A Note on Alternative ‘Theories of the Firm’

JC Spender, Kozminski University & Fordham University

Abstract

Firms are sometimes called the ‘engines’ of our democratic capitalist system. Yet what are firms and why do they exist when market arrangements are available to all? To many they seem perfectly obvious and unproblematic. Yet Coase famously asked this ‘Emperor has no clothes’ question in his youthful 1937 paper (Coase 1937). New Institutional Economics (NIE) provides answers. But it fails to deal with the intuitions Coase and his mentor Frank Knight brought to their analyses. We need to engage ‘Knightian uncertainty’ more fully than has been done thus far. The Note argues for four ‘types of uncertainty’: ignorance, indeterminacy, incommensurability, and irrelevance. Firms are social institutions that enable entrepreneurs to engage all types of uncertainty as they inhabit, even dominate, the economic sphere of life. Firms enable individuals to raise the value of their resources and capabilities. Profit plays little part in the analysis.

The Note

Many seek ‘alternative’ theories of the firm because they find the prevailing theories of the firm (ToFs) unsatisfactory, even offensive. For instance, some believe the ToFs of economists like Milton Friedman, Michael Jensen, and Willian Meckling conduce rapacious profit-maximizing and disrupt firms’ more social/ethical tendencies (Friedman, 1970; Jensen & Meckling, 1976). Of course, the claim that academics’ theories actually influence businesspeoples’ behavior, echoing Keynes’s famous aside about ‘dead economists’, runs counter to management educators’ widespread concern that their theorizing is ‘irrelevant’ to managers’ practice. We agonize about ‘rigor and relevance’ (Birkinshaw, Lecuona, & Barwise, 2016; Fincham & Clark, 2009; Hodgkinson & Rousseau, 2009; Kieser & Leiner, 2009; Kieser, Nicolai, & Seidel, 2015; Lee, 1999; Palmer, Dick, & Freiburger, 2009; Vermeulen, 2005; Worrell, 2009). Despite these doubts and bearing the fable about the cobblers’ children’s shoes in mind, we seem surprisingly reluctant to research our own business and assess the impact of management education on business practice. Thereby we ignore the divide in our
community between those who bemoan our theory’s bad influence and those who wring their hands over its lack of influence. Notwithstanding the heroic assumption that our theorizing is influential, even if often for the worse, the specific ToF being vilified is seldom identified, especially how and where it ‘goes wrong’ or ‘offends’. Critics find blanket criticism easier, especially by presuming private firms always choose to pursue their own interests and maximize shareholder value (MSV) rather than society’s benefits (Willmott, Spender, & et al, 2016). The critics claim a ‘more enlightened’ ToF would lead to ‘better’ business behavior. But which ToF is proposed? It is seldom characterized beyond replacing MSV by ‘more social’ interests, whether or not these can be determined or articulated. The ToF sought is defined more by the ‘wrongs’ of today’s, rather than by any ‘rights’ of tomorrow’s.

It is always risky to define something in terms of what it is not. Absent specification of the wrongs and rights the debate descends into a ‘culture war’. Much of our conversation is ‘political’ shouting, perhaps liberals to one side, neo-liberals to the other. This is too bad since the academic’s most urgent social role is to help clarify the public debate, to open up new language so that the discussion can move forward toward creating better social arrangements. The academics’ other roles, such as passing knowledge on to a new generation or discovering new knowledge, is secondary, especially so when the knowledge being created or transferred does not seem to help those in the ‘real world’ beyond our ‘ivory towers’. For some time, most obviously since WW2, the management academic community has sought to clarify the debate about firms, managing them, and managing their interactions with various stakeholders and others by framing it with rigorous testable theories (Backhouse & Fontaine, 2010). This is a positivist project that has not gone well, at least it has not much clarified or changed managerial practice. Rather it has helped institutionalize business management as a quasi-scientific discipline in spite of its negligible real-world impact. It serves the academic community well but does not clarify the ToF puzzle.

Meanwhile, especially since the crises of 2008 and 2020, there is a pervasive sense business interests everywhere have become excessively powerful, protecting and advancing private shareholders’ interests at the expense of the interests of the ‘rest of us’. Once again, after quieting in the post-WW2 boom years, social inequities have reappeared as major issues in every aspect of social and political life - education, military service, healthcare, incarceration, gender discrimination, etc. Many find parallels with the era of the Robber Barons and the Gilded Age at the end of the 19th century, the times before national political action, such as the Anti-Trust and labor
movements, pushed private investors' powers back towards a more equitable balance of public and private concerns. Thomas Piketty's highly academic 685-page volume of empirical evidence about the 'baked in' divergent rates of return on capital and on labor became a surprising best seller (Piketty, 2014, 2015). Putting racial issues to one side yet again as too intractable, complementary volumes characterize business management as a political matter rather than a scientific one (Boushey, DeLong, & Steinbaum, 2017; Delsol, Lecaussin, & Martin, 2017). Many of those criticizing the current ToFs do so on political grounds, pondering whether capitalism, even liberalism, is increasingly irrelevant. If so the vast majority of the theories underpinning business school (BSch) research and teaching need to be replaced by radically new thinking not yet spelled out.

Some critics appeal to anti-capitalist Marxist or Socialist ideas, enraging the 'free market' devotees. This Note does not attempt a political view to justify the private firm's future political and economic roles, or its demise. Rather, it deals with our attempts to develop a clearly capitalist but more managerially relevant ToF. With luck, this might help clarify the political debate as well, giving private firms' managers and shareholders get a better sense of what might be done to address the rising social concerns about their activity. The Western economies' concerns about monopoly that led to political constraints on firms at the end of the 19th century are rising again with Amazon, Google, and so on. But there are also new concerns about firms' ecological and a-social behavior. Likewise, there is renewed effort to establish business management as a socially validated 'profession', renewing the 19th century effort (Khurana, 2007).

But questions abound. Is the alternative one simply replacing MSV by MEV (More Ethical Values)? We can go with the exhortation be more ethical, but does this get us anywhere near a more relevant ToF? Is the calculus of MSV relevant to the values we are to value more? Are these new values unidimensional and quantifiable or pluralist and qualitative? Is the methodology of economic science relevant to the ToF we seek, or should aesthetics and/or faith be its basis? Are the ethical burdens of business practice the same as those of professional practice elsewhere; medicine, engineering, education, military, etc.? Are institutional contexts not shaped by their 'materiality', as well as their ethos and history? Are their languages and rhetorical practices the same? Surely each institutional context has its unique language and ethical foundation; the Hippocratic Oath versus West Point's "Duty, Honor, Country"? Along these lines this Note accepts the 'separation thesis' in the sense that business management differs from other professional fields. But the case for dismissing the thesis is strong, given many presume ethics are universal.
Against universality we can probe how business institutions differ from the social institutions of the other 'spheres of life'; religion, politics, aesthetics, etc. (Weber, 1970). This surfaces management education’s most serious weakness; that none of the ToFs available help us understand managerial practice beyond its most trivial and generic sense of 'decision making'. On the one hand we have situations in which managers exercise ‘managerial power’ or ‘fiat’. But power pervades society. It is not peculiar to firms, though they are indeed ‘zones of power’, as are all social institutions. On the other hand, we have managers as rational resource allocators and optimizers, setting up relations between people and between people and resources. Again, intendedly rational decision pervades society, making our biases and other ‘cognitive failures’ to be fully rational an aspect of all human activity, not peculiar to firms. Are firms different, and if so, how?

