The Firm as a Subsociety

Claudine Gartenberg[1]
Wharton School
University of Pennsylvania

Todd Zenger
Eccles School of Business
University of Utah

January 5, 2021

We propose a view of firms as subsocieties that are shaped by shared perceptions of purpose and justice. Applying insights from political and legal theory, we discuss how employees are members of these mini-societies who have exchanged certain rights, such as formal authority over their work and compensation structure, for the benefits of membership in the firm. For these employees, expectations arise that are similar to those held by members of civic society at large; namely, that norms of justice and common social aims will be upheld. This view suggests a unique role for those who own and manage firms, as well as a different mechanism from standard theories of the firm for shaping firm boundary decisions. Boundary choices, in addition to their efficiency effects as described by existing theories, are also evaluated by employees for consistency with the espoused purpose and standards of justice of the firm. In general, this subsociety perspective suggests that theories of the firm should account for both societal and efficiency effects when assessing optimal governance of transactions.

Keywords: Theory of the Firm, Firm Boundaries, Corporate Purpose, Justice

[1] Claudine Gartenberg is Assistant Professor of Management at the Wharton School, University of Pennsylvania, 3620 Locust Walk, SHDH 2035, Wharton School, Philadelphia PA 19104; email: cgart@wharton.upenn.edu. Todd Zenger is the N Eldon Tanner Presidential Chair in Strategy and Strategic Leadership at the Eccles School of Business, University of Utah, Salt Lake City, UT 84103; email: todd.zenger@eccles.utah.edu. Claudine Gartenberg acknowledges financial support from the Govil Family and the Wharton School.

In his treatise, *The Wealth of Nations,* Adam Smith makes the insightful claim that the attributes of an effective "well-governed society"—one with the capacity to propagate a nation's wealth (Smith, 1776, p. 378) revolve around that society's capacity to elicit "the cooperation and assistance of great multitudes" in economic production (p. 26).[2] Smith goes on to describe the attributes of societies that enable such cooperation, previewing themes of justice, beneficence, and purpose that have become central to modern conceptions of well-functioning societies and organizations. But despite these discussions of society, the economic argument that has most shaped subsequent thinking is that these great multitudes divide and specialize and then cooperate within firms and through markets, with firms functioning in DH Robertson's words as "islands of conscious power in [an] ocean of unconscious co-operation" (1923: 85).[3] How this separation into conscious, directed cooperation within firms and unconscious, more spontaneous cooperation in markets occurs has subsequently emerged as the framing question that guides the theory of the firm literature.

Yet, arguably in this literature's push to understand the division of cooperation between firms and markets, Smith's early references to society-level attributes and the broader moral sentiments that enable (and disable) productive cooperation have been neglected. Today's theory of the firm says little about how non-monetary sentiments, particularly human desire for purpose and meaning in work (Frankl, 1949; Rosso, Dekas, and Wrzesniewski, 2010; Grant, 2012), as well as preferences for justice in dealings with others (Greenberg and Colquitt, 2013; Fehr and Schmidt, 1999; Ouchi, 1980) shape the separation of cooperative activity into firms and markets. Instead, firms simply emerge from sequenced decisions about how to efficiently organize exchange, "[becoming] larger as additional transactions are organised by the entrepreneur and [becoming] smaller as [the entrepreneur] abandons the organisation of such

[2] Smith provides a nice elaboration of this point in discussing all of the "cooperation" required to produce a woolen coat and concludes: "…without the assistance and co-operation of many thousands , the very meanest person in a civilized country could not be provided, even according to, what we very falsely imagine, the very easy and simple manner in which he is commonly accommodated." (1776: 22-23).

[3] Robertson continues: "The factory system itself, while it involves endless specialization of the work of ordinary men, involves also deliberate co-ordination of their diverse activities by the capitalist employer; and the head of a single big business to-day exercises a width and intensity of industrial rule which a Tudor monarch might have sighed for in vain." (p. 84)

transactions" (Coase, 1937). Consistent with this path of formation, the resulting firm is referenced as a governance mode (Williamson, 1975), an asset owner (Grossman and Hart, 1986), a nexus of contracts (Jensen and Meckling, 1976), or as a sub-economy (Holmstrom, 1999)—essentially the aggregation of transactions that the entrepreneur chooses to manage internally through access to formal authority and asset ownership.

Yet, this economic depiction of firms and markets is both atomistic and asocial. As Granovetter (1985) claims, within markets "actors do not behave or decide as atoms outside a social context" and within firms, they do not "adhere slavishly to a script written for them" (487). To build a theory of the firm with such an "unrealistic utility function does not provide a basis for understanding real organizations" (Simon, 1991: 35). Instead, humans are "socially dependent creature[s]"(Simon, 1991: 30) and the "clever institutional arrangements" described by economists are at best a "functional substitute" for an alternative, entirely social form of control (Granovetter, 1985: 489). To drive home this point, proponents of a more complex, social view of this firm-market distinction highlight cases in which firms generate more conflict and less cooperation than markets (Eccles, 1982, Uzzi 1996). Yet, the challenge for this fully social perspective, however, is that it fails to articulate a basis upon which boundary choices are made. As such, it is more a theory of organization than a theory of the firm, capable of providing clear predictions regarding the owner-manager's frequent make or buy choices. In other words, in this alternative world where social relations reign supreme, why should choices about formal governance or firm boundaries matter? And, why should we observe different empirical patterns of sociality across firms and markets?

One proposed solution is that organizational identification is what uniquely emerges through socialization within firms, as set forth by Simon (1951, 1991), Albert and Whetten (1985), Kogut and Zander (1996), among others. In this view, organizations establish in-groups—a sense of "we" rather than an "I", and, in the words of Simon (1991), this identification with a "'we'…allows individuals to experience satisfactions (to gain utility) from successes of the [business]" (pg. 36). Identification with an organization, however, is not a foregone conclusion. As Marx (1884) early noted, hierarchies are better known for fostering alienation than identification. Stifling bureaucracy, external monitoring and extrinsic

rewards crowd out identification, per self-determination theory (Ryan and Connell, 1989; Ryan and Deci, 2010; Enzle and Anderson, 1993). In that sense, any theory of the firm that appeals to organization identification must explain the mechanisms whereby the authority that firm boundaries afford principals sometimes fosters identification and robust cooperation and at other times foster alienation and complete non-cooperation . Beyond related efforts by Freeland and Zuckerman (2018), to our knowledge, existing theories of the firm have not taken up this task.[4]

Our aim in this paper is therefore to articulate a social theory of the firm or, more precisely, a *societal* theory of the firm, where purpose, justice and firm boundaries all feature prominently in explaining why cooperative sociality (or sometimes its complete obverse) frequently emerges within firms. In our theory, firms are not mere "subeconomies," but "subsocieties" defined by sociality profoundly shaped by the legally distinct form of control accessible within the boundaries of the firm. Our central argument is that firm principals (owner-managers), via their control as employers and asset owners, possess an enhanced capacity to not only shape incentives and allocate decision rights, as currently discussed in economic theories of the firm, but also to shape norms of justice and organizational purpose in ways that are vital to generating (or destroying) cooperation and identification within firms.

A fundamental insight that arises from this perspective is that authority enabled within firms functions as a social amplifier that, depending on how it is exercised, profoundly influences cooperatively-generated economic outcomes for better or worse. We then explore how the choices of the owner-manager around firm boundaries, membership (e.g, hiring and firing), and other strategic decisions, have subsociety effects within the firm that are quite distinct both from those outside its boundaries, and from those predicted by standard economic theories of the firm. In this subsociety view, the boundary choice becomes a decision about whether assuming the firm's expanded authority creates more economically functional social effects than economically dysfunctional social effects. This view implies that the economic outcomes that emerge within firms are fragile to sociality among its participants, and this sociality is in turn fragile to the actions

[4] Nickerson and Zenger (2008) take up a related task in their discussion of envy in arising within firm boundaries as a result of social comparison. We discuss the relationship between our theory and this approach in the discussion.

of the principal, as a principal's actions have amplified influence within firm boundaries. Our belief is that our theory of firms as subsocieties provides a capacity to explain a wide range of strategic actions encompassing vertical integration, acquisitions and mergers, and employee hiring and firing.

