$$

$$> \alpha^1_i \left[\frac{(1-p)}{p} \theta(2v - 1) - (1 - v)\right] + (\alpha^2_i - \theta_e c_e) \left[\frac{(1-p)}{p} (1 - \theta)(2v - 1) - (1 - v)\right]$$

This condition is indeed most relaxed by minimizing $\alpha^2_i$ and $\alpha^1_i$.

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 $D = (Z', Z')$ has $P_i$ executing $D_k$, $\alpha^1_k = 1$, $\alpha^1_{i-k} = 1 - c_e$. 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 $u > B + c_e(1 - \theta_e)(1 - v)$. This concludes the proposition.

**Proof of Proposition 4**: The proof of Proposition 4 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

$$w_1^1 \in [0, B] \cap [u - v(\alpha_1^1 - \theta_{e}c_e), B], \quad (4)$$

$$w_2^1 \in [0, B] \cap [u - \alpha_2^1v, B] \quad (5)$$

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]

---

References