One reason to focus on the ToF, rather than on the political, social, and legal institutions that constrain firms and pressure them to be ‘better’, is that there is no agreement about which ToF best explains why private firms exist in the first place and hence how their activities might be justified given society’s constraints. Many analysts presume firms are inherently legitimate, ‘neutral’ production functions or production arrangements fully justified by the social needs served or the goods and services they provide. Their firm is ‘ethically neutral’, a rational apparatus, a scientifically-designed ‘tool’ in the hands of a management team that chooses to use it towards ends, social or ethical, or otherwise. Gas ovens are fine for baking bricks and bread, but not people. The tool notion puts the analyst outside the ‘neutral’ firm imagined, not part of it. The ethical challenge being condensed into re-orienting the management team’s decision-making about the tool’s use. There are tool issues about making it more efficient, masking the ethical issues around valuing efficiency over, say, ‘ready-to-hand-ness’ or its ‘affordances’ or work-place climate. Likewise, there are ‘internal’ ethical issues when, for instance, agency issues intrude, and greedy managers prioritize personal gain over both shareholder return and social service. Other analysts focus on labor relations, evaluating the individual costs, the bargain of contributions and rewards for those employed, or on other social costs. This leads towards ‘stakeholder theory’ as a ToF. There are various problems with this, such as how to ‘trade off’ the incommensurate costs and benefits of those involved. There are further problems about forecasting future costs and benefits in an uncertain dynamic environment. But even then, stakeholder analysis presumes rather than explains or justifies the firm as an inter-agent arrangement.
The Current Inventory of ToFs

As we know, in 1937 Coase asked the economics community “Why do firms exist?“, pointing out that we had no good answers whenever market arrangements were available (Coase, 1937). New Institutional Economics (NIE) proposed some answers to be discussed and later rejected (Klein, 2000; Klein & Sykuta, 2010; Langlois, 1989; Leathers, 1989; March & Olsen, 1984; Ménard & Shirley, 2008; Powell & Bromley, 2015; Rutherford, 1989, 2011; Vromen, 1995; Williamson, 2000). Notwithstanding, NIE-based ToFs may eventually supplement those ToFs presently in vogue. In BSchs today the leading ToFs are those of management teachers rather than of economists, in spite of the economists’ political power within the school. For the most part management teachers simply assume firms exist, visible all round us and ready to be theorized. They seem little troubled by the doubts about the firm’s ‘nature’ that post-Coasian economists now admit.

Gareth Morgan’s charming Images of Organization offers eight models favored by BSch teachers, variants of the two basic ToFs that underpin management education today: bureaucratic theory and ‘community’ (Morgan, 1997). These are children of Ferdinand Toennies’s distinction between Gesellschaft and Gemeinschaft (Toennies, 1971). Max Weber took up Toennies’s ideas in his analysis of the four modes of rationality he observed in human affairs: means-ends rational action (zweckrational), value-rational (wertrational), traditional (its justification lost in the mists of time), and affective (generated by a charismatic) (Kalberg, 1980). Bureaucratic theory is at the limit of zweckrational; all choices are strictly rational or intendedly so. It has been greatly criticized for leading towards ‘disenchantment’, an appropriate way of characterizing the life of ‘organization man’ (Whyte, 1956), and work as ‘toil’ rather than ‘human flourishing’ (Paulus PP.II, 1981). Contrasted against bureaucracy some presume ‘community’ means individuals ‘fully engaged’ rather than strictly in their organizational roles, choosing to associate and participate, perhaps urged by charismatic leadership, albeit with uncertain consequences or benefits. Others see firms as communities (markets) of a rational sort arising from individuals’ free choices about ‘selling’ their labor for promised returns, wages, promotions, healthcare, etc. under ‘incomplete’ or ‘open-ended’ contracts. At the same time the contexts of power and politics in which these ‘deals’ are made are down-played; for instance, seeking work in a ‘company town’ or merely to get health-cover otherwise unobtainable.

Many critics of current ToFs focus on the ‘dehumanization’ of bureaucratic theory, arguing
community is more 'humane', 'ethical', and socially oriented. Their arguments are often buttressed by the plentiful business-page evidence of business's unethical or anti-social behavior. But presuming bureaucracy is the ToF being considered has analytic consequences. Weber was dismayed by the encroachment of zweckrationalität into most, if not all, 'spheres of life'. Disenchantment is not merely about adopting a strictly instrumental approach, it also reflects a positivist belief that everything important about life can be explained/understood on the basis of reason and logic, that there is no mystery left in business as there is, say, in the aesthetic sphere. The disappearance of other 'rationalities' leaves bureaucracy as an 'ethically-neutral' or abstract tool in the hands of its owners and managers, even when 'agency' problems separate them. The firm has no 'ethical nature'. Or rather, its ethics are abstract, merely those of the persons managing as they guide their tool this way or that. Much of the business ethics literature focuses solely on the ethics of the firm's managers, stripping away any sense of the firm's inherent ethics, echoing the argument that 'guns do not kill, people do'.

Such 'value-free' analysis is paradoxical, presuming findings universal yet still relevant to the specific uncertain 'habitus' being inhabited. With the strawman of bureaucracy as our leading ToF, attempts to show managers actually pursue MSV are fraught with assumptions about whether managers are 'rational' in the zweckrational manner presumed. There are further assumptions about measuring and maximizing shareholder value. Given an economy not yet (if ever) at equilibrium, how might shareholder value be determined in a way that is not contingent on further unknown developments? Surely every valuation is contingent on there being no unforeseen 'disruption' by new technology or government regulation? Thus, MSV is as much a 'straw man' as bureaucracy; neither doing much to explain or clarify managers' practice as they grapple with their firm's uncertain future. The more tractable part of the stakeholder debate is about the distribution of the firm's calculable costs and benefits among the various stakeholders presently identified, to consider why the owners' interests dominate all others' in spite of the corporate law pertaining. In particular stakeholder theory says little about the interests of those whose interests have been 'silenced', excluded from the analysis because they lack the power to take part in the debate among those with power.

In spite of management writers assuming all economists embrace MSV, the most cursory examination shows they do not, far from it. While economists are far agreed on their ToFs, they offer a variety. Unfortunately, this is mostly ignored by business management and ethics writers.
For instance, Fritz Machlup argued for three types of ToF: marginalist, managerial, and behavioral (Machlup, 1967). The marginalist model is abstract, the neo-classical economists’ theory. In contrast, managerialist theories focus on power differences among those taking part in the firm, separating owners, managers, and employees (and customers and suppliers). Some label this ‘principal-agent’ theory. The firm comprises distinct power groups with divergent interests. For instance, managers are presumed have the power to counter the owners’ power, or at least take advantage of their ignorance or inattention and then maximize their own rewards, whatever the consequences for owners and shareholders. One managerial ToF presumes managers strive to maximize firm growth because that increases their personal rewards rather than the owners’. Behavioral theories likewise see firms are made up of divergent interest groups. Their behavior being based on observation rather than on maximization or agency theories. Machlup based his analysis on the Carnegie Mellon authors Richard Cyert and James March, and saw owners and managers negotiating five divergent ‘objectives’; production, inventory, sales, market-share, and profit (Cohen & Cyert, 1965; Cyert & Hedrick, 1972; Cyert & March, 1963; Machlup, 1967:4). In practice ‘profit’ is a exceedingly slippery concept (Weston, 1950, 1954). In similar manner Robert Gibbons argued for four potentially formalizable ToFs; incentive theory, adaptation, property-rights, and rent-seeking (Gibbons, 2005).

**Non-Zero Transaction Costs**

Economists’ ToFs vary widely, neither as simplistic nor as ‘dehumanized’ as the MSV tag suggests. Among these many models one issue stands out, an echo of 18th century French economic thinking and the 2nd Law of Thermodynamics; non-zero transaction costs. The 2nd Law claims every ‘real’ process is irreversible and leads to unrecoverable losses; entropy rises until the system dies a ‘heat death’. If firms are analogous to thermodynamic systems, they can only persist by generating new economic energy enough to cover these losses.

Where does the 2nd Law fit in? Coase is credited with introducing the notion of transaction costs (TCs). But he wrote little about firms or their TCs after his 1937 paper, moving focus onto the broader interactions between law and economics. Meanwhile the paper was ‘rediscovered’ in the 1970s, eventually being brought to more general notice by Oliver Williamson’s *Markets and Hierarchies* (Williamson, 1975). Williamson argued firms are bundles of transactions that incur
non-zero costs. Claiming to 'operationalize' Coase’s intuition, Williamson argued firms exist if and when managers are able to reduce the relevant TCs to less than would arise if those transactions were organized across the available markets. Williamson's ToF was heavily influenced by his Carnegie-Mellon training, seeing transaction costs arising the behavioral causes explored by Cyert, March, and Herbert Simon. Managers were able to reduce these TCs by bringing individuals into the managers’ zones of power or ‘fiat’, especially by preventing employees' ‘opportunism'. Williamson claimed empirical research showed his ‘transaction cost economics’ (TCE) would evolve into a formal theory (Williamson, 2005, 2016).