The organization of the remainder of our paper is as follows. We begin with (1) a brief review of traditional economic (or subeconomy) theories of the firm.[5] We then follow with (2) a series of examples that motivate various aspects of our argument. We then (3) introduce the concept of the firm as a subsociety, discussing the nature of cooperation within well-functioning societies. We focus on (4) two fundamental principles central to promoting social cooperation—justice and purpose—along with the benefits and costs associated with maintaining and violating them. We proceed by exploring (5) how these benefits and costs relate to theory of the firm. In doing so, we present a series of propositions that incorporate these social costs and benefits into the choice of what activities to internalize within the firm, and how these activities are managed. We conclude (6) with some discussion of the implications of our theory.

Firms as Authority-enabled Subeconomies

Since Coase (1937) the theory of the firm literature has generally viewed firms and markets as distinct governance modes for organizing economic production. The choice between these two modes is based on which alternative best mitigates costs that stem from opportunistic and boundedly rational behavior and enables the benefits of both the division of labor and cooperation. A firm therefore serves as a viable, and in fact arguably more common, alternative to markets as a means by which to manage economic exchange between parties. In that spirit, Holmstrom (1999) labeled firms as "subeconomies."

Much of this literature adopts a perspective consistent with Robertson's metaphor that firms are islands of conscious power within a larger economy shaped by markets and prices. Firms are a means of controlled cooperation, "a system of consciously coordinated personal activities" as Barnard (1938: 72)

[5] These theories form the starting point of our analysis, as they are explicit about motivational assumptions and the role of the principal. We address the relationship to knowledge approaches in the discussion section.

described. In this view, hierarchy and fiat are central features, whereby employees "must obey first, then seek recourse." (Summers quoted in Williamson, 1985: 249). While hierarchy and authority live at the core of most economic approaches, theories of the firm highlight different sources (and descriptions) of this distinct form of control.

Transactions cost economics emphasizes the role of fiat within firms in reducing the costs of productive exchange. This fiat is generally conceptualized in two forms. In what Gibbons (2005) calls the "rent seeking theory of the firm" (Williamson, 1971; 1979; 1985; Klein, Crawford, and Alchian, 1978), the use of fiat afforded by integration provides a more efficient resolution to haggling over uniquely generated value (appropriable quasi-rents). Alternatively, in the "adaptation theory of the firm", the argument is that authority may more efficiently facilitate adaptation that requires coordination or cooperation among multiple actors, particularly for those tasks that fall into Barnard's "zone of indifference" for employees (Williamson, 1971; 1973; 1975; 1991; Klein and Murphy, 1988; Klein, 1996). A similar argument is made in knowledge-based theories of the firm, where organizations enable greater cooperation, especially for complex tasks that make extensive use of tacit information (Demsetz, 1988; Conner and Prahalad, 1996; Nickerson and Zenger, 2004).

The property rights view highlights the role of asset ownership as the primary differentiator between firms and markets when exchange contracts are incomplete (Grossman and Hart, 1986; Hart and Moore, 1990). Ownership influences outcomes by defining which party has the residual claims to the output, thereby reshaping incentives of each party. In the case where a buyer integrates its upstream supplier, the firm-market decision changes the identity of that residual claimant from the supplier to the buyer, and in doing so changes the incentives of the supplying party. Integration leads suppliers to divert effort toward what their contract rewards, and away from non-contractible activities whose fruits now go to the buyer and not them.

In Holmstrom (1999)'s expanded articulation of agency theory, authority within firms is tied to the capacity to shape incentives. Firms are "subeconom[ies] in which the executive office has the power to regulate trade by assigning tasks, delegating authority, and delineating principles for how explicit and

implicit incentives are to be structured." (Holmstrom (1999: 76) This power of the executive office to shape the "rules of the game" within firms is particularly advantageous when multi-tasking and other informational asymmetries exist (Holmstrom and Milgrom, 1991, 1994), as the central office can provide weak incentives and appropriately narrow job designs to mitigate these market failures.[6]

These theories, while divergent in their underlying mechanisms, are united in their focus on authority, incentive and information differences as the primary factors that drive the firm-market distinction. As such, though the foundations of these theories differ, they share the "subeconomy" underpinning introduced by Coase (1937), in the sense that boundary decisions are determined by the most efficient organization of transactions which vary in their complexity, uncertainty, and costs to manage.

Illustrations

While these existing theories of the firm have been empirically predictive (Whinston, 2003; Lafontaine and Slade, 2007), there remains an abundance of cases where the causal drivers of boundary choices appear quite unrelated to these arguments. Instead, the observed boundary choices seem to reflect more social or societal-level factors that suggest deficiencies in our current economic explanations. To motivate our theory, consider the following illustrations:

Illustration 1

In 2017, Google launched a partnership with the U.S. Department of Defense focused on "algorithmic warfare" with the stated mission of accelerating the "DoD's integration of big data and machine learning." The initiative, part of Google's larger strategy of repositioning itself as an AI company, would earn Google an estimated $250 million over the medium term. While only a small number of Google employees were directly involved, word quickly spread throughout the company and

[6] Another approach is based on evolutionary logic (Nelson and Winter, 1982) and resource perspectives (Penrose, 1959). This theory of the firm conceives of firms as superior governance mechanisms for establishing routines and tacit knowledge than markets. These routines and knowledge in turn enables firms, under certain conditions, to respond to external conditions more efficiently than markets (Helfat et al., 2009; Jacobides and Winter, 2005; Teece, Pisano and Shuen, 1997; Amit and Shoemaker, 1993).

set off an "existential crisis" during the spring of 2018. More than 4,000 employees signed a petition demanding "a clear policy stating that neither Google nor its contractors will ever build warfare technology."[7] More than a dozen employees resigned. Many engineers refused to work on the project. "Do the Right Thing" stickers appeared across Google offices. "People who signed up to be tech heroes don't want to be implicated in human rights abuses," said one employee.[8]

Employee protests reached such levels that Google's management withdrew the company from the initiative, and issued "AI Ethics Principles" that defined explicit boundaries beyond which the company would not expand its AI business.

Illustration 2[9]

During the late 1990s and early 2000s, power and natural gas markets widely deregulated in the U.S. and Europe. As a result, the use of energy derivatives grew quickly as industry participants sought to take advantage of the liberalized markets. Many incumbent energy firms established internal trading units to trade derivatives both speculatively and to hedge their existing contractual commitments. As the derivatives markets boomed, top traders in these organizations were richly compensated for the high returns of their trades to the firms. While these compensation practices were standard among Wall Street-style operations, they were alien to the energy companies that now housed these groups. Internal challenges arose as employees in the primary, non-trading business grumbled that these "twenty-something college boys" were making millions of dollars off the hard work of the core business. After the collapse of Enron, the trading units were largely shuttered or sold to financial institutions. While the companies explained their decisions by the need to de-risk their operations and refocus on their core business, in private, many executives expressed to the authors a sense of relief at eliminating the large compensation differentials, and associated discord, within the company.

---

[7] https://www.nytimes.com/2018/06/01/technology/google-pentagon-project-maven.html, "Google will not Renew Pentagon Contract that Upset Employees," June 1, 2018, accessed January 19, 2019.
[8] https://www.wired.com/story/why-tech-worker-dissent-is-going-viral/, "Why Tech Worker Dissent is Going Viral," 6/29/2018, accessed January 19, 2019.
[9] Illustrations 2 and 3 arise from the authors' private conversations with senior executives of firms in the featured industry.

Illustration 3

The rise of the biotechnology industry during the 1980s and 1990s led to the entry of numerous small firms that controlled valuable intellectual property but lacked commercialization capabilities. This shift in turn led to a series of mergers within the industry by pharmaceutical firms that correspondingly lacked the specific patents and expertise but had substantial drug development and marketing infrastructure. The merger wave created a dilemma for managers regarding compensation policy toward both the incoming and legacy scientists. The incoming scientists often owned shares of the acquired entity that were closely linked to the specific drug or technology on which the firm was based. In contrast, the legacy scientists were generally paid with flatter incentives based on a combination of salary and shares in the larger firm, whose value reflected a multitude of projects beyond their own. As such, managers were faced with a Faustian choice of either tolerating internal pay differences among similar scientists or harmonizing the pay of the newly acquired group and thereby alienating them. This undesirable set of options was often referenced as a primary explanation for large pharmaceutical firms shifting away from acquisitions and toward greater reliance on alliances and in-licensing.

