International Journal of Economic Behavior and Organization
2018; 6(1): 1-10
http://www.sciencepublishinggroup.com/j/ijebo
doi: 10.11648/j.ijebo.20180601.11
ISSN: 2328-7608 (Print); ISSN: 2328-7616 (Online)

Surplus, Contractibility and Theory of the Firm

Zhongwei Wu

Seix Investment Advisors LLC, Park Ridge, USA

Email address:
zhongwei.wu@sloan.mit.edu

To cite this article:
Zhongwei Wu. Surplus, Contractibility and Theory of the Firm. *International Journal of Economic Behavior and Organization*. Vol. 6, No. 1, 2018, pp. 1-10. doi: 10.11648/j.ijebo.20180601.11

Received: December 24, 2017; **Accepted**: January 10, 2018; **Published**: January 26, 2018

Abstract: A theory of the firm is advanced as a unified property rights for surplus created by cooperative activity. This differs from traditional transaction cost literature which views the firm as a governance structure. Individual purposive entities act in their own interests, within or without a firm. Only contractible inputs are coordinated for joint surplus maximization. Each party determines her non-contractible inputs for her own benefits, which explain many observed firm phenomena. An equilibrium models are developed for the firm and used to explain the questions of firm existence, boundary, and ownership.

Keywords: Firm, Surplus, Contractibility, Equilibrium

1. Introduction

Since Ronald Coase's seminal 1937 paper initiated the transaction cost paradigm in economics (TCE), there has been much development in the theory of the firm. Particularly notable are Oliver Williamson's framework of strategic behavior, as well as Oliver Hart et.al. emphasis on allocation of residual decision rights.

While the TCE paradigm has created useful perspectives of the firm, it has generally not focused on the most fundamental features associated with firms: (a) that it has its own legal standing and property rights, (b) that it differs from other types of organizations such as clubs, non-profits and governments in its profit motive, (c) that it has its own purpose which is distinct from its participants (owners, employees), and (d) the role of the entrepreneur. In particular, the TCE literature primarily focuses on command hierarchy being the reason for a firm. It tends to ignore these other common features.

Starting from the asset specificity observations by Alchian et.al., an alternative theory of the firm can be constructed based on *surplus*. This theory is based on property rights considerations of cooperative activities and will provide a unified theoretical framework for firms with a formal model that gives all the major results of the theory of the firm.

2. Surplus and Integration – A Property Right Theory of the Firm

The TCE literature explains the firm as a hierarchy where fiat by a superior can remove opportunism and bargaining by the various parties in a transaction. It also explains that without the command hierarchy, one party can hold up the other party and appropriate her quasi-rent from any asset investment specific to the relationship. However, this notion is flawed in that: (a) a command structure does not require a firm to exist, consultants can take orders without being a firm employee, and (b) it is simply unrealistic and asymmetric to assume strategic behavior exists outside the firm in a market, but not within a firm between different purposive individuals.

2.1. Asset Specificity and Surplus

Alchian (1972) highlighted that the need for integration could arise from quasi-rent a party generates from her specific asset. Two observations can be made on specific assets, which play a central role in TCE. First, asset specificity is the result of surplus in cooperative activity. Surplus can be defined as the combined income of the parties involved in the cooperative activity over the sum of each party's respective best alternatives income on her asset (or opportunity cost). This definition is exactly equal to the sum

of their quasi-rents. Without surplus, there is no quasi-rent, nor specificity to the asset. Second, the specificity must be *mutual* and both parties must derive quasi-rent for the specific asset for it to be internalized. If any party does not view the other's asset as special, then she can simply contract for the use of a substitute asset at the market price. The asset does not need to be internalized, and hence no firm is needed.

Why do parties engage in cooperative production in the first place? The answer must be they anticipate joint surplus.[1] In static markets, such surplus is exactly equal to the joint quasi-rent of their inputs. In markets with uncertainties, the *expectation* of surplus is necessary for the parties to proceed. If the parties only generate the surplus jointly but not separately, the property right and claim to the surplus is not pre-determined. This *indeterminacy* in property rights leads to an indeterminacy in the *division* of the surplus. To the extent there is no *a priori* attribution of surplus to each party's individual inputs, bargaining comes into the picture.[2] The price of an intermediate product is but a parameter used in bargaining. At the same time, any division must result in quasi-rent for both parties. Hence, asset specificity is a result of surplus, rather than an exogenous assumption, as in the TCE literature. Viewed under this light, the firm itself is in fact a *contractual* solution to the bargaining problem of dividing the surplus. It creates a new entity to own the surplus, and contracts for pre-determined fractional claims by each party in the unified ownership, including possibly in the form of incentive compensations. This reduces misalignment of incentives.

The TCE elements of bounded rationality, uncertainty, information asymmetry, and incomplete contracts are not necessary for this indeterminacy to occur. However, these factors are necessary in deducing that long term contracts alone are ineffective in removing strategic behavior. Therefore, these factors can be summarized into the concept of non-contractibility, and be used in the model developed below. In the real world, incompleteness is the rule rather than the exception. Non-contractibility is the simple idea that the transaction costs of monitoring or measuring certain inputs are so prohibitive that the parties may not attempt to contract for them at all. The firm as a contract reduces strategic behavior by creating a pre-ante specification for the division of surplus. Strategic behavior on non-contractible inputs is thus dis-incentivized by an appropriate surplus sharing scheme. Non-contractibility necessarily means there is no competitive market that can provide substitutes under a price mechanism. Partial contracting may be construed as a blend of contractible inputs and non-contractible inputs. Therefore, the model retains its generality.

2.2. Benefits of Integration

In the context of strategic behavior, the benefit of a firm is thus the integration of the property rights of the surplus. If investments decisions in specific assets can be made by the firm rather than each party individually, the hold-up problem is solved. The *mechanism* of costly ongoing or ex-post bargaining is also removed, so long as one legal entity integrates the ownership of all of the supplier's specific assets, the intermediate product, and the buyer's specific assets. If all specific assets are internalized, only non-integrable or non-specific asset remain external and decisions on them are not-integrated. That is how an intermediate product market transforms into a factor market instead. No hierarchy is necessary for transaction costs to reduce. The common ownership can be achieved ex-ante before any specific investments, or ex-post through a merger, and whether it's a lateral or vertical integration.

The firm thus allows the competing interests from different parties to be unified in common ownership. It also acquires a *distinct purpose* different from those of its owners, in that its surplus creation does not coincide with its owners' surplus completely.[3] The firm's own profit seeking motive naturally emerges from this analysis. This is fundamentally different from the capital structure literature, which assumes it as an *a priori* tenet and the starting point of analysis. It also differs from TCE which views the firm as an organization structure rather than a legal ownership construct. However, the firm does rely on its owners and agents to fulfill its purpose. It is through these agents that its goals are achieved. That means strategic behavior will persist post integration. Analyses of such post integration behavior can explain observed phenomena in actual firms.

Note the firm's surplus here is measured in monetary rather than consumption utility terms, because the firm's output is specifically created for market exchange, not for consumption by its owners. That distinguishes it from other types of organizations such as cooperatives or clubs. This corroborates the Fisher first separation theorem which distinguishes the firm's profit motive from its owners'

1 The business community understands the firm's surplus purpose very well. For example, it is common parlance to distinguish a firm's going-concern value from its liquidation value. There is generally an assumption that assets are worth more when combined by the firm than what they are worth separately. The market value of a successful publicly traded firm is usually much higher than its book value, too. However, book value represents historic rather than current values, hence liquidation value is a more pertinent comparison.