While Coase certainly brought transaction costs to economists’ attention, deserving his 1991 Nobel, Williamson, awarded a Nobel in 2009, may have missed what Coase had in mind. In his Nobel lecture Coase thanked Williamson for bringing his pre-WW2 work to others’ attention (Coase, 1992). But until the end of his long scholarly life Coase argued economists were heading blindly in the wrong direction, pursuing formal theories and thereby failing to address the practical questions he posed in 1937 (Coase, 2002; Coase & Wang, 2012). This amounted to Coase claiming the economists’ inventory, including the contributions of Williamson and the others in the NIE school, still had no ToF that met the needs of real-world managers better that Morgan’s hoary staples. Coase never directly criticized Williamson’s work. Williamson’s reasoning was in CMU’s neo-bureaucratic tradition, based on the comparative costs of creating order as close as possible to mechanical perfection. Rigorously comparing order arising from the invisible play of free market forces against order imposed by managerial fiat, typically in response to the other employees’ cognitive limits and failings. Managers, it seemed, did not suffer opportunism or bias. Order is presumed a virtue in and of itself, perhaps spiritual, but also economic, reflecting neo-classical economics’ notions of equilibrium and maximized economic benefit.

The irony is that order by fiat, drawing on an entirely different power-centered tradition, is also read as an economic virtue, without any justification (Wiebe, 1967). This echoes Scientific Management’s 19th century reasoning and methods (Spender, 1996). Order and the resulting organizational efficiency contrasts against and supersedes market efficiency. Theorists choose between aiming at perfect markets or at perfect organizations, deeming them equally valid economic aims. Firms only exist when markets fail. This raises questions about the different ToFs’ ethics, denying any presumed neutrality. The economist’s choice of ToF is unavoidably ethical. Absent total knowledge of the way things are now and might be later there is no ‘neutral’ option.
The battle raging today is between the ethics of free markets and the ‘economizing’ ethics of command and control; yet in the second managers decide for others who are excluded from the decision-making.

There is a different way to read Coase’s thinking about transaction costs. Few commenting on Coase’s ‘theory of the firm’ note Frank Knight’s *The Economic Organization* (Knight, 2013). This book is not highly cited, even in comparison to the dwindling citations to Knight’s *Risk, Uncertainty, and Profit* (Knight, 1921). Knight’s 2003 book was assembled from teaching notes prepared while at the University of Iowa around 1923, immediately after gaining his PhD. This was a hugely important period in the development of Knight’s thinking about the nature of economics, especially about its ethics. He eventually took up a position wholly opposed to Lionel Robbins’s classic argument that economics was the study of the optimal use of scarce resources, potentially formalizable mathematically (Robbins, 1932).

Knight, in contrast took up an ‘institutional’ position. For him economics was the study of the alternative institutional arrangements that society contrived to enable ‘the economy’ to function. He argued the economists’ challenge was always to come up with something better than ‘common-sense’, especially the practice-based commonsense of businesspeople and politicians. Knight’s 2003 book implies contrasting the way markets and firms facilitate economic interaction. Coase became an institutionalist in the Knightian tradition. He was a student at LSE in 1932 when Knight’s lecture notes were being circulated in mimeo form among the economics students. Though Coase never cited them it is reasonable to assume he read them or heard other students, such as his friends Abba Lerner and Ron Fowler, talking about them. Coase wrote that he learned most of his economics from his fellow students. This claim is strengthened by knowing the only LSE classes he took that touched on economics were those from his mentor Arnold Plant, by no means a mainstream economist (Coase, 1991a, 1991b, 1991c, 1991d). While Coase attended some of Knight’s classes during his visit to Chicago in 1934, he made little of them. His critiques of Knight’s ideas in the 1937 paper remind one of dogs catching the postman’s leg and refusing to let go. Something stuck in his craw. Eventually, of course, Coase stated Knight’s thinking influenced his own more than did any other economist.

We can surmise Coase regarded ‘transaction costs’ as ‘institutional’ phenomena rather than as computable costs in a rigorous maximizing model. Note transactions costs would be un-
differentiable from ‘factor costs’ if they were computable, as Williamson presumed. Methodology is at issue here, Coase thinking institutionally, Williamson thinking positivistically. Coase intuited transactions costs were generally, at least in part, non-computable, the practical costs of creating institutions, more precisely those of institutionalization. Such costs are excised or ignored in formal models. Where did Coase get these institutionalist ideas? Coase’s oeuvre has almost no discussion of his own ‘methodology’; he does not answer this query directly. But he did not start by presuming bureaucracy and design. Rather, his 1937 paper presumed firms arose from entrepreneurial judgments. Note he mentions entrepreneurship over 20 times, transaction costs not once. He sketched a theory of entrepreneurship, albeit ignored in our current entrepreneurship literature. He also presumed managerial power (fiat) and employee subordination, alluding to Batt’s Law of Master and Servant (Batt, 1933; Coase, 1937:391). In these ways Coase’s under-articulated institutionalism shifted from formal economic theorizing, which he plainly never much cared for, and into entrepreneurial judgement and sense-making, non-formal modes of human thought, agentic and imaginative rather than objective and computational. Neither Knight’s nor Weber’s influence was cited. Yet one of the handful of remarkable papers Knight wrote during his Iowa years was The Limitations of Scientific Method in Economics (Knight, 1924). This borrowed from Weber’s analysis, most specifically alluding to the distinction between zweckrational and wertrational, the source of today’s contrast of objective and interpretive thinking. He also alluded to traditional and charismatic modes of reasoning. Knight was the first to translate Weber’s work into English. Knight’s Iowa papers fleshed out his earlier claim that economic thinking necessarily went beyond zweckrationalität to embrace the other modes of thought.

Institutionalism

Douglass North argued ‘institutions’ are the arrangements free individuals ‘create’ to engage, ameliorate, or perhaps exploit, the uncertainties in a specific sphere of their shared lives (North, 1971, 1990). The institutionalization process is not clear; institutions are not ‘designed’ with total knowledge to hand. Nor are they simply emergent, without agentic or entrepreneurial intervention. They are creatures of some mixed middle ground, of partial ‘fiat’. They are also communal rather than an individual’s artifact. Our market society’s institutions are a long way evolved from those of their primitive predecessors such as the medieval guilds. Karl Polanyi argued three kinds of social characteristics must be present before a market society can emerge:
'redistributive', 'general reciprocity', and 'householding', the last seeing the family as the fundamental unit of economic activity, as did Knight (Polanyi, 2002).

A market society has freedoms that give it vitality and openness, retaining an ongoing engagement with uncertainty. It can never be wholly institutionalized. Likewise, no existing culture, national or professional, can be fully described or theorized as if there was an analytic vantage-point outside it. Nonetheless, even if incompletely understood, a ‘society’ implies an identifiable sense of stability and coherence, so comprising a plurality of institutions and persistent un-institutionalized areas of activity. These may become more institutionalized later and go on to influence other existing institutions. DiMaggio and Powell’s popular story is of institutional influence rather than the processes of institution-creation Weber labeled the ‘institutionalization of charisma’ (DiMaggio, 2001; DiMaggio & Powell, 1983; Jepperson, 1989; Kraakman, 2001; Powell, 2001; Powell & DiMaggio, 1991; Weber, 1968). Institutional theorists recognize their literature has neglected the processes of creating institutions (Greenwood, Oliver, Sahlin, & Suddaby, 2008). In part these lacunae persist because important aspects of Weber’s approach to institutionalism, on which much current institutional theorizing is based, have been lost to contemporary scholarship.

Aside from his notion of charisma and institutionalization, Weber presumed six discrete spheres of life - religion, the economy, politics, aesthetics, the erotic, and intellectualism (Oakes, 2003; Weber, 1970). These are not fixed axiomatic features of every society; more historically established. They were a methodological convenience, not a ‘theory of human society’. Weber’s intuition was that a society of heterogeneous individuals could not be examined as a coherent objective phenomenon. Society was infinitely dynamic. Only the institutionalized aspects of a society were stable enough to be worth examining. Institutionalization renders an otherwise incomprehensible society comprehensible as islands of meaning are created. Contra this, positivistically-inclined analysts sometimes presume a particular axiom provides a total concept of society, a tendency evident among laissez faire enthusiasts who see society as a market arising either unbidden or humanly ‘designed’ from the twin axioms of ‘perfect markets’ and rational self-maximizing individuals. Of course, such a market is not a feature of human life or an inhabitable society, it is simply an anarchic dystopia arising out of a presumed model of the individual.