Illustration 4[10]

Following the September 11 terror attacks, law firms—particularly those based in New York City, where the most severe of the attacks occurred—noticed that many of their staff began to question their career paths. "We all kind of felt what I was doing didn't seem significant in the scheme of the world. It wasn't personally satisfying," attested one attorney to the study authors (Carnahan *et al.,* 2017). Firm departures – particularly to careers outside of law – substantially increased in the years that followed. This turnover, however, was notably lower within firms that had integrated pro-bono services into their firm. As another attorney stated, "'[Pro bono work] makes you feel alive. And that you are doing something worthwhile.'…[lawyers] reported greater satisfaction from promoting social reform or helping a

[10] This example is adapted from Carnahan, Kryscynski, and Olson (2017)

disadvantaged client than from wrangling over money." (Rhode, 2003: 447 as quoted in Carnahan *et al.*, 2017).

Illustration 5

In 2017, Skydance, a media company founded in 2010, expanded from live action films into animation. To fuel this expansion, in early 2019, Skydance hired John Lasseter, the co-founder of Pixar, to head the new division. Pixar had been responsible for some of the most successful animated films in history, including Toy Story, Cars, and Incredibles. Skydance CEO David Ellison characterized Lasseter as "a singular creative and executive talent whose impact on the animation industry cannot be overstated"[11] and as "responsible for leading animation into the digital age." The decision, however, created substantial controversy within the company. Lasseter had been forced to leave Disney (Pixar's acquirer) in disgrace over sexual harassment allegations involving a multi-year history of abusive treatment of female employees at Pixar.

At Skydance, the decision was perceived as executives implicitly sanctioning his behavior and "the mood after the announcement was one of shock and anxiety."[12] In one account, audible gasps could be heard when the news went out, with female staffers crying in their offices. One visible fallout was actress Emma Thompson's well publicized decision to pull out of the animated film, "Luck." In a searing letter to the Skydance CEO, she shared her perspective on the decision's harm to employees: "If John Lasseter started his own company, then every employee would have been given the opportunity to choose whether or not to give him a second chance. But any Skydance employees who don't want to give him a second

[11] https://www.washingtonpost.com/arts-entertainment/2019/02/26/i-can-only-do-what-feels-right-emma-thompson-releases-blunt-letter-about-ex-pixar-chief-john-lasseter/, "Emma Thompson releases blunt letter about why she won't work with ex-Pixar chief John Lasseter", Washington Post, February 26, 2019, Accessed February 26,2019.

[12] https://www.latimes.com/business/hollywood/la-fi-ct-lasseter-skydance-testcase-20190124-story.html, John Lasseter is attempting Hollywood's biggest #MeToo comeback. How's that going?, January 29, 2019, Accessed February 19, 2019.

chance have to stay and be uncomfortable or lose their jobs. Shouldn't it be John Lasseter who has to lose HIS job if the employees don't want to give him a second chance?"[13]

Illustration 6 (See Appendix A)

Illustration 7 (see Appendix A)

Summary

These seven illustrations (two of which are discussed in Appendix A) share important themes, which we summarize in Table 1. They focus on broad organization-wide social responses to a firm's expansion of activities, including hiring decisions, and associated boundary decisions. We observe firms choosing to jettison, suspend or continue owning and operating business lines based on how these businesses influence broad societal constructs, particularly organizational purpose and perceptions of justice. In these vignettes, Google was forced to scale back operations where expansion decisions violated some significant portion of the collective employee beliefs about what Google's purpose is or should be. Energy and pharmaceutical companies chose to outsource activities to avoid discord from internal pay imbalances. Law firms expanded their engagement in pro bono activities to provide their attorneys with a sense of meaning beyond their daily work, while in contrast Amazon and IBM discontinued portions of their business that violated employees' and other stakeholders' beliefs about their organization's appropriate purpose. Skydance, in branching into a new business, experienced substantial organizational costs due to a hiring choice that violated norms of fairness and justice of the larger enterprise. Finally, Disney was forced to grapple with how and whether to expand due to the consequences for perceived fairness that expansion and related incentive choices entail. We suggest that these vignettes are mere illustrations of broader social forces that shape the boundaries and functions of organizations and are poorly explained by existing theories of the firm. We turn now to developing our theory of firms as subsocieties.

[13] https://www.latimes.com/entertainment/la-et-mn-emma-thompson-john-lasseter-skydance-20190226-story.html, "Emma Thompson's letter to Skydance: Why I can't work for John Lasseter", Los Angeles Times, February 26,2019, Accessed February 26,2019.

<< Insert Table 1 here >>

Firms as "Well-Governed" Subsocieties

From Subeconomy to Subsociety

Our theory begins with the assumption that firms are more than simply loci of transactions. They are collectives of individuals, and as collectives, they are subject to the range of social forces that characterize human interaction. As the illustrations above demonstrate, these social forces do not arise randomly, but instead arise in patterns shaped by the boundaries of firms and the choices of those who manage them. We acknowledge that we are far from the first to characterize firms as social groups. Field-defining works such as Barnard (1938), Simon (1947), and Cyert and March (1963) focused primarily on social dynamics inside organizations. These social approaches later evolved into theories of firms as political coalitions (March 1962), and vehicles for collective sense-making (Weick, 1995) and identification (Albert and Whetten, 1985).

However, while powerfully descriptive of the influence these social forces carry, much of the research in this tradition has not focused on the "make or buy" question or, stated alternatively, how precisely the existence of the firm and its boundaries bear on these observed social outcomes, and vice versa. In this sense, these are social theories of organization and not theories of the firm. Theories in economic sociology, in fact, have explicitly challenged the entire premise that firm boundaries matter, arguing that the same sociality exists within firms may exist outside them. The "clever institutional arrangements" that economists emphasize are at best a "functional substitute" for deeply embedded social relations (Granovetter, 1985) that may exist rather equivalently inside and outside firm boundaries. While this approach has been increasingly challenged (Nee and Ingram, 1998; Bourdieu, 2000), the revised theories remain at best loosely connected to the theory of the firm.

One notable exception in which a social theory of the firm has been developed lies in the idea of organizational identity and firm boundaries. Herbert Simon in his later work (Simon, 1991) characterized the essential advantage of firms over markets as deriving from a combination of authority, rewards,

coordination (all covered by prior economic theories of the firm) and organizational identification. As he states:

> "Identification with a 'we'…allows individuals to experience satisfactions (to gain utility) from successes of the unit…Of course, identification is not an exclusive source of motivation; it exists side by side with material rewards and enforcement mechanisms that are part of the employment contract. But a realistic picture of how organizations operate must include the importance of identification in the motivations of employees." (Pg. 36).

Identification shapes cognition and attention of organization members. As Ocasio writes, "By serving as the backdrop of strategic deliberation", organizational identity shapes "both the issues and problems that firms consider and the appropriateness of the answers and responses to those issues and problems" (Ocasio 1997, p. 198). It also drives discretionary behavior that cannot be specified in formal contracts. "Identity creates a set of expectations about appropriate behavior for a particular organization" (King, Felin & Whetten, 2010: 295). Kogut and Zander (1996) translated insights about identification into a knowledge-based theory of the firm. In this view, which builds off works of Kant, Smith, Durkheim, and Weber, the primary benefit of firms is their coordination advantage in the use and production of knowledge. In that sense, the theory is more cognitive than social, as are related theories based on knowledge and learning mechanisms. It does, however, appeal to identification as the basis for why workers choose organizations over contractual relationships, proposing that "the boundaries of firms demarcate qualitative changes in the reservoir of social knowledge available to economic agents (i.e., people) because coordination and learning are developed within the organizational context of shared identities…firms provide a sense of community by which discourse, coordination, and learning are structured by identity." (Pg. 503). Those transactions inconsistent with shared identities are therefore likelier to be outsourced.