2 Alchian and Demesetz (1972) alludes to the attribution indeterminacy as the "metering problem." However, it still views the firm as a production function. This analysis is different in at least three key aspects: (1) it treats the firm's purpose as surplus creation rather than joint production, with technology innovation as a special case, (2) it explicates a general model of non-contractible inputs rather than a limited notion of shirking, and (3) it develops fully the consequences of joint surplus and non-contractibility with a model to answer the firm boundary questions. Incentive structures are emphasized. Because there is no *a priori* unique way of *attributing* that surplus to each individual input, strategic behavior cannot be avoided, but only minimized. This is done by integrating and removing bargaining on certain inputs, and optimizing incentive structures on the other non-integrable inputs. This analysis subsumes the production function/shirking analysis as a special case.

3 Spulber (2009) emphasizes the firm as a transaction institution whose objectives differ from those of its owners. Grossman and Hart (1986) assumes the firm as ownership of assets, for the purpose of granting one party a control mechanism to make the decision in situations not contracted for. Instead, the firm is derived here as an entity independent from owners and a legal construct as a consequence of surplus creation.

consumption preferences.[4]

This explanation of integration differs from Grossman and Hart (1986), and Hart and Moore (1990) (GHM). While the GHM theory is called a "property-rights" theory, it actually presupposes *disparate* profit centers by the two parties even after integration, and hence views property rights as *not* unified in a firm. That line of thought follows from TCE reasoning – if the benefit of integration is hierarchy/command, one needs to justify how a fiat mechanism *overcomes* each party's individual profits seeking. GHM accomplishes this by assuming an owner through fiat can affect the incentive structure of both parties for non-contractible ex-ante decisions. Thus by assuming command is the *only* mechanism that can change the outcome, the theory predicts who owns the firm is important.[5]

In reality, different divisions generally do not appropriate their own profits, and command can create strategic behavior of its own. More generic assumptions are needed. In the model below, the *unification* of surplus is the motivation for integration. The sharing in this unified property rights in surplus creates the *incentives* for the parties to improve their respective non-contractible behavior. Second, strategic behavior persists in a firm and each party continues to maximize their own quasi-rent by choosing the non-contractible portions of their decisions.

The GHM theory's key conclusion is that the party making the more important non-contractible decision should retain ownership of the firm, through her exercise of residual decision rights. However, the original GHM theory leaves no room for joint ownership, and it took much development in the literature, and many additional assumptions to derive a more realistic ownership result. Gattai and Natale (2015) provides a good survey of this very long journey. In the sections below, an analogous but stronger result will be derived. It will come from the fact that the right incentive structure is needed for a party to improve her non-contractible behavior.

2.3. Firm Boundary

From the asset specificity discussion, it has become clear that the key ingredient of the firm is not necessarily ex-ante specific investment. Rather, it is the fact that cooperative activity creates surplus over individual endeavors. Surplus creation leads to specificity and bargaining. Therefore, the phrase *surplus specificity* will be used to denote any input that is specifically needed for surplus creation over its substitutes.

With this in mind, one can conclude a firm's boundary should be set exactly to internalize all assets that are (a) surplus specific and (b) can be integrated. In contrast to GHM, integration is defined here as not just ownership, but also the making and customization of assets, intermediate products or other inputs by the firm. This means a firm's external inputs should consist of only two kinds: (1) non-integrable inputs, such as labor, which broadly includes unskilled, skilled, managerial or entrepreneurial with varying degrees of specificity, and (2) non-specific inputs, such as fungible monetary capital. Labor cannot be integrated because the input can only be rented not owned, and it consists of a purposive entity's non-contractible behavior. Specific capital (possibly including machinery, plant and intellectual property) needs to be retained by the firm.[6] Non-specific capital can be purchased with monetary capital. Therefore, one can assume the only form of external capital input is monetary. It is the internalization mechanism that reduces the number of items that has to be priced or bargained over to a minimum. This is where the firm sets its boundaries and the inputs that remain external become known as *factors* distinct from other types of assets.

In the firm literature, there is a plethora of related terms. TCE and GHM often use terms such as assets and investments. Cheung (1983) characterizes the firm as transforming an intermediate product market into a factor market, but does not answer how intermediate products differ from factors. The analysis so far clarifies what should be internalized in the firm: any integrable input specific to surplus creation.[7] Note here specificity is defined in economic terms rather than by physical or technological characteristics. Thus, the terminology of assets, investments, intermediate products, factor, labor, and capital are reduced into a unified concept of input.

2.4. Incentive Structure

With a few noted exceptions, the firm literature historically has not focused on the special role of the entrepreneur (s). In the context of the firm, the role of the entrepreneur (s) can be understood as a *concentrated source* of surplus creation. Entrepreneur (s) are those that see a new way and its potential over the status quo (surplus). Without the

4 The Fisher first separation theorem results if (a) owners consume a negligible amount of the firm's product, and (b) the use of the output is exclusive. One counter example is the open source software community. Software users and creators tend to have significant overlap. The use of the software output is also non-exclusive. The open source community is successful without an exclusive property rights structure for the intermediate product or specific investments. There is no monetary measure of surplus (profits), either.

5 The GHM model also makes assumptions in that (a) the owner intervention constitutes ex-post contractible variables such as quantities to maximize total ex-post profits for both parties, yet (b) the owner cannot command all quasi-rent from the other party, instead a Nash bargaining solution results where the owner gives up 50% of the total surplus to the other party. This second assumption is critical in deriving the paper's result in the choice of ownership, because the 50% sharing results in sub-optimal decision by the parties on their ex-ante decisions. However, the fundamental thesis in the TCE literature is hierarchy reduces bargaining. If one assumes bargaining still holds under a hierarchy, and then relaxes the assumptions of symmetric information and zero cost ex-post re-contracting, then the Nash bargaining result does not hold. In fact the hierarchy holds no advantage over the market in that case, and surplus will dissipate. Instead, the importance of ownership is derived by analyzing each party's own maximizing incentives post integration in this paper.

6 Since physical capital is only integrated when it is related to surplus creation, the make-or-buy decision is readily derived here. Surplus specific inputs that cannot be substituted should be made rather than bought. Non surplus specific inputs are bought not made. For example, firms will not need to make office supply or transportation vehicles as these are not surplus specific.

7 After integration, the factor inputs may still not be priced exactly at the same level as its best alternative use due to efficient wage considerations. Efficiency wage can be considered a form of asset specificity and quasi-rent.

entrepreneur function, parties would be putting their inputs in their respective best alternative use. That is the reason for the entrepreneur's usual position of being near the top of a command hierarchy (*control*) – because the knowledge of surplus creation needs to be transferred directionally from the entrepreneur (s) to other participants in the endeavor. That is also the reason that entrepreneur (s) tend to be the primary driver of new firm formation. Surplus creation can come from many sources: (a) creating entirely new products, (b) finding new demand, (c) finding new uses for existing capabilities, (d) capturing monopoly rent, and (e) capital allocation, among others. The commonality here is surplus creation.