Whenever it seems useful to say ‘society’ exists (countering Margaret Thatcher’s famous quip ‘there is no such thing as society’) other ‘models of the individual’ must be adopted. Society's
institutions lie within various 'spheres of life', characterized by the various roles these differently defined individuals adopt; seekers of faith, of gain through exchange, of power and order, of beauty, of erotic satisfaction, or seekers of more persuasive ideas. Using the term 'society' then points to a sense of coherence across these various spheres, an obviously problematic claim. Each sphere is characterized by its own institutions, most of all by its languages. Each institution presupposes an embedded ethic in the sense that an 'ethic' always implies an existing society or, more specifically, an underlying ethic of the institutions constituting that society. For instance, 'politics' is the debate among those with socially legitimated power who strive to create widespread social arrangements that constrain others perhaps intending to maximize their own power rather than society's capabilities. They pursue a coherent and dominant orientation in the political sphere of life. That might be to maximize social control or to minimize crime or to prioritize the political power of specific religious beliefs, hence the importance of the principle of separating Church and State. In our liberal democratic society, maximum viable personal freedom is often proposed as the highest political goal, exemplified for most of us by the 'four freedoms'. But freedom is paradoxical, synthesizing freedoms 'to' as well as freedoms 'from'. Thus, laissez faire's dispersive tendencies must be supplemented by complementary arrangements that conduce collaboration and socialization.

Politicians pursue many different objectives, perhaps contradictory, demanding subtle trade-offs between principle and practice, between competing aims. Arrangements in the political sphere almost certainly 'irritate' those in the other spheres, for the spheres can never be isolated from each other, even under the most repressive political systems (Luhmann, 1995, 2002, 2003; Moeller, 2012). A modern society is a plethora of interacting institutions that conduce both stability and predictability, but also lead to change in themselves and the society's other institutions, all driven by their contradictions and irritations. For example, a free press facilitates both stability and change in society, one reason dictators work to lessen the press's freedoms to express views contrary to their own yet still demand they express their own. To shut down the press entirely is to elevate the influence of rumor and people's irrepressible hunger to communicate through whatever other institutional arrangements they can find, as banging the pipes when in jail.

Our current democratic capitalism harbors many different institutions in its various spheres of life. Among these, private business firms are of considerable, if not overwhelming, importance to a capitalist socio-economy. They are our economy's 'engines'. But their existence is contingent on
the arrangements in our system’s other spheres; legal, political, and religious especially. Ultimately a ‘market society’ is defined as one ‘dominated’ by the doings in its economic sphere. As noted earlier, Weber saw this as ‘disenchancing’ society, preferring ‘zweckrationalität’ over other modes of reasoning. But rationality is far from being the firm’s sole characterizing ethic; many institutions claim rationality (Lecky, 1910). A viable ToF requires something that differentiates firms from other rational institutions. Profit provides a clue but is too problematic to serve our purpose.

Knight saw a social analyst needed some basic ‘theory of economic activity’ before s/he could consider how the existence of an economy interacted with a society’s other spheres. That led to his belief (and to Coase’s) that a possible rigorous theory along the marginalist lines Alfred Marshall indicated. It would be based on two axioms: (a) our desire for more of what we value and (b) our tendency to value that less as we got more. Knight felt additional axioms, likewise rooted in human nature, would be needed before a ‘science’ of economics could be developed. In the meantime, until that theory appeared, he saw institutional approaches as ways of clarifying democratic society’s economic sphere. Hence, he asked a question an order of magnitude more general than the one Coase asked. Rather than Coase’s ‘why firms?’ Knight asked, ‘why an economy?’, the more so later in his life as he transitioned from economics back into ‘social philosophy’. Given this orientation it is no surprise that Knight wrote little about firms. Likewise, Coase moved on from his 1937 focus on the ToF as he absorbed Knight’s thinking more deeply, focusing more on modern society and the interaction of its law and economic institutions. Institutional methods followed when the analyst concluded no existing economy could be comprehended in toto, as neo-classical economists presumed. Knight argued his ‘useful’ economics was the study and comparison of the various persisting modes of institutionalization. Coase followed, though at a safe distance.

Rather than seeking a single all-encompassing ToF, today’s economists and management writers might make better progress towards a useful clarifying economics of establishing and managing firms by following Knight’s lead and contrasting the alternative economic arrangements arising our socio-economy. He identified seven and, with the possibility of a rigorous theory of economics in mind, focused attention on ‘free enterprise’ (Knight, 2013:19). These days we compare cooperatives, collectives, not-for-profits, worker-owned firms, Mittelstand, and many other business arrangements. The comparison can also lead to discussing alternative ‘political’ arrangements of power within the firm; top-down versus bottom-up, etc. But how to know which is ‘best’? How can we evaluate or compare these alternative arrangements’ economic consequences
when there is no known connection between political order and economic capability? In spite of our biases and passions, there is little evidence that tight managerial control is more successful, economically, nor are more open and less ‘dehumanizing’ modes. As mentioned, Williamson and other NIE theorists echo the bureaucratic idea that ‘economizing’ is about order, command, and control reducing the costs arising from employees’ failure to be completely rational. There is no evidence that supports this assumption. Jensen and Meckling touched on this as ‘monitoring’ costs they presume measurable and computable (Jensen & Meckling, 1976). Order may seem like rationality applied, but again does not differentiate firms from other instances of institutional order. Management may be the process of institutionalization, but something other than mere command and control or order must be the institution’s ethos.

**Knightian Uncertainty**

Coase may have intuited one route towards a realistic and useful ToF would be to focus on the non-computable transaction costs of managing. While he noted the costs of finding trading partners, legal services, and so on, he knew the subtleties of engaging others into the business. Employees might be engaged through ‘subordination’, but this would not work for the firm’s legal advisers. Coase had no business experience, but he trained in ‘managerial accounting’, known in the UK and Empire as ‘cost and works accounting’ (Wang, 2014). Given the division of labor, managers needed to understand how to manage the division of costs and benefit to best effect. For instance, which products in the firm’s product line were generating the best return? Which operations were most costly and ripest for improvement or outsourcing? Coase understood well the difference between financial accounting, the creation of snapshot balance sheets required by investors and tax authorities using ‘generally accepted accounting principles’ (GAAP), and managerial accounting, focused on internal resource allocation choices that required judgments about resources’ incommensurability and production flow timing. More fundamentally he may have been reflecting the managerial work of assessing and engaging the situation’s Knightian uncertainties and the non-measurable costs of doing this.

There are many debates about what Knight meant by ‘uncertainty’, especially as contrasted against ‘risk’. One often advanced idea is that Knightian uncertainty applies to circumstances that cannot be measured and computed, where there is a lack of quantifiable knowledge. This negative
definition is not much help, its weight depending on what it is not. Scholars of positivist disposition struggle to bring Knightian Uncertainty (hereinafter KU) into the realm of calculation nonetheless, sometimes with sophisticated mathematics and statistics without bothering too much with what Knight had to say on the matter. Their conclusions run counter to Weber’s (and Knight’s) thinking. As John Locke remarked, “the faculty which God has given Man to supply the want of clear and certain knowledge in cases where it cannot be had is judgement … The mind sometimes exercises this judgement out of necessity, where demonstrative proofs and certain knowledge are not to be had; and sometimes out of laziness, unskillfulness or haste, even when demonstrative proofs are to be had” (Locke, 1928:298).

KU calls forth our imagination. This cannot be defined or fitted into a causal model. We know imagination only through our own experience of it. We can only impute it in others, observe its effects but never the phenomenon itself as we experience it. The act of imagination can produce a sense of closure or certainty in situations otherwise under-specified, contradictory, or paradoxical. Contra positivists’ sense of what academic activity should be about, generating generalizations, dealing with peoples’ imagination explores what is specific to the act, what differentiates it from every other such act. It is also embedded in an instance of time. Generalizations imply classes of instances, time-free. An individual’s act of imagination is more ‘art’ than deduction, the application of a universal. Knight saw economic activity as an art-form (Knight, 1923a). At first sight this is simply tautological, judgment is ‘illogic’, what logic is not, and so cannot have a logical explanation - perhaps nonsense. The imagination cannot be theorized, no matter how many claim to know how to excite it with a ‘whack on the head’ or create the ‘prepared mind’. Like human agency, imagination is ‘a cause without a cause’, a boundary condition to our logical reasoning; the remarkable native ability we have to develop confidence in conclusions beyond demonstrable reasoning.