Our work follows this approach in spirit, however with an important difference. While we acknowledge that organizational identification is a core aspect of any social theory of the firm, we also hold that it leaves unanswered a fundamental sociological challenge: that identification is far from assured within firm boundaries. Many organizations in fact induce a strong sense of alienation among their employees, rather than the warm glow of identification discussed by Simon and others. We choose to

focus, therefore, on the specific social mechanisms associated with firm boundaries and the resulting managerial considerations.

Our view is theories of the firm should be expanded by characterizing firms as subsocieties shaped by authority accessible within firm boundaries. In advocating this view, while we echo the social critiques of economic theories of the firm—critiques that collectively highlight the substantive role that sociality and identification play in shaping economic outcomes—our aim is to develop a *societal theory of the firm,* one that identifies the significant role that authority within firm boundaries plays in shaping patterns of sociality (for better or worse), and the related role that sociality plays in shaping boundaries.

What is a (sub)society?

The concept of commercial enterprises as societies actually far predates our modern understanding of firms. The earliest known shareholder companies from two thousand years ago in the Roman Republic carried the society label. They were referred to as *societas publicanorum* and performed a range of commercial activities, from military supply to agricultural production and construction (Malmendier, 2009). Building on this millenia-old view, we similarly view firms as miniature societies circumscribed by firm boundaries, operating within a larger social context. We refer to these as "subsocieties," adopting the prefix convention of Holmstrom (1999), who labeled firms "subeconomies" because of their operation within a larger economic context. In this subsection, then, we discuss the concept and attributes of a society and why these attributes are usefully applied to firms.

The term "society" itself can be defined broadly as an ordered community of individuals formed for a particular purpose or activity.[14] This view originates from the Latin terms *socius*, or "companion," and *sociatas*, or "union for a common purpose."[15] Originally referencing autonomous tribes of the Italian Peninsula that allied themselves with the Roman Republic, a *socius* comprises a community of individuals

[14] A society, according to the Oxford English Dictionary, has two related definitions. First, a society is "the aggregate of people living together in a more or less ordered community." Second, a society is "an organization or club formed for a particular purpose or activity." We combine these two definitions into the one used above.
[15] OED and Cambridge Dictionaries, https://www.wordsense.eu/societas/

that has ordered itself toward a particular aim. The nature of this aim might vary widely from political or military to commercial, but critical to the composition of an effective society is a willingness to subsume individual desires into a larger society-level purpose or objective.

Modern firms still meet this definition of society. As Selznick (1948: 25) describes, firms are "rationally ordered instruments for the achievement of stated goals" that are on the one hand "an *economy*" and on the other "an *adaptive social structure*" with both formal and unwritten rules. We similarly recognize firms as not only sub-economies but sub-societies directing their power toward a common aim. In the context of firms, the aims are commercial in nature, rather than political, and the members are not citizens or subjects, but employees. This concept of firms as subsocieties is not just a relabeling of Holmstrom: a society is a fundamentally broader construct than an economy. Economies are systems to regulate production and trade of goods and services. In contrast, societies, as described above, are ordered communities of individuals with a common aim, characterized by (ideally) peaceful cooperation.

What constitutes a "well-governed" (sub)society?

While this term "well-governed society" originated with Adam Smith, many other Enlightenment figures, as well as political theorists since then, have concerned themselves with how to construct a functional civic society. Within this tradition, functional society is often contrasted with a "state of nature" with less than desirable cooperative outcomes, wherein humans interact without the benefit of formal government and civic membership. In one of the earliest discussions, Thomas Hobbes argued for a "Leviathan"—a powerful political force that orders society and wrests it from an otherwise dark state of nature where culture and commerce fail to exist and "the life of man [is] solitary, poor, nasty, brutish, and short" (Hobbes, *Leviathan,* 1651). In this view, the "state of nature" renders cooperation infeasible because man's inherent character inevitably leads to uncertainty and violence. Societies with centralized leadership, by contrast, provide a peaceful means of dispute resolution such that individuals willingly subordinate their self-interest to reap the benefits of cooperation. While this view of the society-nature dichotomy has evolved substantially since Hobbes, the general insight from this "social contract" school of

political theory is that well-functioning societies exist when members willingly exchange their rights for the benefits of cooperation as enabled by the institutional umbrella of government.

This insight then raises the following question: what are the main features of well-governed societies that enable them to avoid the ills that beset individuals caught in a disordered "states of nature", and instead enable this exchange of rights that support cooperation? This question lies at the core of political theory from Thomas Hobbes to John Rawls. Two features commonly arise as answers: i) a sense of justice, and ii) a sense of common purpose. While these two features are by no means a comprehensive set, their prominence prompts us to consider each in turn.

Justice in well-governed societies

In one of the clearest articulations of the requirements or obligations of a "well-ordered society," political philosopher John Rawls states that

> Among individuals with disparate aims and purposes, a shared conception of justice establishes the bonds of civic friendship; the general desire for justice limits the pursuits of other ends. One may think of a public conception of justice as constituting the fundamental charter of a well-ordered human association. (pg. 5)

In other words, one of the primary attributes of a well-ordered society is the maintenance of this "public conception of justice", what Rawls elsewhere calls "the first virtue of social institutions" (1971:3). Even Adam Smith, one of the original advocates of free markets, recognized the importance of justice in society, viewing it as "the main pillar that upholds the whole edifice" and its removal as destructive to the "immense fabric of human society". (*Theory of Moral Sentiments, Chapter 3, Para 4).* Advancing a similar theme, Gouldner (1961) references fairness or what he terms the norm of reciprocity as one of only two social agreements universally held by societies across time and cultures (Ouchi, 1980).[16] Economic institutions that fail to generate perceptions of equity whether through formal or informal means simply fail (Ouchi, 1980; Husted and Folger, 2004). This view also corresponds to extensive work in social psychology on the role of equity as a precursor to trust.

---

[16] Simmel's claim is that social cohesion demands patterns of reciprocity and that "all contacts among men rest on the schema of giving and returning the equivalence." (1950: 387).

As is evident to anyone who has managed employees, however, establishing fairness perceptions within firms as sub-societies is no simple task. Fairness is assessed individually, but through social comparison. Rewards personally received are compared with those that other relevant referents receive (Homans, 1961), with relevant referents defined as those similar or proximate along any number of different dimensions including spatial proximity, degree of interaction, and the availability of information (Festinger, 1954; Kulik and Ambrose, 1992; Gartenberg and Wulf, 2017; Obloj and Zenger, 2017). These social comparisons tend to be invidious or asymmetric in that individuals compare themselves to those receiving higher rather than lesser rewards (Martin, 1981; Nickerson and Zenger, 2008), making efforts to vary pay, even when linked to performance, particularly difficult.

If individuals reach conclusions that their treatment is unfair, they respond with efforts to remedy this condition (Adams, 1965). Individuals may reduce their efforts, realigning their efforts to match the magnitude of inequity felt (Adams, 1965). They may seek to alter the outcomes that they or others receive, actively voicing their options to those in authority positions (Hirschman, 1970; Nickerson and Zenger 2008). They may simply depart from the unjust subsociety (Hirschman, 1970; Kaczperczyk and Balachandran 2018; Carnahan, Agarwal, Campbell, 2012). Of course, all such behaviors clearly undermine employee cooperation in pursuit of an organization's goals or purpose. Behaviors within a subsociety are thus profoundly shaped by broadly shared perceptions of justice.

Purpose in well-governed societies

But justice alone is insufficient to compose a well-ordered society. Even Rawls, the father of modern theories of justice recognized societies ideally possess "a vision of the way in which the aims and purposes of social cooperation are to be understood. (pg. 8)" Thus, in addition to a shared sense of justice, well-functioning societies have a defined purpose. The seeds of this idea also extend back at least to Adam Smith, who argued that each actor within a society possesses "a principle of motion of its own, altogether different from that which the [designer] might choose to impress upon it." Therefore, if a society's designer aligns these divergent interests so as to:

> "coincide and act in the same direction, the game of human society will go on easily and harmoniously, and is very likely to be happy and successful. If they are opposite or different, the game will go on miserably, and the society must be at all times in the highest degree of disorder."