This has further implications in the distribution of ownership. If surplus creation is inherently a dynamic and uncertain process, the more relevant a party is to surplus creation, the more that party should bear its risks and rewards in the form of residual ownership (equity). That corroborates with the anecdotal pattern that entrepreneurs usually have large ownership stakes in the firm. In cases where a large number of participants are surplus specific, ownership is often diffuse, e.g., employee stock option plans in technology firms.

2.5. Command and Information Barrier

Lastly, like TCE, one also needs to address the command hierarchy observed in the firm. First, why does a firm deploys hierarchy? The answer must be that it is beneficial for surplus creation. A new unskilled employee joining a firm has no surplus specificity. She will take orders from superiors because she is being trained to be part of the firm's surplus creation mechanism. On an ongoing basis, her taking orders may facilitate the continuous evolution and adaptation of that mechanism. It is also through command that unskilled (non-specific) labor input becomes more specific. Therefore the direction of command and control must be from more specific inputs (such as the entrepreneur) to less specific inputs (such as new employees), a conclusion from GHM.

Second, as any firm's "secret sauce" creates an inherent incentive for competitors to copy, firms erect information barriers at its boundary so that the surplus creation mechanism is not easily divulged. Information barrier, therefore, again, is a result of the surplus creation nature of the firm, rather than an exogenous assumption.

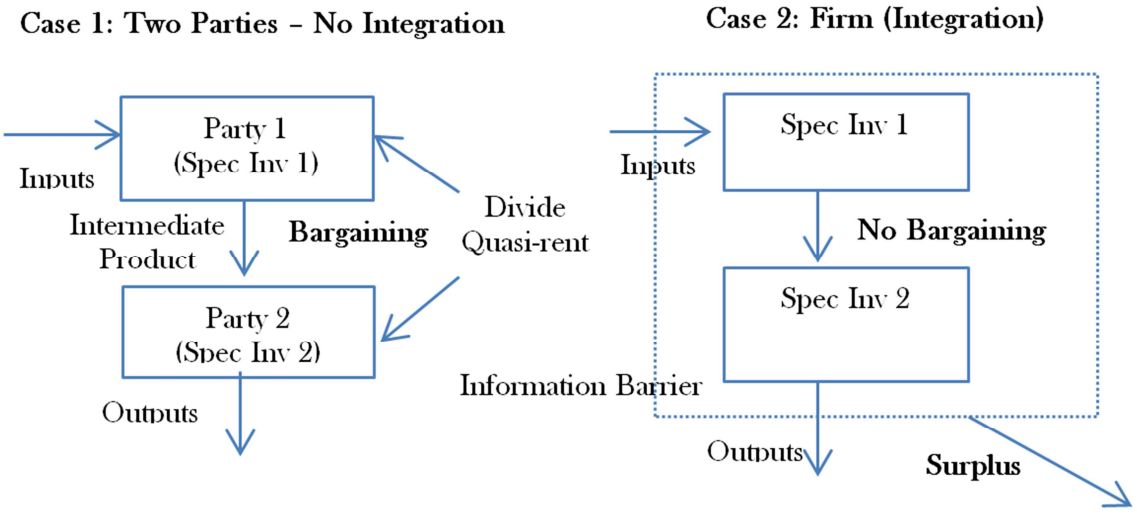

Figure 1. *Firm Integration of Surplus Removes Bargaining.*

3. Formal Models of the Firm

In this section a simple model of surplus creation will be constructed for strategic behavior, firm boundary and ownership. The main theorems from the model are: (1) existence of firm, (2) independence of firm from owners, (3) non-profit maximization of firm, (4) the party of higher impact should have more ownership, (5) the higher surplus specific party should be at the top in a hierarchy, but a hierarchy is still less efficient than joint ownership, and (6) firm boundary is set at surplus specificity of zero, all from typical diminishing return (concavity) assumptions.

Assumption 1: The following are baseline assumptions

(i) there are no taxes,

(ii) for *contractible inputs*, there are no cost of contracting, including price discovery, negotiations, contracting, monitoring and enforcement; therefore, the inputs are jointly determined by the parties to maximize surplus; this determination may involve a hierarchy in which one party has authority over another, this assumption can be justified in that contracting costs can be directly subtracted from the income and optimization can fully take account of such costs precisely because they are contractible by assumption,

(iii) for *non-contractible* inputs, the costs of contracting is prohibitive, therefore each party sets her non-contractible inputs to maximize her individual quasi-rent or income.

Each party involved is denoted by an index, such as k or l, for example, $k, l \in \{f, g\}$ where f and g denote different

parties, or $k, l \in \{K, L\}$ where K denotes capital and L denotes labor. The inputs of various parties are denoted as i, i_f, i_g etc, where i, i_f, i_g are inputs mutually observable and contractible at zero cost, with no information asymmetry. Here no distinction is made between ex-ante and ex-post inputs. The reason is simply that in cooperative activities, interactions are typically repeated. Artificial differentiations of ex-ante and ex-post inputs can serve to create contrived scenarios to derive a specific result that is already desired by the theorist.

Generally lower case letters will be used to denote inputs or intermediate outputs, and upper case letters to denote income or costs, while both lower and upper case letters will be used to denote the parties. For the input levels under specific equilibrium conditions, a superscript is added to denote it as a specific value, e.g., i_k^* for the optimal level for i_k. The total surplus for a joint activity among multiple parties is denoted by S. The surplus of a joint endeavor can be written as $S(i_f, i_g) = Y(i_f, i_g) - Y_f(i_f) - Y_g(i_g)$, where Ydenotes income from the joint endeavor, and Y_k denotes the income from party k 's best alternative income in their individual endeavors (opportunity cost). Obviously, each party's opportunity cost $Y_k(i_k)$ is not dependent on the other party's input. A division of the joint income by the parties is denoted by $Y_k^{\#}$, with $Y(i_f, i_g) = Y_f^{\#}(i_f, i_g) + Y_g^{\#}(i_f, i_g)$. Because i_f, i_g are contractible, the parties jointly maximizes S by simultaneously determining the optimal inputs, $\frac{\partial S}{\partial i_f} = 0$, and $\frac{\partial S}{\partial i_g} = 0$ at equilibrium levels i_f^* and i_g^*.[8] However, the division into the $Y_k^{\#}$ is indeterminate. In the deductions below, a 2-party formalism is used, but the multiple party analysis will exactly analogous.

Some additional assumptions are warranted as is typical in economics:

Assumption 2: All of the income $Y, Y_f, Y_g, Y_k^{\#}$ are assumed to have the normal concavity properties which corresponds to diminishing return assumptions: $\frac{\partial Y}{\partial i_k} > 0, \frac{\partial^2 Y}{\partial {i_k}^2} < 0$. In addition, no input leads to no income $Y(i_l > 0, i_k) > Y(i_l = 0, i_k)$ and vice versa. Cooperative efforts are accretive, for any two inputs, $\frac{\partial^2 Y}{\partial i_f \partial i_g} > 0$.

Two important concepts from before are now defined explicitly:

Definition 1: *Quasi-rent* is a measure of asset specificity at the equilibrium: $\Delta Y_k = Y_k^{\#} - Y_k$. If an input k is not asset specific, then $\Delta Y_k = 0$.