Weber’s pluralism led to ‘interpretive’ or ‘meaning-making’ approaches, presuming an individual exercising hir imagination active under conditions of KU. In such situations, bounded by the facts known and so pointing towards what is not known, our imagination or judgment steps in to ‘construct’ actionable meaning. But it is not unconstrained. It cannot be ‘whatever you want it to be’. It presumes facts and is constrained by them, the resulting absence indicating an ‘opportunity space’ bounded by what is known (Spender, 2014). In a pluralistic universe KU may also refer to the irritations we sense between the spheres of life enabling us to draw on our experience of other
spheres and acts of imagination therein. Lord Kelvin famously believed science depended on measurement and calculation alone, creating islands of scientific reasoning and rationalist comfort in the confusing and incompletely grasped universe we inhabit. Knight argued economics could not be useful if so conceived (Knight, 1924). Indeed, a completely understood world would have no economics, no bargaining based on differing interpretations of economic value, thus neither markets nor firms. The term 'knowledge' would lose its meaning, for that depends on experiencing 'not knowing', of discovering impediments to reaching our imagined goals. Knowledge is always 'bounded'. Economics presupposes the imagination active between 'knowing' and 'not knowing', different people arriving at different valuations whether or not these are based on scarcity or taste. Even dedicated scientists sometimes claim 'truth' lies less in facts than in the imagined beauty or simplicity of Occam's Razor. The canons of beauty and simplicity are not those of logical reasoning. This edges us towards defining entrepreneurship as the capacity to draw on 'knowledge' from the non-economic spheres of life, to engage the KU appearing in the economic sphere. This risks tautology once more; when entrepreneurs are defined as those with the capacity or 'trait' to be entrepreneurial, the old tautology that we know leaders by their ability lead, but in no other way, such as education, class, race, upbringing, inter-personal skills, etc.

While some may read Coase's 1937 paper as a failure to define entrepreneurship, it actually masked his intuitions about the practical impact of KU and how managers deal with it. Certainty drives rigorous analysis. In contrast KU shifts the emphasis from reasoning to acting, from reasoning to judging, thereby from the general to the specific - what some call the distinction between the nomothetic and the idiographic (Allport, 1962). Practice is what, where, and how gives rise to how we experience, one instant at a time. Time is central, not excised; every instant differs from every other. We may try to explain our experience as partially 'covered' by some general law, but something of the visceral immediacy of experience always escapes. A judgement is not the application of a general 'covering' principle. It is the imagination's interplay of the known and the not-known, and so experienced, 'known' in a different way. Conversely general principles arise as the situation's specifics evaporate and leave the generalization exposed, especially when experiences are aggregated. Knight argued fiercely against positivist methods that sought only the nomothetic. Usefulness meant being able to relate the general with the specific, to synthesize mutually incommensurate modes of thought with society itself.

The first step is to put the claim that valid human knowledge is always and only 'scientific',
about the reality 'out there', in its proper place. It is no more than a claim, a working assumption, a judgment about what it means to 'know'. Thus, following Weber - and the Greek philosophers - knowledge is what our imagination make or create as we reflect on our experience, an 'interpretive' notion. While KU cannot be rigorously or logically defined - by definition - we can illuminate it by pointing to our various practices as we collide with it in the pursuit of our goals. How do we set about 'making knowledge'? The positivist notion, presupposing a logical and thus knowable reality, frames the practice of making knowledge as one of discovering the supposed reality. This type of KU can be characterized as 'ignorance' of the knowable facts.

The method here is positivist science, in part by generating and testing hypotheses. Managers labor to discover the relevant business facts, to overcome their ignorance. But note the process involves bridging between the idiographic and the nomothetic, the instance of the test experience and its relation to the universal of a law. Most of positivist disposition are happy with Karl Popper's 'falsification' idea - the 'logical asymmetry' between the hypothesis that 'all swans are white' and the appearance of a black swan that 'falsifies' the hypothesis. Unfortunately, neither Popper nor his followers paid much attention to the illogicality of logical asymmetry (Ayer, 1971). We are stuck with having no direct certain knowledge of reality, as Descartes reminded us, nor do we have any language, even logic itself, that is free of our assumptions. Human language cannot ever mirror or 'correspond' with reality. As Knight pointed out, language is the most fundamental of all social institutions, without which society is not possible. We try to capture experience with language and must always fail to do so fully. As socialized beings we have a plurality of languages available to us, some more incommensurate than others. Thus, falsification is meaningful only when the language of the experiment or 'test' is incommensurate with the theorizing language in which the hypothesis is framed. Otherwise what we can say about the test is stuck within the tautology that the language of the theory limits what can be said. The relation between the experiment as a test 'instance' and the hypothesis's 'generality' is deeply problematic, always a matter of the researcher's judgment, never of logic, as the Duhem-Quine thesis showed (Boylan & O'Gorman, 2003; Sawyer, Beed, & Sankey, 1997).

The point here is that neither knowledge nor KU can escape human language. We use language to judge everything we claim to know. The types of KU are types of judging, arising because it is not given to us to know anything 'for certain' or, as Giambattista Vico put it, 'to enter God's Mind'. When we judging and frame a situation as one of ignorance of what we can know - 'when does the
flight leave?’ - we are ‘researching’, the practice of leveraging systematically from what we judge we know to frame the specific knowledge-absence/s encountered - we Google or call the airline. While academics, especially those of a positivist disposition, judge all uncertainties to be matters of ignorance, ordinary people and managers recognize other types of KU. Given Knight’s argument that economics is about interaction between ordinary individuals, even when shaped by social institutions, there is another fundamental type of KU; the ‘indeterminacy’ arising when two people with different views and values interact. As Martin Shubik argued, indeterminacy is the overarching metaphor of game theory (Shubik, 1954, 2002). Managers often engage indeterminacy by negotiating with the partners giving rise to it. Persuasion overtakes research. Note the distinction between situations of ignorance and of indeterminacy is itself a judgment. When a game-theoretic ‘strategy’ has been discovered, the indeterminacy collapses into ignorance of that strategy, as happens when one party knows the others’ bargaining strategy and limits (Fisher & Ury, 1981). The value of the distinction lies in guiding managers’ practice. In some situations, managers judge negotiation with others to be more fruitful than research into ‘facts’.

Third and fourth types of KU arise because all human knowledge is held in language, and every language strikes a balance between what is assumed (its axioms) and what can then be said, deduced from or contingent on those assumptions. Languages are ‘useful’ rather than merely entertaining when they are fit to real-world purposes, to the specific challenges the living face. But their plurality means, as with the Duhem-Quine thesis mentioned earlier, that their usefulness lies not in their ‘objectivity’ and what they capture about reality, than in how their incommensurability helps us frame what we do not know about the situation’s specifics and yet act purposively or ‘mindfully’. Hence the value of the educational method of ‘contrast and compare’, the ancient dialectics of colliding different aspects of what we know in the pursuit of deeper insights that might suggest more effective practice.

The value of dialectical discussion, even in those despised business meetings, lies in the insights generated as different views are ‘batted about’, especially when the process leads to new language that seems to grasp the situation and managers’ intentions better. Revolutions are born with novel language; absent new language change is impossible. Those in firms have to take part in meetings or risk losing contact with its idiosyncratic language, for that is a dynamic, ever-moving train. The senior managers’ primary task is to create, shape, and control the use of language within the firm, for that continuously shapes the attention and imagination of those participating. Thus, the third
type of KU, after ignorance and indeterminacy, is ‘irrelevance’. It arises from the lack of a practice-relevant language for capturing the experiences of the situation’s knowledge-absences and for transforming the imaginative responses into innovative practice. The fourth type of KU follows close behind. It is ‘incommensurability’. If action is to be reasoned and mindful rather than happenstantial, the plurality of languages being judged relevant to the situation must be synthesized into simple 'go, no-go' statements. Consider the plurality of anxious discussions in advance of D-Day - weather, troops’ morale, weapon preparedness, layers of attack and support, enemy preparations, enemy intelligence, and so on; all condensed into Eisenhower’s ‘Go!’