Organizations scholars emphasize a similar theme of purpose alignment. Barnard (1968:42-43) suggests that any "formal system of cooperation requires an objective, a purpose, and aim." Selznick (1957) suggests a leader's responsibility is to define "the mission of the enterprise" such that widely "accepted values …infuse the organization." Mayo (1945) focuses on mechanisms that provide this alignment, highlighting both "objective inducement" and social efforts that lead to "changing states of mind." Indeed Ghoshal and Moran (1996) go further to state that the primary advantage of firms over markets is "purposive adaptation" (pg. 33), which enables firms to coordinate in new directions absent price signals and existing markets, as well as direct the discretionary behavior of its employees.

While economic theories have focused on such alignment through incentives, sociological and management literatures tend to emphasize purpose alignment through socialization processes (Ouchi, 1980; Etzioni, 1965) in which employees assume the values of the organization and adopt shared feelings of solidarity (Durkheim, 1933; Barnard 1968). This same narrative underpins the organizational identity literature's discussion of the processes through which individuals within an organization adopt a common identity (March, 1991; Whetten and Godfrey, 1988). If employees embrace an organization's values, goals, and purpose as part of their identity, then workers are willing to provide consummate performance in pursuit of that purpose (Akerlof and Kranton, 2005).[17]

However, purpose is more than simple goal alignment. It is the "why" behind a social group that imbues these groups with meaning, and meaning, as Frankl (1962) emphasizes, is a primary motivator for

[17] We view organizational purpose as distinct from goals. Purpose speaks to the "reason for which something is done" (Oxford Dictionaries), and typically "reaches beyond profit maximization" (Henderson and Van den Steen, 2015). At the same time, purpose need not be pro-social in orientation, but must simply provide a shared set of beliefs that allow employees to feel that the work they perform has meaning and collective impact (Gartenberg et al). Ultimately, a well-functioning organization is a well-functioning subsociety—one that is imbued with a shared purpose and one that provides a reason for work that inspires employees to look past formal incentives and job designs in calibrating their behavior.

individuals. Consistent with this logic, research suggests that organizations in which employees share a strong sense of purpose outperform their peers (Gartenberg, Prat and Serafeim, 2019).

In summary, two conditions are necessary (although perhaps not sufficient) for a "well-ordered" society: a shared sense of purpose and effective administration of justice. Sub-societies with these two attributes will be likelier to have members with a sense of identification that motivates them in a common direction. Conversely, sub-societies lacking clear purpose and rife with injustice will be likelier to have alienated members who exert perfunctory behavior at best and destructive behaviors at worst.

Responsibilities of central authority in a well-governed subsociety

What then is the role of the central authority in composing a well-governed subsociety? The leader's role is to use central authority to shape perceptions of justice and purpose in ways that prompt individuals to identify with the collective group and willingly take actions in the interests of that collective, even at personal cost. The role of leaders is to foster this collective identity by shaping purpose and justice in ways that enable productive and peaceful cooperation among people. In this sense, leaders occupy one side of the social contract, the foundational mechanism through which individuals surrender key rights and subject themselves to the rule of law (or the employment contract) and the duties of citizenship (or employment), in exchange for the benefits of membership. By their actions, leaders ensure the social contract is either upheld by providing benefits of membership or violated by failing to provide these. In other words, the leaders' responsibilities are to ensure that their authority is used to uphold the society's side of the contract. Depending on their actions, leaders elicit a corresponding response from members. When the leader's end of the social contract is upheld, members identify with the society and act in its interest. If the contract is violated, members will respond in ways that undermine the society's effective function.

Importantly, the preconditions for peaceful cooperation in societies are broader than those for efficient trade in economies. Hence, the responsibilities of principals are broader in societies than economies, and principals must consider the maintenance of purpose and justice when making decisions, as well as efficiency and performance.

Legal basis for firms as subsocieties

In the preceding section, we argued that firms are subsocieties, subject to the same social forces that buffet societies at large. The role of principals, therefore, is broader than the role that would entail were firms merely subeconomies. Along with standard efficiency and performance objectives, principals are also responsible for establishing "well-ordered" societies with clear purpose and shared perceptions of justice. By adeptly attending to or neglecting these societal functions, principals either foster a sense of identification that inspires cooperation, or generate alienation that prompts great organizational dysfunction.

However, our argument so far has been based largely on analogical reasoning—an assertion that firms are like societies, as collectives of individuals aiming to cooperate toward common goals. Therefore, by analogy, they are subject to the same reasoning and logic as societies as a whole with their principals challenged to maintain these subsocieties in a well governed state. In this section, we deepen our argument to propose that firms – by their very legal construction –facilitate an exchange of rights that empower those who lead firms with the capacity to build and shape subsocieties by managing shared purpose and justice.

Central to the theory of the firm is the argument that within their legal boundaries, firms replace the dispute resolution of courts with more flexible dispute resolution enabled by hierarchy (Williamson, 1991). As Freeland and Zuckerman (2018) similarly argue, a firm's legal structure generates a *rights hierarchy*, whereby disputes between members are adjudicated up the hierarchy, with the "firm" as a legal personage at the apex, holding power to wield, delegate and revoke these rights. A central outcome of the rights employees cede to firms is the capacity to flexibly resolve disputes surrounding how work is performed, what behaviors are appropriate, and what distribution of rewards is equitable. Within hierarchies the resolution of these disputes can be "justified through reference to what '*ought' to happen based on the larger social purpose by which such rights are justified*." (Freeland and Zuckerman, 2018: 146, emphasis ours.)

The firm's legal boundary enables this rights exchange and demarcates members of the subsociety from non-members, in much the same way that the formation of a state with borders demarcates citizens from non-citizens and enables an exchange of rights and responsibilities between citizens and the state. It is precisely the rights transferred to the firm that empower the principals to effectively shape purpose and justice. What are these transferred rights, and how do they relate to these social responsibilities? Here, we discuss three rights that, together, provide principals with the power and responsibility to shape these and other societal aspects of an organization.

First, in accepting employment, employees grant employers a capacity to control and shape how work is performed. In fact, the litmus test in the US Tax Code that legally separates an employee from a contractor is "whether the alleged employer …has the *right* to control …the details of the alleged employee's work" (Pitts v. Shell Oil Co., 463 F.2nd 331 (1972), as cited in Masten 1988: 186). Following similar logic, independent contractors represent employers "only as to results of [their] work and not as to the means whereby it is to be done" (56 Corpus Juris Secundum 45; Restatement of Agency (2nd), 2, 14, as cited in Masten 1988: 186). While both contractors and employees can be terminated for failing to accomplish work-related goals, only employees can be terminated for simply engaging in behavior or conduct deemed inconsistent with a manager's beliefs about how work is to be done. When entering an employment relationship, employees accept an implied duty to "yield obedience to all reasonable rules, orders, and instructions of the employer" (53 American Jurisprudence, as cited in Masten 1988: 185). In other words, upon entering an employment contract, individuals consent to be guided not just in their tangible work products, but also in their conduct, loyalty, and focus in the course of creating that work.

Second, in accepting employment, employees cede rights to the employer to determine compensation and its basis, as well as resolve disputes. The courts are extraordinarily reluctant to "pierce the veil" and weigh in on matters internal to the firm, including the fairness of internal rewards. Yet, the courts are quite willing and indeed expected to adjudicate contractual conflicts that play out in markets. Within firm boundaries, managers can disconnect rewards received from current output with impunity

from the courts, allowing the firm to more fluidly manage equity across time, and generating what Ouchi (1980) has referenced as serial rather than spot equity or justice.

Third and finally, employers assume control of voice rights for the firm, which they can selectively delegate. Legally firms are "juridical persons" (Blair, 2004) with legal personalities and identities (Robe 2011; Freeland 2016) that are shaped by those authorized to express the firm's voice. Consequently, these legal personalities and the expressions of particular authorized employees bound to it through an employment contract, speak as though they speak for the firm. And what they speak clearly shapes perceptions of purpose and justice among employees. As Freeland and Zuckerman (2018: 148) note: "because employees are agents who act on behalf of the firm, even very low-level actors often have at least limited power to make binding commitments for the firm or speak on its behalf." Precisely because of this distributed voice, the firm has strong incentives to tightly control expressions of that voice, ensuring its use remains consistent with principles of purpose and justice defined by the owner. Of course, no such voice rights extend beyond the boundaries of the firm, e.g. the firm does not hold voice rights to speak for its suppliers, nor can the firm delegate such voice rights to suppliers.