Definition 2: *Surplus specificity* measures how important an input is to the surplus generation. It is the difference of surplus at the equilibrium vs. setting that input at zero: $\Delta S_k = S(i_k^*) - S(i_k = 0)$. *Differential surplus specificity* is defined as $\frac{\partial S}{\partial i_k}$. Surplus specificity assumes that the input i_k cannot be substituted by another input.

These two distinct concepts are equivalent. Under only contractible inputs, if there exists surplus $S > 0$, and both inputs are surplus specific, $\frac{\partial S}{\partial i_k} > 0, k \in \{f, g\}$, then each party has quasi-rent. This is easily seen because by definition $S = \Delta Y_f(i_f, i_g) + \Delta Y_g(i_f, i_g)$. If either $\Delta Y_f \leq 0$, or $\Delta Y_g \leq 0$, one party will transition to the best alternative use of her input and the surplus generation will not occur. Therefore, $S > 0 \rightleftarrows \Delta Y_k > 0$. However, there is indeterminacy in the split of the quasi-rent $S = \Delta Y_f + \Delta Y_g$. This can be resolved by an upfront contract as the inputs i_f, i_g are fully contractible. Furthermore, based on prior discussion of specificity being mutual, it is also assumed that one party f does not have an alternative to contract another party g without that party earning the same quasi-rent $\Delta Y_g > 0$.

Next, as discussed before, surplus further depends on additional inputs a_f, a_g, which are non-contractible. For example, one can think of i_f, i_g as the quantity of input, and a_f, a_g as the quality of input.

Assumption 3: a_f, a_g are not separately compensated for other than in the division of surplus, since they are not mutually observable; and they are not factored in the opportunity costs $Y_k(i_k)$,

Assumption 4: a_f, a_g incur only private costs in C_f, C_g, with $C_k(0) = 0, \frac{\partial C_k}{\partial a_k} > 0$, $\frac{\partial C_f}{\partial a_g} = 0$, and in all regions near equilibrium, cost is convex $\frac{\partial^2 C_f}{\partial {a_g}^2} > 0$; in addition, C_k is not observable to the other party.[9]

Assumption 2 also applies to a_f, a_g as well, and surplus starts at 0, $S(a_f > 0, a_g) > S(a_f = 0, a_g)$ and vice versa. Now the joint surplus is revised: $S = Y(i_f, i_g, a_f, a_g) - Y_f(i_f) - Y_g(i_g) - C_f(a_f) - C_g(a_g) = \Delta Y_f(i_f, a_f) + \Delta Y_g(i_g, a_g)$, where $Y_k^{\#} = d_k Y - C_k$.

The condition for surplus maximization now also requires simultaneously $\frac{\partial S}{\partial a_f} = 0$, $\frac{\partial S}{\partial a_g} = 0$, in addition to $\frac{\partial S}{\partial i_f} = 0$, $\frac{\partial S}{\partial i_g} = 0$ at equilibrium levels i_f^*, i_g^*, a_f^*, a_g^*, where $\frac{\partial S}{\partial a_k} = \frac{\partial Y}{\partial a_k} - \frac{\partial C_k}{\partial a_k}$. However, because a_f, a_g are not contractible, S may not be maximized. Let us examine three cases.

3.1. No Integration

Since a_k is not contractible and only compensated in the surplus division, there is an incentive for each party to engage in bargaining, such as a misrepresentation of the

8 It is also possible that there are additional constraints on the range of inputs i_k, upon which the optimal solution might be on the boundary of the feasible set. This possibility is not considered here.

9 One can construe non-contractible input a_k as private "effort", and non-contractible cost C_k as the private cost of that effort. However, this set up can also model the alternative scenario where a_k is a private strategic behavior, and -C_k>0 represents private benefits (pecuniary or not). Such strategic behavior can include all manners of agency problem discussed in the literature, including obfuscation, misrepresentation and influence. In that scenario, the private quasi-rent is still the same, while joint income suffers from strategic behavior with $\frac{\partial Y}{\partial a_k} < 0, \frac{\partial^2 Y}{\partial {a_k}^2} < 0$, while the party gains private benefit of $\frac{\partial(-C_k)}{\partial a_k} > 0$, with $\frac{\partial^2 - C_k}{\partial {a_k}^2} < 0$. The equilibrium analysis leads to the same non-optimal result, with $d_k < 1 \rightarrow \hat{a_k^*} > a_k^*$, the party over uses the non-contractible input to derive her private benefit while only sharing part of the cost. This model hence incorporates the scenario of private benefit and shared cost in Jensen and Meckling (1976) Section 2.2 as a special case.

value of a_k and its cost C_k. The cost now includes bargaining which can be denoted as b_k. The quasi-rent for each party is reduced by the bargaining, $\Delta Y_k(i_k, a_k) = Y_k^{\#}(b_f, b_g) - Y_k - b_k$.[10] Bargaining can be in a transfer price negotiation from party f to party g, with an intended effect on the division of income Y into $Y_k^{\#}$. It is reasonable to make the following assumptions:

Assumption 5: bilateral bargaining has these effects: (a) bargaining benefits oneself $\frac{\partial Y_k^{\#}}{\partial b_k} > 0$, but (b) it hurts the other party $\frac{\partial Y_f^{\#}}{\partial b_g} < 0$.

It is easy to derive that the equilibrium under this assumption is bargaining will continue to go up until it eats up the entire surplus so that $\Delta Y_k = 0$, and hence at equilibrium $S^- = 0$, while the split of income $Y(i_f, i_g)$ is still indeterminate. So the solution to the problem is the following simultaneous equations:

(i) the contractible inputs are optimized: $i_f^- = i_f^*, i_g^- = i_g^*$,

(ii) bargaining takes up all the quasi-rent: $b_f^- = Y_f^{\#} - Y_f$, $b_g^- = Y_g^{\#} - Y_g$, and, $\Delta Y_k^- = 0$,

(iii) non-contractible inputs a_k^- are generally not at their optimal levels a_k^*; because of the additional activity b_k, neither party has an incentive or need to optimize a_k.

3.2. Firm Integration and Unified Ownership of Surplus

Now if one assumes a firm integrates the ownership of the surplus creation process and surplus specific inputs i_k. The surplus at the firm level is simplified back to $S_m = Y(i_f, i_g, a_f, a_g) - Y_f(i_f) - Y_g(i_g)$. The C_k are private so not included in S_m, while the net joint surplus is $S = S_m - C_f - C_g$.

Each party will share a fixed portion of the surplus of the firm at a ratio of d_k with $1 \geq d_k \geq 0, \sum_k d_k = 1$. This division of surplus can be done either through a division of the residual rights through ownership of the firm (equity), or through a formulaic division by contract (e.g., a commission or bonus), or even non contract reward (e.g., experience and reputation). Now party k's income $Y_k^{\#}$ consists of three parts, the opportunity cost of $Y_k(i_k)$ for selling her input i_k into the firm, her share of the firm surplus $d_k S_m$, and her private cost of putting in her private effort $C_k(a_k)$.[11] Therefore, $Y_k^{\#} = d_k S_m - C_k(a_k) + Y_k(i_k)$, $\Delta Y_k = Y_k^{\#} - Y_k = d_k S_m - C_k$, and $S = \Delta Y_f + \Delta Y_g = S_m - C_f - C_g$, as expected.