Knight’s uncertainty covers all types of knowledge-absence as they arrest action towards the goals chosen; both ignorance of what is taken as factual, and the result of divergence between individual interests (indeterminacy) and ways of expressing them (incommensurability and irrelevance). TheToF implied is the idiosyncratic language that characterizes a specific firm in a specific situation at a specific time with specific resource availabilities and shortages. It bridges between thought and action. It characterizes the firm as a rhetorical artifact, management as a rhetorical practice focused on persuasion rather than direction, instruction, or proof. Thus, for instance, the resource-based view (RBV) is deeply deficient in addressing only parts of theToF, ignoring the difficulties of establishing the specifics of the situation or of building the language that enables those comprising the firm to grasp the resources’ value in that situation. The RBV cannot escape being trapped in the tautology of VRIN, mirroring the traitist notions mentioned earlier, judging some resources 'strategic' because they are judged valuable to the firm rather than on the basis of their price or creation/acquisition cost. (Kraaijenbrink, Spender, & Groen, 2010). RBV authors often claim Edith Penrose's work as their theory’s foundation yet ignore Penrose’s famous comment that the value of resources to the firm lay in the services provided not their cost (Pattit, Pattit, & Spender, forthcoming; Penrose, 1959:25). Thus, Penrose did not offer aToF that addressed Coase’s question. She presupposed the firm’s existence and focused on its growth as its management team created more-useful languages with which to discuss the business situation's uncertainties and thus growth possibilities. But this undoubtedly reveals Penrose had a certain alignment with Knight, especially with his distaste for positivist thinking.
A Language-Based Theory of the Firm

This excursion into KU helps establish a ToF of the firm as an idiosyncratic natural language that shapes the participants’ attention, reasoning, and imagination. This contrasts with ‘formal’ ToFs that presume the firm is a bundle of known resources or ‘capabilities’ to rationally allocate known resources with known consequences, so denying imagination a place in the analysis.

Bureaucratic theory highlights the rational allocation of resources of known value towards known needs, supported by command and control measurement. Order is presumed to lead to productivity. Community-based ToFs are not so easy to connect to the economic sphere of life. While encouraging a non-abusive workplace seems like a ‘good’ in itself its economic consequences are still hotly debated. Post WW2 management education’s most comfortable canard, allowing it to pillory Scientific Management and 19th work, is that happy workers are more productive; for which there is no conclusive evidence. Notwithstanding, for community-oriented ToF enthusiasts the taken-for-granted causal link is from motivation to productivity.

Morgan’s two principal metaphors underpin the vast bulk of management education. Neither engages KU nor non-zero TCs. The search for alternative ToFs is driven by what these two models cannot speak about. One possibility is to focus on the contrast between them and the idea of synthesizing their incommensurate implications, seeking ‘trade-offs’. Every trade-off is idiographic, specific to time, persons, and place. Trade-offs are never general, for at that point the models traded merge into one. This makes for the most important moments in science, as when James Maxwell brought electricity, magnetism, and light into a single coherent model, or when Albert Einstein brought mass and energy together. In the social sciences we await the fruit of Max Weber’s project to synthesize economics and sociology (Weber, 2019). But the idea that scientists are dealing with a plurality rather than a single ‘covering law’ can be illuminating (Courvisanos, Doughney, & Millmow, 2016; Holcombe, 2008; Rescher, 1993).

One version for ‘managers as enquiring scientists’ is this author’s pluralist FOPS model (Simon, 1952; Spender, 1976). It framed strategizing as synthesizing financial, organizational, people, and society notions of the business situation. Another is the Balanced Scorecard that framed strategizing as trading-off between financial, internal organization, learning and growth, and customer objectives (Kaplan, 2010; Kaplan & Norton, 1996). A focus on synthesizing shifts the methodological emphasis from the positivist program of seeking universal covering laws to an
institutional or interpretive one where academics work to clarify where the actor’s imagination might be most fruitfully projected. Note the methodology chapter Michael Porter seemingly excised from his 1980 book that appeared later as two papers that showed his switch from his early academic identity as a mainstream economist to an interpretivist strategist, *en route* to becoming an adviser to governments (including Gaddafi’s) without declaring his politics (Porter, 1983, 1991). He contrasted the theorizing of his first self against the ‘framing’ of his second (Spender & Kraaijenbrink, 2011).

Knight’s exploration of uncertainty and its economic implications opens up further alternatives. Thus far the core appeal is to the 2nd Law of Thermodynamics, non-zero TCs, and the firm’s need to create new economic value if it is to exist and persist. Most presume a relationship between creating new value and being profitable, though that turns out to be more complicated than it seems. But Knight also saw ethics as both foundational and complementary to rigorous theorizing. Knight wrote all economic decisions are ultimately matters of ethics. In spite of being almost completely ignored by business ethics writers, Knight wrote about the ethics of economics from the very first to the very last (Knight, 1916, 1966, 1967). Since he saw economic activity as individuals acting on their judgment, as this Note fleshes out, it was shaped by their ethics. While ethics is often defined as systematizing, defending, and recommending concepts of right and wrong conduct, it draws attention to the beliefs an individual might have about what is right and what is wrong, and how this play out against social norms. The ethics of economics might well be grounded solely in the last, as implied by judging whether a firm’s purpose or economic activity meets society’s legitimate needs. Knight dismissed this approach with the first intuition in his PhD thesis, that the economic system shapes individuals’ and thus society’s needs at the same time as it helps them get met (Knight, 1921). This brought his institutionalism into play immediately as he asserted economic needs are institutionally defined, neither religiously, politically, or philosophically ‘natural’, nor biologically so.

Knight had little time for those who tried to bring the ethics of the non-economic spheres of life into the economic sphere, seeing that maneuver as a kind of ignorance that a close reading of Weber would cure. Thus, he found no value in the business ethicists’ familiar triad of deontology, virtue, and consequence. A great deal of his writing focused on the ethics and ethos specific to the economic sphere. Given the notion of economic needs above, the economic sphere’s ethics could not turn on the non-economic needs being addressed. If society chooses guns over butter, so be it.
That was a matter for the other spheres, political or religious perhaps. In many of his writings, such as *The Ethics of Competition*, Knight found the ethics of games informing, aligning him somewhat with Johan Huizinga's *Homo Ludens* (Huizinga, 1955; Knight, 1923b). Knight considered players, economic or otherwise, competed for the personal satisfaction and public acclaim resulting. Games were carefully created situations of uncertainty that could not be reduced to ‘strategies’ by game theorists. That would ‘disenchant’ them, rob them of their *raison d’être*. The ethics of all games includes ‘sticking to the rules’, ‘giving the under-dog a break’, ‘knowing when to stop’, and other homely and common-sense notions well known without any scholars’ help. Games generate uncertainties of all the KU types, especially when played by teams that need context-specific languages to shape their collaboration. Referees also need to know appropriate language if they are to manage the game. Game theory seeks to ‘disenchant’ our games, to excise the mystery that makes watching them so compelling.

Suggesting ‘playing the game’ as an alternative KU-driven ToF seems relevant to many real-world situations in industries such as oil, hedge funding, airlines, and other industries that are ‘competitive’ and not dominated by monopolists or government fiat, as in a planned economy. The game’s rules frame the firm’s language and relations with the situation. Porter’s popular 5-forces model can be reframed as an institutionalized game of rent-seeking within a set of rules defining customers, suppliers, new entrants etc. and government regulation, the glaringly missing 6th force (Porter, 1979, 1980; Spender & Kraaijenbrink, 2011). But the personal and public consequences of playing games do not deal with the most fundamental aspect of the economic sphere’s ethos, the need to cover non-zero transaction costs. The economic sphere is not one of abstractions and ideas. The Knight/Coase economy is in-the-world, a practice that engages most of the citizens in a capitalist democracy, increasingly facilitating and constraining their lives. These two features, the need to cover non-zero TCs and the concept of property, differentiate the economic sphere from the others. While games may explore the interplay of the participants’ ideas and strategies, they do not generate economic value, nor do they embrace property even though, as Adam Smith might have observed, star players are often rewarded handsomely, even grossly.