We view this compilation of rights transferred to the employer as empowering the firm to manage both the consistency of its actions, including its employees, to a set purpose as well as manage consistency in adhering to principles of justice. The firm's right to control employee actions and not merely employee outputs allow the firm to promote behaviors consistent with a stated purpose. The capacity of those who manage the firm to speak for the firm and to more broadly control voice rights provides a vehicle to articulate a firm's purpose and to further reinforce behaviors consistent with it. Moreover, the authority to determine the distribution of rewards largely unfettered by the intervention of the courts provides a broad capacity to manage equity within the firm. Of course, the fact that these rights have been ceded by employees to employers elevates the firm's responsibility to wield these rights in responsible ways that build a well-governed subsociety.

Implication 1: Social amplification within subsocieties

While the authority derived from employment provides an enhanced capacity to reinforce purpose and promote justice in ways that increase employee identification and cooperation, this need not be the outcome. As the saying goes: "with great power comes great responsibility." Effective exercise of these rights supports organizational identification and widespread cooperation. Failure to wield these rights responsibly yields alienation and disengagement. In this manner, the legal boundary of the firm functions as a social amplifier that depending on the choices made by those who direct the subsociety leads to widely varying behavior that is either highly functional or highly dysfunctional.

Importantly, unlike knowledge-based theories of organization, our theory of the firm assigns a central role to the leader who wields this enhanced control to shape an efficient form of governance. Moreover, unlike traditional subeconomy theories, the choice to integrate does not depend solely on how well governance is matched to dyadic transaction or contract attributes. Rather, the integration decision allows an exchange of rights that enhances the capacity to shape purpose and justice and its resulting sociality within the firm and hence generates a lever that simply does not exist outside the firm. Depending on how this lever is then wielded, an array of outcomes is possible.

Owners and managers who lead firms must account for this social amplification when making a wide range of decisions. These decisions include who to employ, how to assign tasks, how to structure compensation, and which customers to target, among others. Employees of the firm will evaluate all such decisions for their consistency with principles of justice and purpose. Those who transact with the firm as non-employee contractors may well recognize decisions that violate justice and purpose, and may even react negatively, especially if these violations impact them personally. However, the responses from employees of the firm—those who have ceded rights of commitment, control, and voice in exchange for an expectation of fairness and justice to prevail within their sub-society, will be more wide-reaching. In particular, employees of the firm—those who are sub-society members will be focused on evaluating the full breadth of decisions involving all other employees and not just themselves, evaluating all such decisions for consistency with purpose and justice. Actions that are deemed unfair or inconsistent with purpose and justice ripple through the society undermining the sub-society's capacity to generate

organizational identification and cooperation. Actions that are deemed fair and consistent with purpose and justice reinforce the sub-society's effectiveness in promoting identification and cooperation. In this sense, the legal boundaries of the firm defined by employment ensure that the social influence of managerial action is amplified within the subsociety and comparatively muted without. Again, this is not to say that non-employees are indifferent to the consistency of decisions with purpose and justice, but differences in the social-legal contract with the firm generate a very different response pattern. We thus make the following general proposition:

> *Proposition 1: Managerial decisions have an amplified social effect on the subsociety as defined by a firm's legal boundaries. Actions that violate principles of purpose and justice undermine organizational identification and cooperation, while actions that reinforce principles of purpose and justice have widespread positive influence that promotes identification and cooperation.*

Implication 2: Boundaries of the firm as subsocieties

While existing theories of the firm focus on boundary choices made so as to economize on the transaction costs associated with a particular exchange, we suggest based on the prior discussion that boundary decisions must also account for the effects of these decisions on shared perceptions of justice and purpose that shape the social well-being of the subsociety. Boundary decisions are therefore a joint function of efficiency effects covered in the standard theories and the social effects arising from our subsociety perspective. To develop this logic, we first review the baseline predictions from standard theories of the firm. We then add our firm as subsociety predictions, and lastly overlay these predictions to develop a set of unified predictions that account for the two perspectives together.

Boundary predictions from subeconomy perspectives

The intuition for the firm-market boundary choice in subeconomy theories is that firms are more prevalent in contexts where transactions are more difficult to execute, often because the transactions demand specific forms of investment. Figure 1, reproduced from Gibbons (2005), illustrates this idea. The y-axis represents the effectiveness of the transaction, while the x-axis represents the difficulty of managing that transaction, where transaction difficulty is a general term that encompasses several theoretical ideas, from asset specificity to effort non-contractibility to optimal incentive design. In general, we observe that

simpler transactions can be more effectively managed regardless of whether within or outside firms. But the central insight is comparative. What we observe is that markets generally outperform firms when transaction execution is easy, but that this advantage gradually erodes and ultimately reverses as difficulty increases. In this sense, the boundary of the firm is shaped by transaction difficulty.

<< Insert Figure 1 here >>

An underlying—although not always stated—assumption in this work is that the difficulty of a transaction is an underlying attribute of the transaction itself: whether the attributes that drive transaction output are ultimately contractible, firm-specific, or have other features that govern difficulty.

Social amplification and transaction effectiveness

Our subsociety view of firms shaped by shared perceptions of purpose and justice suggests a very different mechanism may also influence firm boundary decisions. Boundary choices, in addition to their efficiency effects, are now also evaluated by employees for consistency with the espoused purpose and principles of justice. Expanding the scope of the firm may involve transacting with new customers or engaging in activities that some view as violating the firm's purpose or integrating a business with policies perceived as unfair relative to the existing business. In this sense, boundary decisions have social amplification effects that must be accounted for by those responsible for choosing firm scope.

Our motivating illustrations were of this precise form. For instance, law firms expand into pro bono work, rather than either requiring their personnel to focus on higher margin corporate work or allowing their personnel to take time off to devote to pro bono work on their own. The choice to provide pro bono services in-house was made not because these activities were especially difficult, firm-specific, non-contractible or any of the other standard explanations of internalization decisions. Instead, offering pro bono work imbued the firms as a whole with purpose, particularly after a devastating tragedy. Google exited businesses in both China and the US not because these transactions suddenly became simple to outsource. Instead, they exited because employees revolted for being associated with the work. Energy firms exited speculative energy trading not because integration incurred excessive transaction costs, but

rather because—at least in part—energy traders received such outsized compensation relative to the rest of the organization. These are instances of social amplification through boundary decisions. In these illustrations, the effectiveness of the scope choices by law firms after 9/11, Google, and energy firms was influenced by the social response of the firms' members, positive in the case of the law firms and negative in the case of Google and energy firms.

Whether this response is positive or negative is determined by the consistency of the transactions with the espoused purpose and principles of justice of the subsociety. Actions that are socially consistent – such as allowing lawyers to provide pro bono services within their communities – strengthen the sense of purpose within the organization. Actions that are inconsistent – such as hiring John Lasseter despite the concerns of the female employees – weaken the sense of justice within the organization.

Figure 2 illustrates this idea. As with Figure 1, Figure 2 compares the relative effectiveness of markets and firms, but this time as a function of the social consistency of the transaction with the underlying firm. The x-axis represents this social consistency of the transaction with the firm. As demonstrated by the illustrations above, consistency can be positive and reinforcing (e.g., law firms after September 11) or negative and violating (e.g., Google), to varying degrees. The center of the x-axis reflects the indifference point, where the transaction is neither consistent nor inconsistent with the underlying purpose and justice of the firm. As in Figure 1, the y-axis refers to transaction effectiveness.

The higher slope in Figure 2 of the curve representing the firm relative to the market graphically depicts the social amplification associated with firm boundaries. Starting at the indifference point in the center of the x-axis and moving right, the figure shows how transaction effectiveness within firms increases relative to markets the more they reinforce the purpose and justice within the firm. Conversely, as we move to the left and transactions increasingly violate the espoused purpose and justice of the firm, their effectiveness within firms decreases relative to markets. Expanding the boundaries of the firm into transactions that reinforce justice and purpose, therefore, are associated with higher organizational effectiveness, while those that violate purpose and justice are associated with lower effectiveness.