Given the upfront profit division d_k is fixed, bargaining will not increase one's share of the surplus and so will not exist. The equilibrium solution will now consist of the simultaneous equations:

(i) the i_ks are contractible, leading to $\frac{\partial S}{\partial \hat{i_f}} = 0, \frac{\partial S}{\partial \hat{i_g}} = 0$ as before,

(ii) the parties will each optimize her own quasi-rent on her non-contractible input, $\frac{\partial \Delta Y_k}{\partial \hat{a_k}} = 0$, or $d_k \frac{\partial Y}{\partial \hat{a_k}} - \frac{\partial C_k}{\partial \hat{a_k}} = 0, k \in \{f, g\}$.

Since the parties do not own all the benefit of their private unobservable input $\hat{a_k}$, it is less than optimal: $d_k < 1 \rightarrow \hat{a_k} < a_k^*$, due to the assumption that Y is concave on a_k. The degree of sub-optimality is influenced by the ownership percentage d_k (and no Nash bargaining is assumed as in GHM).

However, since there is no bargaining, each party still reaps a positive quasi-rent. From Assumption 2 and assuming $d_k > 0$, then $\hat{a_k} > 0$, $\hat{S} > 0$ and so is the quasi-rent $\Delta \hat{Y_k} = d_k S_m - C_k > 0$. This result is better than the bargaining scenario. Therefore, the need for property right integration is hereby proved with non-contractible inputs and bargaining:

Proposition 1 Firm Existence: Under Assumptions 1 to 5, $\hat{S} > S^-$, and $\Delta \hat{Y_k} > \Delta Y_k^-$. Or expressed in words: (a) Under non-integration, surplus and quasi-rent will be fully dissipated by bargaining, (b) with a unified property right of the surplus, and non-zero sharing, the parties will engage in cooperative activity which result in a non-zero joint surplus, and each party will also have a non-zero quasi-rent due to non-zero sharing. This is the reason for the firm.

Proposition 2 Firm and Owner Objectives Are Different: The firm's profit maximizing level of input (i_k^*, a_k^*) differs from the actual equilibrium where each party optimizes her quasi-rents $(\hat{i_k}, \hat{a_k})$. In addition:

Corollary 2.1 Non-maximization of Firm Profits: non-contractible inputs cause the firm not to maximize profits. $\hat{S} < S^*$ at the firm equilibrium.

Next, one asks the question of what division d_k of surplus would render the highest joint surplus. Intuitively, if the surplus specificity of one input dominates, $\Delta S_f \gg \Delta S_g$, then one wants $\hat{a_f}$ to be much closer to the optimal a_f^*, than $\hat{a_g}$ is to the optimal a_g^*. That implies, $d_f \gg d_g$.

For a more detailed analysis, one can use the comparative statics method. Given a particular ownership right division d_f and $d_g = 1 - d_f$, the conditions $\frac{\partial Y}{\partial i_k} = 0$ and $d_k \frac{\partial Y}{\partial a_k} - \frac{\partial C_k}{\partial a_k} = 0$ set the equilibrium solution of $\hat{i_f}, \hat{i_g}, \hat{a_f}, \hat{a_g}$. Therefore, the non-contractible inputs $\hat{a_f}(d_f), \hat{a_g}(d_g)$ are derived as functions of each party's ownership d_f, d_g. They are solutions to the equations $\frac{\partial C_k}{\partial a_k} / \frac{\partial Y}{\partial a_k} = d_k$.

With the concavity assumption of the income $\frac{\partial Y}{\partial a_k} > 0, \frac{\partial^2 Y}{\partial {a_k}^2} < 0$ and the convexity assumption on the cost $\frac{\partial C_k}{\partial a_k} > 0, \frac{\partial^2 C_k}{\partial {a_k}^2} > 0$, one can see that as d_k increases, $\hat{a_k}$ must increase. This means the functions $\hat{a_f}(d_f), \hat{a_g}(d_g)$ are monotonically increasing in d_f, d_g, intuitively more ownership leads to more efforts. One can rescale one of the functions $\hat{a_g}(d_g)$ so that $\frac{\partial \hat{a_f}}{\partial d_f} = \frac{\partial \hat{a_g}}{\partial d_g}$ at all points $d_f = d_g$, or equivalently $\hat{a_f}(d) = \hat{a_g}(d) = \hat{a}(d)$. This allows the

10 Also, one can assume the bargaining cost manifest itself in reducing non-contractible inputs a_k. If there is no connection between the input a_k and her quasi-rent, both parties will reduce it to zero where surplus is reduced $S^- = S(i_f, i_g, a_f = 0, a_g = 0)$. This is less than the integration outcome with $a_k > 0$. Proposition 1 will still result.

11 The Y_k cost can be in the form of the firm paying her a wage equal to her opportunity cost.

parties' non-contractible input functions to be evaluated on the same scale, and any sensitivity analysis to be comparable.

Assumption 6: non-contractible inputs are also concave: $\frac{\partial^2 \hat{a_k}}{\partial {d_k}^2} < 0$.

Now S can be written as a function of d_f, d_g, $S(d_f, d_g) = S(\hat{a_f}(d_f), \hat{a_g}(d_g))$. The task is to maximize S as a function of independent variables d_f, d_g, subject to the condition $d_f + d_g = 1$, The Lagrange multiplier is set as $S - \lambda(d_f + d_g - 1)$. This leads to $\frac{\partial S}{\partial d_f} = \frac{\partial S}{\partial d_g} = \lambda$, which is equivalent to $\frac{\partial S}{\partial \hat{a_f}} \frac{\partial \hat{a_f}}{\partial d_f} = \frac{\partial S}{\partial \hat{a_g}} \frac{\partial \hat{a_g}}{\partial d_g} = \lambda$.

If one input's surplus specificity is higher than that of the other input $\frac{\partial S}{\partial \hat{a_f}}(a_1, a_2) > \frac{\partial S}{\partial \hat{a_g}}(a_2, a_1)$ for any specific level of a_1, a_2. It can be written as $\frac{\partial S}{\partial \hat{a_f}}(\hat{a}(d_1), \hat{a}(d_2)) > \frac{\partial S}{\partial \hat{a_g}}(\hat{a}(d_2), \hat{a}(d_1))$, or $\frac{\partial S}{\partial d_f}(d_1, d_2) > \frac{\partial S}{\partial d_g}(d_2, d_1)$, because $\hat{a_f}, \hat{a_g}$ are of the same functional form, under the condition $d_1 + d_2 = 1$. Now start from $d_f = d_g = 0.5$. Because $\frac{\partial S}{\partial \hat{a_f}} > \frac{\partial S}{\partial \hat{a_g}}$, and $\frac{\partial \hat{a_f}}{\partial d_f} = \frac{\partial \hat{a_g}}{\partial d_g}$, it is not the equilibrium point. If the firm moves in the direction of decreasing d_f and increasing d_g, because of the concavity of $\hat{a_k}$, $\frac{\partial \hat{a_f}}{\partial d_f} \uparrow$, $\frac{\partial \hat{a_g}}{\partial d_g} \downarrow$. Similarly, because of the concavity of $S(\hat{a_k})$ and the monotonicity of $\hat{a_k}(d_k)$, $\frac{\partial S}{\partial \hat{a_f}} \uparrow$, $\frac{\partial S}{\partial \hat{a_g}} \downarrow$. Therefore, $\frac{\partial S}{\partial d_f} = \frac{\partial S}{\partial \hat{a_f}} \frac{\partial \hat{a_f}}{\partial d_f} \uparrow$, $\frac{\partial S}{\partial d_g} = \frac{\partial S}{\partial \hat{a_g}} \frac{\partial \hat{a_g}}{\partial d_g} \downarrow$. It is moving further away from the equilibrium point. Therefore, the equilibrium point must be on the side where $d_f > d_g$.