Knight’s intuitions about the differences between resources (defined by price in some market external to the firm) and the services they provide within the firm were revealed in his vigorous debate with the ‘Austrian’ economists over the nature of ‘capital’. One the one hand, it is a store of the fruits of labor, on the other an ‘eternal fund’ (Knight, 1935a, 1935b, 1935c). The processes of
production transform capital into another form of capital mirroring the First Law of Thermodynamics, the conservation of energy. Though neither the Austrians nor Knight are clear about where capital comes from in the first place. Penrose cited Knight’s *Risk, Uncertainty, and Profit* a couple of times, without explanation or critique, though her supervisor (Machlup) had engaged Knight in the capital debate, somewhat acrimoniously, and covered several of Knight’s works in the methodology lectures Penrose probably took from him at John Hopkins (Machlup, 1935, 1978). Penrose wrote her famous sentence about resources and services without citing Knight though virtually the same language can be found in Knight (Knight, 1957:xxvii; Penrose, 1959:25).

**Services**

But what is ‘service’? Any dictionary will serve up several suggestions. Services are provided by servants, including public services by public servants. ‘The service’ points to the military sphere of life, marked by ‘duty’. Work done as bidden by a master. The duty a tenant is bound to provide. The act of waiting at table or a paid member of a household. Professional services provide useful labor that does not result in tangible products. A central thread is that of ‘incomplete contract’, when the precise duty or labor to be performed is to be specified later, the antithesis of ‘spot contract’ what takes place outside ‘the immediately-clearing market’. A second is the absence of a good which can be separated from its provider and so marketed as goods can. Contracts for future service are best understood in institutional terms, contingent on a plurality of relevant and thus contingent laws (Horwitz, 1992; Steinfeld, 1991). The Knight/Coase/Penrose notion of service is (a) conditional on the firm as a social institution rather than as a bureaucratic machine or a society, (b) time- and mind-full, seen in a longer legal perspective than a spot relation or market exchange, and (c) a relationship between particular individuals rather than classes of individuals. The meaning of ‘service’ cannot be anchored on market equilibrium, as is the economic meaning of a ‘resource’ when tied up with its market price. The complementary ‘production’ meaning of a resource is tied up with ‘use’, but that meaning is contingent, specific to the firm which is as free to misuse the resource as to make good use of it. A hand-drill is of no use to the owner who does not know how to use it. Conversely, those with a hammer ready-to-hand see everything as a nail. Likewise, Coase began his *The Problem of Social Costs*, the paper that complemented his 1937 paper and secured his Nobel, by implying economic activity was always reciprocal, open, between specific
individuals (A and B) rather than between an individual and the impersonal market (Coase, 1960:2).

The openness of the service situation accentuates the need for the participants to borrow or create a language to facilitate and govern their interaction. Individuals A and B, and any others involved, have to develop a language in which they can interact by stating their negotiating positions and demands, deal with any relevant ignorance, and synthesize the competing values with ‘trade-offs’. Contract is a major element of the law. The processes are illustrated in Lianna Farber’s analysis of medieval trading (Farber, 2006). The participants identify what and how they value the property or service, must find the moment to consent to the exchange, and observe how the new language developed might ripple out into the community as, for instance, the first traders arrive with furs, establish a language and trading pattern that eventually appears fully institutionalized as customary. Thus, management's work on new languages to frame new possibilities radically re-institutionalizes their firm as a locus of idiosyncratic language. Managers are ‘institutionalizers’, especially when innovating, maybe acting rhetorically to emphasize their own charisma (do it because of me), or new disruptive technological possibilities (because it is more efficient), or because there are new customers and markets to be spoken to with, say, compelling advertising.

This suggests a dynamic, evolutionary, language-based ToF. Which is fine if the firm is understood as language alone, or more specifically, language that best illuminates the dynamic inter-individual relations that actually comprise the firm as a complex of practice. But this picture fails to address the non-zero TCs of all real-world practice that demarcate the economic sphere. Knight’s rejection of market-based valuation, his embrace of the fluidity and subjective nature of economic value suggests value-creation may be better understood as personal re-valuation. When A and B have developed their own ‘deal’ they have reached a win-win point at which both have revalued upwards what they exchanged, otherwise there would have been no deal. The total economic value available to them is increased. New property has been created; learning and the acquisition of property converge.

This extends Knight’s subjective view of value. Instead of economic gain being at another’s expense, as when property is rivalrous, Knight’s increase arises as the parties’ interaction lifts both to a new level, for instance to better appreciate the value of some natural resource. Again, this might seem to arise at Mother Nature’s expense, oil extracted and burned makes some people
richer but is an irreversible loss in the Earth's total resource inventory. Knight's view embraces our discovering how to make more productive use of the oil, that there is no objective valuation, no absolute limit to its usefulness to us, say as polyethylene rather than as heat. The argument for allowing the economic sphere to arise in a society is that economic activity can lead to learning how to make better use of all manner of property which is thus never entirely rivalrous, always basically non-rivalrous, only made rivalrous by us and our institutions. We can never know the ultimate use-value. This radically undercuts the neoclassical view of economics as the study of choice under scarcity (Robbins, 1932).

Knight's analysis complemented that of John R. Commons, especially the 77 page analysis of the concept of 'transaction' in his Legal Foundations of Capitalism, a book Williamson appears to have cited only once in his oeuvre, and then only cursorily and without comment (Commons, 1924; Madhok, 1996; Spender, 2018; Williamson, 1996a, 1996b; 1996c:386). Ignoring Commons's lengthy analysis, Williamson based his interpretation of 'transaction' on the briefer, popularizing discussion in Commons 1934 (e.g. Williamson, 1996a:371). Williamson concluded a transaction was 'an exchange' - such as might occur in any market (Williamson, 1975:124). The NIE notion is that economic transactions can occur within firms as readily as within markets. There is something exceeding strange about this, given markets are mechanisms of distribution while firms are mechanisms of disaggregation, distribution, production, and aggregation. NIE presumes they share some under-defined middle ground with, even given at their extreme firms are contexts of total power and markets contexts of its total absence.

The huge difference is that Commons articulated an institutional view in which every transaction is embedded in a five-fold social interaction between parties A and B, influenced by the presence of parties A' and B', the 'opportunity parties' who would have done a deal had A or B dropped out, plus C, the authority presiding over the deal, establishing and adjudicating the parties' various rights, duties, and obligations (Commons, 1924:65). This institutional analysis re-defined 'property' as a set of specific reciprocal relations between individuals, not as a tangible object standing apart from it owner, scarce, rivalrous, and ready to be exchanged. Thus Williamson 'disenchanted' Commons's (and Coase's) analyses with his CMU-based positivism and impolite claims to 'operationalize' them (Spender, 2018).

The institutional approach is an alternative ToF that synthesizes nomothetic and idiographic
elements and brings ethics into its core. Its merit is that it can illuminate management practice; its weakness is that it does not dictate that, as so many positivist theorists aspire to. Thus ‘the market’ is not an institutional concept despite Knight’s acceptance of this commonplace of neoclassical economic theory (Knight, 2013:23). Yes, there may be a physical or virtual market, a legislated and administered place that facilitates inter-individual negotiation. But ‘the market’ is not a notion that can illuminate the economic sphere for that comprises a seething community of heterogeneous individuals engaged all the modes of KU that actually spur their activity. These days the power of neoclassical economic discourse is so great that we have imbibed its notions of market and property. Yet, as Commons’s analysis showed, they are of little practical use to the very managers engaged in wealth generation. Where is the market so many talk about? Wall Street? Chicago? On the financial pages? In some high-speed trading virtual space? The market is an imagined linguistic concept, part of an economics discourse that does not understand itself for therein we find no answers to Knight’s question about the economy's existence, nor Coase’s questions about the firm.

Economic activity is institutional and time-contingent on individuals’ sense of and engagement with the KUs of their reciprocal relations with others, absent which they would not engage their attention and imagination and there would be none of the inter-individual interaction to be shaped by the idiosyncratic language that characterizes ‘the firm’. As suggested, their economic activity creates ‘economic value’ by persuading participants to revalue the assets and services they provide existing firms whether as customers, suppliers, or labor. The economic sphere presumes a notion of property that can be revalued within the economic process. Economic activity presumes property rights. Protected by laws and aspects of the legal sphere of life, in the economic sphere property’s nature is of relations rather than objects, even quasi-objects like patents. Coase discussed how a cave can be of use to many different parties. While these users may be subject to legislation, the economic issue is what the various prospective users are prepared to pay (Coase, 1959:25). While the law sets boundaries, it does not determine agents’ practice.