<< Insert Figure 2 here >>

This leads us to our second proposition:

> *Proposition 2: All else equal, firms are likely to internalize transactions that reinforce principles of purpose and justice, and less likely to internalize transactions that violate those principles.*

Social amplification, transaction difficulty, and firm boundaries

We now consider an integrated view of the subeconomy and subsociety theories of firms, as depicted in Figure 3. Along the vertical axis is transaction difficulty, the key determinant of boundaries under the subeconomy view. Along the horizontal axis is transaction social consistency, the key determinant of boundaries under the subsociety view. The contents of each quadrant represent predictions of governance mode as a function of the degree of each lever.

The diagonal quadrants contain the straightforward predictions: If the transaction is both difficult and socially consistent with the firm, it is likely to be internalized (upper right quadrant). Similarly, if the transaction is both easy and is socially inconsistent with the firm, it is likely to remain external to the firm.

The off-diagonal quadrants depict divergent subeconomy and subsociety effects. When the transaction is both difficult and socially inconsistent (upper left quadrant), managers must weigh the efficiency effects of internalizing the transaction against its social costs. The Google AI illustration above, for example, falls into this quadrant and represents the managerial challenge of weighing potentially attractive scope expansion against the potentially harmful social impact on employees. The flip side of this situation arises when the transaction itself is simple but reinforces societal principles of the firm (lower right quadrant). The post 9/11 legal firms' expansion into pro bono work is an example of this condition, whereby the pro bono work itself did not represent a transaction of particular difficulty, but reinforced the aspired purpose of these firms. This latter condition presents management with the need to weigh the potential inefficiency of internalizing these transactions with the potential social benefits to the members of the firm.

<< Insert Figure 3 here >>

This discussion above then presents the following additional propositions:

Proposition 3: When transactions are relatively more efficient under market governance, firms will internalize them nonetheless if the benefits of reinforcing principles of purpose and justice outweigh the added transaction difficulty of integration over outsourcing.

Proposition 4: When transactions are relatively more efficient under firm governance, firms will outsource them nonetheless if the costs of violating principles of purpose and justice outweigh the reduced transaction difficulty of integration over outsourcing.

Discussion and conclusion

Our aim in this paper has been to propose that firms are not mere subeconomies—aggregations of internally managed and sequentially assembled transactions, but also subsocieties, subject to the same social forces as the larger societies in which they are embedded. These subsocieties are demarcated by formal legal boundaries that empower those who own and manage them to shape perceptions of shared purpose and justice—principles that are vital to promoting (or undermining) shared identity and inspiring cooperation. The actions of those who manage these subsocieties are interpreted for their consistency with shared purpose and principles of justice, with consistency encouraging identification and inconsistency prompting alienation.

While there are a range of actions that shape perceptions of purpose and justice, we particularly highlighted decisions about boundary choices—decisions to add activities to or subtract activities from the firm. With this focus, we explored how broader sub-society spillovers create a social amplification effect through which boundary decisions are evaluated. Given this social amplification, activities that might be more efficient inside the firm under standard theories of the firm may in fact be externalized if their integration would incur sufficiently high social costs. Conversely, integration decisions that appear to be less efficient than transacting via market exchange may remain attractive to integrate, if integration enables social benefits, as in Carnahan *et* al (2017)'s example of pro bono work within law firms following the 9/11 terror attacks.

Our argument highlights the central role of formal, legal boundaries in empowering actors to shape subsocieties. In doing so, we seek to compose a theory of the firm distinct from the tradition of purely social theories of organization, going back to Barnard, Selznick, March and Simon, and other pioneers of organization theory. One branch of these social theories of organization emphasizes the

organizing role of social networks (or dyadic relationships) and makes strong claims that the sociality that operates within or across networks generates trust, information exchange, and cooperation that essentially substitutes for formal organization. In emphasizing this sociality as a substitute for formal organization, the theories do not address the central enabling role that firm boundaries and associated legal rights within the firm play in empowering those who architect these social relations. In contrast, our theory emphasizes the important power that legal boundaries bestow to firm principals to explicitly shape sociality within the firm, and the corresponding responsibility that accompanies that power.

A second branch focuses on the role of organizational identification in driving differences in behavior within and across firms. In this research, organizational membership shapes both perceptions of external conditions, as well as perceptions of "who we are" and "what we do." However, this literature offers little explanation as to why shared identity co-aligns with firm boundaries beyond recognizing that social interaction is more prevalent within firm boundaries (Porac et al., 1995). This framing implies that organizational identity is a mere correlate of socialization patterns, rather than an organizational property that results directly from the legal entity. Unlike this prior work, our aim is to demonstrate how formal authority, as distinctly enabled by the legal entity, shapes the subsociety with the firm and thereby influences the composition (or destruction) of organizational identity.

A third branch focuses on knowledge and capabilities-based theories of the firm, wherein firms are more efficient means to use and produce dispersed, tacit knowledge. Although this branch is closer in spirit to a cognitive theory of the firm, it does incorporate several motivational aspects. Nelson and Winter (1982: 107-112) conceived of routines as organizational "truces…defined by a shared tradition arising out of the specific contingencies confronted and the responses of the parties to those contingencies," which parallels a relational contracting view of the firm as discussed by Gibbons and Henderson (2013). Kogut and Zander (1996) appealed to identity arguments to explain why members of organizations would be more motivated to coordinate using tacit knowledge than individuals outside of an organization context: "boundaries demarcate qualitative changes in the reservoir of social knowledge…because coordination and learning are developed within the organizational context of shared identities." (pg 503). We view this

branch as complementary to our concept of firms: organizations may indeed be superior at developing knowledge and capabilities, given the establishment of organizational truces and shared identities. However, these truces and shared identities both are affected by the strength of purpose and justice within the organization.

Our theory most closely relates to several recent social approaches to the theory of the firm. Nickerson and Zenger (2008) have focused on the effects of envy and social comparison on firm scope. In their view, firm boundaries demarcate social reference groups within which individuals compare their own standing and performance along salient dimensions. These social reference groups can induce comparison costs within firms, particularly when the comparisons are perceived as inequitable. Our work relates to theirs in the sense that social comparison is an important dimension along which justice is assessed by individuals. Our framing of firms as subsocieties is broader and encompasses more than strictly social comparison. Our focus is on broader principles of justice—as well as purpose—that principals must account for when making boundary choices. Our effort is also related to recent work by Freeland and Zuckerman (Freeland 2016; Freeland and Zuckerman, 2018). While we have highlighted relevant connections in developing our theory, it is useful to highlight both similarities and differences with this work. Both approaches place the role of legal boundaries and allocation of rights within the firm hierarchy as foundational in differentiating firms and markets. The approaches differ, however, in the implications drawn. Whereas Freeland and Zuckerman focus on voice rights and how these are affected by boundaries, we focus on societal expectations and obligations that result from the legal structure and boundaries.

One critique of our work is that our arguments around purpose and justice simply reflect additional factors that shapes transaction difficulty, along the lines of other transaction specific factors such as asset specificity, adaptability, and contractability (Gibbons, 2005). On the one hand, we acknowledge the possibility of this simple additive approach. A transaction's consistency with purpose and justice could be just another transaction difficulty to factor into some joint optimization logic, along the lines of Figure 2. On the other hand, we see value in isolating these effects in order to understand how

they operate, in much the same way that alternative theories of the firm have been composed focused on asset specificity, adaptation, and contractability. Moreover, we view our theory of the firm as addressing more than simple boundary decisions. It is unique in highlighting the empowering role that firm boundaries play in allowing managers to broadly shape an organization's social function.

We recognize that our focus on purpose and justice as the principal's primary levers to shape sociality – and the primary determinants of the social consistency of a transaction – may miss other important social levers and determinants of a well-functioning subsociety. We make no claim that purpose and justice represent an exhaustive list. However, in our collective experience in both teaching and consulting to companies, we have found that these two determinants account for a large majority of the subsociety issues we have observed within firms. Moreover, as discussed in the paper, these two social constructs appear to be the most foundational construct highlighted in political economy as determinative of well governed societies, as well as pervasive across a wide range of literature in social psychology, organization theory, and sociology.