Proposition 3 Ownership and Incentive: When the non-contractible inputs are of the same scale $\hat{a_f}(d) = \hat{a_g}(d)$, if $\frac{\partial S}{\partial a_f}(a_1, a_2) > \frac{\partial S}{\partial a_g}(a_2, a_1)$, then $d_f > d_g$ at the maximal surplus S. This can be roughly stated as: the party with more surplus specificity in non-contractible inputs should retain more ownership of the surplus.

Note also that more residual rights lead to more quasi-rent. Thus, these two related concepts of surplus specificity and quasi-rent are re-united. One sees that the party with more surplus specificity will have a higher division of the surplus, and consequently a higher quasi-rent.

What does this imply about the boundary of the firm? Let us consider the situation in which a new party l with non surplus specific inputs becomes part of the cooperative activity. For the contractible input i_l, lack of surplus specificity means the firm can simply pay $Y_l(i_l,)$ without owning it. Similarly, if the non-contractible input has no *surplus specificity* $\Delta S_l = 0, \frac{\partial S}{\partial a_l} = 0$, the optimal solution for the input is set at $\hat{a_l} = 0$. This is because $\frac{\partial S}{\partial d_l} = \frac{\partial S}{\partial a_l} \frac{\partial a_l}{\partial d_l} = 0 < \lambda$ at any d_l. If any other non-contractible specific input $\hat{a_k}(d_k)$ has surplus specificity, $S(d_k)$ is monotonously increasing in d_k (or, monotonously decreasing in d_l). Thus $d_l = 0$ is the optimal boundary solution to maximize S. As a result, party l with contractible input i_l can be paid its opportunity cost $Y_l(i_l)$ with no sharing in the surplus. By definition, party l is now outside the firm boundary with no ownership stake, serves as a factor input, and has a price mechanism for its input set at its opportunity cost. Hence:

Proposition 4 Firm Boundary: $\Delta S_l = 0, \rightarrow d_l = 0$. If the surplus specificity of her inputs is zero, it will not be integrated into the firm, nor will its owner retain ownership in the firm. Or, alternatively, the firm sets its boundary at non-integrable inputs with no *surplus specificity*.[12]

3.3. Hierarchy

Under the assumption of hierarchy, one party (say f) retains fiat, and the other party g accepts command. In this scenario, party f will appropriates all surplus S, and the other party (say g) gets only its opportunity cost Y_g. This is equivalent to firm integration above where $d_f = 1, d_g = 0$. Based on this assumption, actual integration of ownership in a firm is not important because party g is only paid her opportunity cost.

The equilibrium solution under this scenario will be the simultaneous solutions of:

(i) the contractible inputs are still set as before: $\frac{\partial S}{\partial \tilde{i_f}} = 0, \frac{\partial S}{\partial \tilde{i_g}} = 0$,

(ii) for the non-contractible inputs, party f optimizes her non-contractible input because she reaps all surplus $d_f = 1 \rightarrow \frac{\partial Y}{\partial \tilde{a_f}} - \frac{\partial C_f}{\partial \tilde{a_f}} = 0$, while Party g gives no effort because she reaps no benefit $\tilde{a_g} = 0$.

Which one of the two parties $\{f, g\}$ should have authority in the hierarchy? This is determined to be the choice that would result in a higher total surplus. It is determined by the part of S that is sensitive to a_k, $Y(i, a_f, a_g) - C_f - C_g$. One of the C_k is zero by definition. Therefore, f should be in charge if $Y(i, \tilde{a_f}, 0) - C_f(\tilde{a_f}) > Y(i, 0, \tilde{a_g}) - C_g(\tilde{a_g})$. Hence:

Proposition 5 Hierarchy: Party f should be at the top of the hierarchy vs. party g, iff party f has the higher surplus specificity to its non-contractible input $\Delta S_f(\tilde{a_f}) > \Delta S_g(\tilde{a_g})$.

The GHM result is hereby reproduced, but under more generalized assumptions, and less general than Proposition 3.

Is this result better than the case of firm integration above? It depends on the *surplus specificity* of respective parties. From Proposition 3, for any non-contractible integrable input with surplus specificity, the optimal allocation involves a non-zero ownership of the surplus, $0 < d_k \leq 1$. Therefore:

Corollary 5.1 Inefficiency of Hierarchy: Hierarchy without profit sharing is in general *not* the optimal solution, for any non-contractible input with surplus specificity, $\tilde{S} < \hat{S}$.

This result differs from the hierarchy paradigm espoused by TCE, or GHM. Specifically, the model implies the more surplus specific party retains more residual rights. A full

12 Lateral integration and conglomerates are questionable precisely because of lack of obvious surplus specificity to the joining of two unrelated businesses.

hierarchy is the limiting case where one party has full surplus specificity and the other party does not. In realistic situations, both parties typically share some residual rights, possibly in the form of incentive payments (promotion and raises).

In general, there can be a much more detailed study of the form of the income function Y, and its structural relationship with the inputs i_k. For example, if the income function acquires uncertainty that depends on contingencies out of the parties' control, and the inputs i_k can denote features more general than the quantity of input, such as decision making, the study can lead to a deeper understanding of the allocation of *decision rights*[13] and resultant firm behavior.

In the simplest case, some decisions are made exclusively by only one of the two parties. A higher surplus specificity by one party implies that party has the more important decisions. This could be the reason for command. This conclusion is the same as what is sometimes called the "adaptation" theory of the firm, e.g., in Simon (1951). Authority is allocated by its result on surplus, or its surplus specificity. That connection between surplus specificity and decision rights is not explicated in the model.

If one assumes the income function dependence on the parties contractible inputs $Y(i_f, i_g)$ is such that certain decisions are exclusive, $i_f + i_g = 1, i_k \in \{0,1\}$, then the equilibrium point will consist of $i_f = 1, i_g = 0$ iff $Y(1,0) > Y(0,1)$. Therefore:

Corollary 5.2 Allocation of Authority: Authority on exclusive decisions in a firm is allocated to the party that whose input results in the higher surplus. Or in common parlance, competence should rise to the top.

If one views each party's input as specialization, decision making can be similarly viewed as a specialized skill that the surplus generation mechanism relies upon. Specialization favors the allocation of task to the party who is better at a particular task.