There is little indication that Knight interacted with Commons, his elder by 23 years, or paid much attention to his labor-oriented institutionalism. It may be that Knight never read Commons’s analysis of ‘transaction’, precipitated Commons’s response to the work of Wesley Hohfeld, an American legal scholar (Cook, 1919). Knight was more interested in the philosophical and sociological aspects of economics than in the legal aspects that fascinated Coase. Hohfeld transformed US corporate law, pointing out that a firm’s ‘property’ comprised a bundle of
reciprocal rights, duties, and obligations rather than ‘real property’ or tangible exchangeable objects (Andrews, 1983; Fiorito, 2010; Hohfeld, 1913, 1917; Schlag, 2015; Vatiero, 2010). Title was always at the pleasure of the State.

Coase’s 1960 paper suggested familiarity with Hohfeld’s thinking, perhaps he encountered it on his trip to the US in 1948 to study the administration of broadcasting (Wang, 2014). Broadcasting was deemed a public service in the UK where the BBC managed the State monopoly. It was very different in the US where, after some misadventures, the government chose to administer broadcasting by auctioning the rights to use particular frequencies, now called bandwidth, to private interests. When Coase returned to the UK and suggested the BBC Trustees try making a market in bandwidth it was considered a joke in extremely poor taste (Coase, 1961). The BBC operated by charging all radio owners a fixed license fee; many chose to keep their ownership quiet (and still do after the BBC demanded TV viewers get a similar license - currently $195 p.a.).

Hohfeld’s approach was far more practical, avoiding the scofflaws, reinforcing Coase’s intuitions about the subtleties and importance of researching relations between a nation’s laws and its economics. Every arrangement was a new interaction of the legal and economic spheres, ethically burdened. With somewhat similar concerns, Knight often pointed out that slavery was grounded beyond the economic sphere, for in our democratic society no-one is able to sell themselves or others into servitude, a legal and a political matter. Once stripped of their legal and political citizenship, and so transformed into property, slaves became a commodity to be bought and sold, and thereby revalue-able. It is seldom appreciated that slaves were among the most promising capital investments in mid-19th century America (Cooke, 2003; Rosenthal, 2018).

**What is ‘The Firm’?**

The Knight/Coase/Penrose firm is an institutional ToF comprised of inter-individual relations in the economic sphere. That much seems obvious. Less obvious is that such firms cannot be illuminated by the management educators’ most popular metaphors - bureaucracy and community. Yes, a firm is ‘organized’ is thus ‘an organization’ and so susceptible to ‘organizational analysis’. But there are organizations in many spheres, religious, legal, political, military and so on. Only in the economic sphere must dealing with non-zero TCs be fundamental.
The implications are commonsense. Those engaged in the firm's practices are collaborating, just as Commons's parties - A, B, A', B' and C - are collaborating in 'revaluing'. But Commons did not consider non-zero TCs. The Knightian and Coasian firm is a different entity that must 'siphon off' part of the value being created by the individuals' collaboration to sustain itself. It is like a State that taxes the citizenry to support the provision of its legal, political, and social services, its public goods and services. Whereas a State may attempt to hold and protect its monopoly on the use of violence, the underpinning ethos of the political sphere, a firm establishes its own domain of power with the intent of control, as is well recognized, but also a domain of value-extraction (Hay, Lister, & Marsh, 2006).

The process of firm institutionalization parallels the creation of the division of labor that, as Adam Smith suggested, is fundamental to 'the enterprise', and the complementary power-based administration that brings resulting parts into a whole. Given a firm exists only because it engages carefully selected uncertainties, strategizing being the process of selecting them, these uncertainties must be similarly divided and distributed to the individual level. The firm exists as these individuals engage these mini-uncertainties with their imagination. There are wide variations across the firm. But each individual internalizes hir own part. While both Knight and Coase argued that absent KU there would be no firms, they did not follow up by noting that since the firm is nothing but ongoing dynamic relations between individuals being shaped by management's rhetorical practices, the KUs must ultimately be borne by these individuals.

To be an employee, individuals must bring this parsed uncertainty into their lives and thereby become part of the economic sphere, specifically the region or 'island' within the economic sphere (the economy) that demarcates the firm as a zone of language, power, and practice. Anyone who has worked in a firm knows the multi-faceted and ethically-challenging nature of this embrace. Coase's notion of 'subordination' missed its dynamic nature for employees can never give up their sense of self and so must deal with the dynamics of its collision with their role as employee (Coase, 1937:391). There is a balance between every individuals' engagements with the alternative spheres, dubbed the 'work/life' balance. Presuming the embrace of KU onerous or 'dehumanizing' misses the satisfactions of artistic production, the sense of competence at work well done, that animated Knight's arguments; work can be fun and fulfilling for aesthetic reasons, showing interaction between the aesthetic and economic spheres. Work can be a place of companionship, inspiration, succor. Invariably there is tension between the ethos of the economic sphere and that
of whatever sphere anchors the individual’s sense of self, perhaps political, perhaps religious. Just as the firm persists so long as there is KU in the economic sphere that its management is able to engage, to the employee’s engagement persists only so long as there is some uncertainty to be engaged. Being an employee is a condition of persistent uncertainty and anxiety, even when satisfying. Will the job disappear as a result of technological change or owners’ whim or competition? The arbitrariness of managerial power is a perpetual reminder of KU and the anxiety of inhabiting the economic sphere.

Clearly the processes of disenchantment roar ahead everywhere. Public services are being privatized, public resources plundered, public goods removed as public sector funding declines; everywhere the eagle eye of the haves scan the environment for needs of the have-nots that can be turned into business opportunities and monetized. Economizing is all. As McCloskey argues so powerfully the rise of capitalism has reduced the number of those living in poverty to historic levels (McCloskey, 2006). But there is little of the necessary examination of the role or nature of firms and thus of the precise mechanisms of these historic socioeconomic changes. There is screaming need for a different ToF that illuminates both the plusses and the ecological, social, and personal minuses of which we are becoming aware. The Knight/Coase/Penrose ToF is a huge step forward for it goes beyond explaining ‘the firm’ as an increasingly powerful apparatus to monetize others’ labor to the benefit of owners and their capital (Jordà, Knoll, Kuvshinov, Schularick, & Taylor, 2017). We now see it is also a legally and socially legitimated apparatus that monetizes the imagination of many to the benefit of the few.

Concluding Comments

This Note seeks alternative theories of the firm. The impulse driving the search is that the ToFs in use, underpinning the current discussion in BSch and elsewhere, are wanting. The recent attention to 'business ethics' has once again raised questions about why firms exist and whether they are 'neutral tools', whether the 'separation thesis' applies. If so, there are singular ethical burdens are about how the tools are used, perhaps for private gain rather than the public good. Most of the business ethics literature assumes this neutrality, leading its authors to examine managers’ choices in a triad of incommensurate analytic frames; deontological, virtue, and consequence.
The weakness here is that a real firm cannot be merely instrumental, objective, or neutral. The firm is not neutral, politically or ethically. A capitalist democracy's 'engines' are institutionalized into its political, aesthetic, religious and other spheres - unless, of course, the firm being analyzed is no more than a figment of the rigorously inclined analyst's imagination, thereby carefully distanced from the real world, an abstraction framed to be consistent, perhaps, with neoclassical economics and the 'blackboard theorizing' Knight and Coase protested.

The Knight/Coase/Penrose ToF treats a firm as a unique institution that inhabits a particular society's economic sphere. It is an idiographic notion, not a nomothetic one. Entrepreneurship, the process of creating the firm, is a process of unique institutionalization, perhaps the application of unique charisma, as Weber explored, or perhaps of design that goes beyond disciplined application of the known to embrace the situation's KU imaginatively. Leadership is achieving such institutionalization, quite different in the various spheres of life. Leadership's principal mechanisms are rhetorical rather than purely calculative. Its most proximate aim is the creation of situationally specific language that can attract and direct the imagination of those to whom the situation's KU can be parsed effectively.

All of this is riven with ethical dilemmas and trade-offs. There are the commonplace ethical dilemmas about directing the firm as a tool. There are subtler questions about sustaining its relationships with society's other institutions. There are questions about the nature, use of and limits to the inter-personal power within the firm. Bureaucracy is often welcomed precisely because it sets boundaries to the domains and practices of managerial power. But at a deeper level there are dilemmas about the dynamic relations between the individuals' selves and the duties, rules, and obligations of the roles they are assigned.
Bibliography


Langlois, R. N. (1989). What was Wrong with the Old Institutional Economics (and What is Still Wrong with the New)? *Review of Political Economy, 1*(3), 270.