Richer decision rights analysis can be developed. For example, how diffuse decision rights can be and how that relates to the function Y is a subject in itself. In industrial specialization, surplus comes from meticulous division of labor, and each specialist makes few decisions and accepts authority from a single source. Decision making is highly concentrated in that scenario. In innovation driven surplus creation, individual initiatives can be the source of surplus and decision rights can be diffuse. Coordination and aggregation mechanisms need to be sophisticated and dynamic in such situations. There can also be situations where information needed for surplus creation is diffuse and spread out amongst many participants. Many decisions need to be made by the parties closest to the relevant information. In that case, the decision rights can be diffuse as well. This may be fertile ground for investigation. However, more analysis of decision rights allocations will not be pursued here due to the rich literature already present. This discussion is very different from Williamson (1971, 1975) in that the allocation is seen to be driven by the benefit in surplus creation rather than to remove haggling by fiat, which is instead done through the firm's unified property rights structure.

Note the model has also given strategic behavior operational meaning. So long as there are multiple self-interested parties involved in cooperative activity, each party will engage in maximizing behavior to her own benefit by choosing her non-contractible specific inputs. This causes the equilibrium to deviate from the maximal obtained when there is only one party or all inputs are contractible. This deviation is denoted by transaction costs in Williamson (1971, 1975), and by agency costs in Jensen and Meckling (1976). They are really the same phenomenon that by nature is very different from that of contracting cost described by Coase (1937).

Definition 3: The *cost of strategic behavior* in a cooperative arrangement is defined as the difference between (a) the total surplus generated if there are no non-contractible inputs or there is only one party, and (b) the total surplus generated with each party's individual maximizing goals; more specifically, in the context of a firm, this is called *agency costs*.

For example, in the case of no integration, the cost of strategic behavior is $C^- = S(i_k^*, a_k^*) - S(i_k^-, a_k^-) = S(i_k^*, a_k^*)$. In cases of integration, the agency costs are: $C^{\wedge} = S(i_k^*, a_k^*) - S(i_k^{\wedge}, a_k^{\wedge})$ and $C^{\sim} = S(i_k^*, a_k^*) - S(i_k^{\sim}, a_k^{\sim})$ respectively. Viewed under this lens, the problem of surplus maximization is equivalent to that of the agency cost minimization used in the literature, e.g., Jensen and Meckling (1976). Note, however, the concept of agency cost does not imply Pareto non-optimality. It is comparing the actual equilibrium to an "ideal" equilibrium that is not possible given the conditions of non-contractibility and multiple parties.

This concludes the study of the general equilibrium model of the firm without risk. The study of firm with risk, preference and incentive schemes beyond ownership can be taken up in a later paper.

4. Empirical Considerations

The analysis suggests several lines of empirical inquiry that can test the thesis. First is firm creation. One can examine the twin requirements of surplus creation and ownership integration. The association of firm creation with a stated purpose of excess profits should be apparent. The need for a firm in entrepreneurial activities should be linked to a cooperative process that involves multiple parties that creates surplus.

The theory predicts that if firm's surplus creation is a dynamic process with an evolving character that depends on a wide swath of employees, then ownership should be widespread. That is a testable hypothesis.

Second, the realm of corporate events can be examined. TCE focused on vertical integration and there have been empirical studies that confirm the relationship between asset specificity and integration. However, these studies have not

13 Note the distinction of decision rights vs. specific decisions here. What is allocated is the right to make determinations in cases of unspecified contingencies. This is a simplifying model which does not specify uncertainty.

necessarily clarified whether it is governance structures or ownership that cause integration.[14] In addition, lateral mergers, corporate spin-outs, joint-ventures should be subjects of studies, and the motivations as they relate to profits and asset specificity can be studied. The relationship of boundary changes to surplus specificity is a second testable hypothesis.

Third, to contrast the unified property right theory of firm versus the governance structure theory of the firm, studies of transaction costs in firms vs. market should be conducted. In the case of vertical integration, price bargaining is removed due to common ownership rather than fiat. Therefore, a study of conglomerates that allow divisions to retain profit centers and deploy internal transfer pricing should relate governance to transaction cost reduction of the selective intervention paradigm.

Fourth, firm dissolution can occur either due to failure or voluntary agreements by internal parties. In the first case, it should be due to negative surplus (or profits). In the second case, either one party is no longer surplus specific, or the surplus creation opportunity is no longer extant.

Lastly, that command should flow from more surplus specific input owners (skilled, managerial) to less surplus specific input owners (unskilled) should be a testable hypothesis as well.

5. Conclusion[15]

In this study, the firm is analyzed as an entity of surplus creation and unified ownership of that surplus and its associated specific assets. Instead of the firm as a governance structure, indeterminacy in the division of surplus from cooperative activities can be contracted by a unified ownership. The firm is a legal construct of property rights.

To summarize, the arguments for integration and setting a firm boundary are as follows. Surplus creates asset specificity, indeterminate division of surplus creates bargaining, which leads to surplus dissipation. Firm is a common ownership and property rights construct to remove that indeterminacy and bargaining. Hierarchy is needed for surplus creation in the transmission of knowledge or the optimal allocation of authority. Information barrier is created to keep the surplus creation capability within the firm. Non-integrable and non-specific inputs will be external and using the market price mechanism as factors. However, strategic behavior as defined by each party's individual maximizing activity persists in what is commonly known as agency costs. Such behavior is modelled as non-contractible inputs. Therefore, those who are most responsible for surplus creation should get the most residual rights. The primary driver of new surplus creation is entrepreneurship. Hierarchy should not be assumed to reduce opportunism and selective intervention is an incorrect paradigm.

This analysis of the firm opens up the firm's internal structure. It unifies the transaction cost paradigm/firm boundary question with the incentive structure paradigm/corporate finance question, resulting in unified equilibrium models that derive results in each framework. This analysis does not, however, discuss the firm's organization or governance structure. This is not the result of neglect. Organizations and governance structures are indispensable in cooperative activities.[16] Its study should also take individual's strategic behavior as a starting point in understanding the issues of governance structure, information flow and influence. There is extensive research in this area already.

As a final thought, strategic behavior is universally prevalent in any human interaction. This paper presents a framework to study such interactions. By separating individual's actions into contractible and non-contractible portions, interactions that have both cooperative and strategic components can be incorporated. This method allows us to analyze the firm as a purposive entity without always assuming a profit maximizing result. By examining each party's motive, rather than assuming only the collective objective is relevant, one can get useful theorems.[17] Cooperation happens when its benefits out-weigh the cost of strategic behavior. The firm is but one such example.

In the TCE literature, strategic behavior and agency costs are commonly looped in together with transactions costs. However, it is really a quite different phenomenon. Individual motive should be the starting point in analyzing human interactions, rather than as an exogenous quantity that is termed cost. The task of economic study is to explicate such motives, by giving up simplifying assumptions such as the profit maximization purpose of the firm. In this paper, the equilibrium model puts

14 For example, see Macher and Richman (2008).

15 In this paper, other potential economic motivations to integrate are ignored. Regulatory arbitrage is a full category in itself: (a) tax arbitrage includes sales tax consolidation, transfer prices setting across borders, inversion into lower tax jurisdictions, tax loss carryforward capture, and (b) other regulatory arbitrage such as jurisdictions choice exist for legal or regulatory oversight reasons as well. However, the discussion also sheds light on other commonly accepted arguments of integration. Economies of scale: it is not inherently clear that scale production requires integration. However, under the lens of asset specificity, each party owns specific asset in a more productive combined effort. Therefore, unified ownership of the production process is indicated in economy of scale. Technology integration: flow process does not require integration *per se* as Williamson (1975) discussed. Financing Efficiency: this argument is really the same argument as the economy of scale, applied to the financing activity of a firm. Due to information barrier, financing is a costly activity of a firm. However, the benefit of scale in financing has to be examined in relation to the agency cost of integration in this case. In-depth study is needed to compare the common examples of General Electric and Berkshire Hathaway with noted failures of diversified conglomerates in the 1980s. Pricing Power/Monopoly: to create pricing power by acting in unison, integration is not automatically indicated. For example, non-integrable inputs such as labor can still exercise monopoly power by creating unions. However, integration of ownership still reduces strategic behavior and competitors do benefit from split ownership in a larger company with monopoly power. This is again a form of surplus creation.

16 Williamson distinguishes mechanism design as ex-ante while governance structure as ex-post.

17 In surplus generating cooperative activity with non-contractible inputs, input owner will never reap the full benefit of her input. In the model, that is represented by partial division of surplus, $d_k < 1$. In Hart (1986), it is represented by Nash bargaining solution where each party reaps only 50% of the surplus. In the Marshallian sharecropping analysis, it is the less than full share of crops the tenant receives. They all represent essentially the same phenomenon that is behind strategic behavior, opportunism, and the agency problem. What is provided here is a unified analytical framework for this phenomenon.

firm and market on equal footing, by making symmetric assumptions of individual's incentive, regardless of ownership structure. That is different from most TCE literature.

These models assume zero contracting costs while analyzing the cost of strategic behavior in detail. The next step in model development is to incorporate the trade-off between contracting costs and strategic behavior costs on the variables a_k.[18] This will take the transaction cost economics paradigm one step further.[19] In addition, the analysis shall also incorporate features of misrepresentation of each party's true preference and her non-contractible behavior.

References

[1] Alchian, Armen A., and Harold Demesetz. 1972. "Production, Information Costs, and Economic Organization." *American Economic Review*, 62 (5): 777-95.

[2] Homstrom, Bengt. 1979. "Moral Hazard and Observability." *The Bell Journal of Economics*, 10 (1): 74-91

[3] Coase, Ronald H. 1937. "The Nature of the Firm." *Economica*, 4 (16): 386-405.

[4] Coase, Ronald H. 1960. "The Problem of Social Cost." *Journal of Law and Economics*, 3 (1): 1-44.

[5] Cheung, Steven N. S. 1969. "Transaction Costs, Risk, and the Choice of Contractual Arrangements." *Journal of Law and Economics*, 12 (1): 23-42.

[6] Cheung, Steven N. S. 1983. "The Contractual Nature of the Firm." *Journal of Law and Economics*, 26 (1): 1-21.

[7] Demetz, Harold. 1967. "Toward a Theory of Property Rights." *The American Economic Review*, 57 (2): 347-59.

[8] Fisher, Irving. 1930. *The Theory of Interest*. New York: Macmillian.

[9] Gattai, Valeria and Natale, Piergiovanna. 2015. "A New Cinderella Story: Joint Ventures and the Property Rights Theory of the Firm." *Journal of Economic Surveys*: doi:10.1111.joes.12135.

[10] Gibbons, Robert. 2005. "Four Formal (izable) Theories of the Firm?" *Jouranl of Economic Behavior and Organization*, 58 (2005): 200-45

[11] Grossman, Sanford J., and Oliver Hart. 1986. "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration." *Journal of Political Economy*, 94 (4): 691-719.

[12] Hart, Oliver. 2011. "Thinking about the Firm: A Review of Daniel Spulber's *The Theory of the Firm*." *Journal of Economic Literature*, 49:1, 101-13.

[13] Hart, Oliver. 2008. "Coase Lecture: Reference Points and the Theory of the Firm." *Economica*, 75: 404-411.

[14] Hart, Oliver. 2009. "Hold-up, Asset Ownership, and Reference Points." *Quarterly Journal of Economics,* 124: 267-300.

[15] Hart, Oliver and Moore, John. 1990. "Property Rights and the Nature of the Firm." *Journal of Political Economy*, 98: 1119-1158.

[16] Holmstrom, B., Milgrom, P., 1991. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics, and Organization*, 7, 24–52.

[17] Jensen, Michael C., and William H. Meckling. 1976. "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure." *Journal of Financial Economics*, 3 (4): 305-60.

[18] Klein, Benjamin, Robert G. Crawford, and Armen A. Alchian. 1978. "Vertical Integration, Appropriable Rents, and The Competitive Contracting Process." *Journal of Law and Economics*, 21 (2): 297-326.

[19] Knight, Frank. 1921. *Risk, Uncertainty and Profit*. Boston: Hart, Schaffner & Marx.

[20] Macauley, Stewart. 1963. "Non Contractual Relations in Business: A Preliminary Study." *American Sociological Review*, 28 (1).

[21] Macher, Jeffrey and Barak Richman. 2008. "Transaction Cost Economics: An Assessment of Empirical Research in the Social Sciences." *Business and Politics*, 10.1 (2008): Article 1.

[22] Mirlee, James. 1976. "The Optimal Structure of Incentives and Authority Within an Organization." *The Bell Journal of Economics*. 7 (1): 105-31.

[23] Modigliani, Franco, and Merton H. Miller. 1958. "The Cost of Capital, Corporation Finance and the Theory of Investment." *American Economic Review*, 48 (3): 261-297.

[24] Muller, D. and Schmitz, P. W. 2014. "Transaction costs and the property rights approach to the theory of the firm". *CEPR Discussion Papers*: 10207.

[25] Simon, Herbert, 1951. "A Formal Theory of the Employment Relationship." *Econometrica*, 19, 293-305.

[26] Spulber Daniel. 2009. *The Theory of the Firm: Microeconomics with Endogenous Entrepreneurs, Firms, Markets and Organization*. Cambridge: Cambridge University Press.

[27] Stiglitz, Joseph. 1974. "Incentives and Risk Sharing in Sharecropping." *Review of Economic Studies*, 41 (2):219-55.

[28] Williamson, Oliver E. 1971. "The Vertical Integration of Production: Market Failure Considerations." *American Economic Review*, 61 (2): 112-23.

[29] Williamson, Oliver E. 1975. *Markets and Hierarchy: Analysis and Antitrust Implications*. New York: Free Press.

[30] Williamson, Oliver E. 2000. "The New Institutional Economics: Taking Stock, Looking Ahead." *Journal of Economic Literature*, Vol. XXXVIII, 595-613.

[31] Williamson, Oliver E. 2009. "Transaction Cost Economics: The Natural Progression." *Nobel Prize Lecture*.

18 For example, see Holstrom and Milgrom (1991).

19 As an example, even with the manager/entrepreneur (s) sharing the residual rights of their surplus with outside monetary investors, the efforts devoted to information gathering and monitoring in the investing community is still tremendous. The entire investment and investment banking industry is devoted to this activity. That the fees of these activities represent a significant percentage of GDP is no accident. While the CAPM tells us that idiosyncratic risks for any particular firm does not command a risk premium, it is precisely surplus/profit seeking activity in the investment community that controls the entry and exit of firms into the capital markets. Without these activities, the assumption of CAPM will fail as adverse selection will dominate the capital markets with firms that do not generate surplus or warrant capital input.