The overriding aim of this paper is to demonstrate how formal authority, as distinctly enabled by the legal entity, shapes the subsociety within the firm and, in doing so, induces social responses of its members to managerial decisions. In this sense, our proposed theory parallels the work in the economic "Theory of the Firm" tradition, but highlights social rather than economic mechanisms in determining the costs of decisions. It is our view that this characterization of firms as subsocieties – with the attendant impact on decisions within and about firm boundaries – can explain important empirical regularities not well characterized by existing theories of the firm. By doing so, we hope to inspire future work on the societal factors that influence firm behavior both within and across formal boundaries.

References

Adams, J.S., 1965. Inequity in social exchange. In *Advances in experimental social psychology* (Vol. 2, pp. 267-299). Academic Press.

Akerlof, G.A. and Kranton, R.E., 2005. Identity and the Economics of Organizations. *Journal of Economic perspectives*, *19*(1), pp.9-32.

Alchian, A.A. and Demsetz, H., 1972. Production, information costs, and economic organization. *The American economic review*, *62*(5), pp.777-795.

Amit, R. and Schoemaker, P.J., 1993. Strategic assets and organizational rent. *Strategic Management Journal*, 14(1), pp.33-46.

Barnard, C.I., 1968. *The functions of the executive* (Vol. 11). Harvard university press.

Barnard, C.I., 1968. *The functions of the executive* (Vol. 11). Harvard university press.

Blair, M.M., 2004. The neglected benefits of the corporate form: Entity status and the separation of asset ownership from control. *Corporate Governance and Firm Organization: Microfoundations and Structural Forms, Anna Grandon, ed*, pp.45-66.

Cappelli, P., 2020. "Contracting, Employment, and the Duty of Care: A Research Agenda", Working paper.

Carnahan, S., Agarwal, R. and Campbell, B.A., 2012. Heterogeneity in turnover: The effect of relative compensation dispersion of firms on the mobility and entrepreneurship of extreme performers. Strategic Management Journal, 33(12), pp.1411-1430.

Carnahan, S., Kryscynski, D. and Olson, D., 2017. When does corporate social responsibility reduce employee turnover? Evidence from attorneys before and after 9/11. *Academy of Management Journal*, *60*(5), pp.1932-1962.

Cheung, S.N., 1983. The contractual nature of the firm. *The Journal of Law and Economics*, *26*(1), pp.1-21.

Coase RH. 1937. The nature of the firm. Economica 4: 386–405.

Conner, K.R. and Prahalad, C.K., 1996. A resource-based theory of the firm: Knowledge versus opportunism. *Organization science*, *7*(5), pp.477-501.

Demsetz, H., 1988. The Theory of the Firm Revisisted. *JL Econ. & Org.*, *4*, p.141.

Durkheim, E., 1933. The division of labor. *Trans. G. Simpson. New York: Macmillan*.

Elsbach, K.D. and Kramer, R.M., 1996. Members' responses to organizational identity threats: Encountering and countering the Business Week rankings. *Administrative science quarterly*, pp.442-476.

Etzioni, A., 1965. Dual leadership in complex organizations. *American Sociological Review*, pp.688-698.

Festinger, L., 1954. A theory of social comparison processes. *Human relations*, *7*(2), pp.117-140.

Freeland, R.F., 1996. The myth of the M-form? Governance, consent, and organizational change. American Journal of Sociology, 102(2), pp.483-526

Freeland, R.F. and Zuckerman Sivan, E.W., 2018. The problems and promise of hierarchy: voice rights and the firm. Sociological Science, 5, pp.143-181.

Gartenberg, C. and Wulf, J., 2017. Pay harmony? Social comparison and performance compensation in multibusiness firms. *Organization Science*, *28*(1), pp.39-55.

Gartenberg, C., Prat, A. and Serafeim, G., 2019. Corporate purpose and financial performance. Organization Science, 30(1), pp.1-18.

Gibbons, R. and Henderson, R., 2012a. What Do Managers Do? Exploring Persistent Performance Differences among Seemingly Similar Enterprises.

--------------------------, 2012b. Relational contracts and organizational capabilities. Organization science, 23(5), pp.1350-1364.

Granovetter, M., 1985. Economic action and social structure: The problem of embeddedness. *American journal of sociology*, *91*(3)

Gouldner, A.W., 1961. Anti-Minotaur: The myth of a value-free sociology. Soc. Probs., 9, p.199.

Grant, R.M., 1996. Toward a knowledge-based theory of the firm. *Strategic management journal*, *17*(S2), pp.109-122.

Grossman, S.J. and Hart, O.D., 1986. The costs and benefits of ownership: A theory of vertical and lateral integration. *Journal of Political Economy*, 94(4), pp.691-719.

Hart, O. and Moore, J., 1990. Property Rights and the Nature of the Firm. *Journal of Political Economy*, 98(6), pp.1119-1158.

Helfat, C.E., Finkelstein, S., Mitchell, W., Peteraf, M., Singh, H., Teece, D. and Winter, S.G., 2009. Dynamic capabilities: Understanding strategic change in organizations. John Wiley & Sons.

Henderson, R. and Van den Steen, E., 2015. Why do firms have" purpose"? The firm's role as a carrier of identity and reputation. *American Economic Review*, *105*(5), pp.326-30.

Hirschman, A.O., 1970. Exit, voice, and loyalty: Responses to decline in firms, organizations, and states (Vol. 25). Harvard university press.

Hobbes, T., 1980. Leviathan (1651). *Glasgow 1974*.

Holmstrom, B., 1999. The firm as a subeconomy. *Journal of Law, Economics, and organization*, *15*(1), pp.74-102.

Homans, G.C. and Behavior, S., 1961. Its elementary forms. Social Behavior, 119(3), pp.488-531.

Husted, B.W. and Folger, R., 2004. Fairness and transaction costs: The contribution of organizational justice theory to an integrative model of economic organization. *Organization Science*, *15*(6), pp.719-729.

Jacobides, M.G. and Winter, S.G., 2005. The co-evolution of capabilities and transaction costs: Explaining the institutional structure of production. Strategic management journal, 26(5), pp.395-413.

Jensen, M.C. and Meckling, W.H., 1976. Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of financial economics*, *3*(4), pp.305-360.

Kacperczyk, A. and Balachandran, C., 2018. Vertical and horizontal wage dispersion and mobility outcomes: evidence from the Swedish microdata. *Organization Science*, 29(1), pp.17-38.

Katz, E., & Lazarsfeld, P. F. (1955). Personal influence: The part played by people in the flow of mass communication. Glencoe, IL: Free Press.

King, B.G., Felin, T. and Whetten, D.A., 2010. Perspective—Finding the organization in organizational theory: A meta-theory of the organization as a social actor. *Organization Science*, *21*(1), pp.290-305.

Klein, B., 1996. Why hold-ups occur: the self-enforcing range of contractual relationships. Economic inquiry, 34(3), pp.444-463.

Klein, B. and Murphy, K.M., 1988. Vertical restraints as contract enforcement mechanisms. *The Journal of Law and Economics*, *31*(2), pp.265-297.

Kogut, B., 2000. The network as knowledge: Generative rules and the emergence of structure. *Strategic management journal*, *21*(3), pp.405-425.

Kogut, B. and Zander, U., 1992. Knowledge of the firm, combinative capabilities, and the replication of technology. *Organization science*, *3*(3), pp.383-397.

Kogut, B. and Zander, U., 1996. What firms do? Coordination, identity, and learning. *Organization Science*, 7(5), pp.502-518.

Kulik, C.T. and Ambrose, M.L., 1992. Personal and situational determinants of referent choice. *Academy of Management review*, *17*(2), pp.212-237.

Malmendier, U., 2009. Law and Finance" at the Origin". *Journal of Economic Literature*, *47*(4), pp.1076-1108.

Simon, James G. March Herbert A. Organizations. 1958.

Martin, J., 1981. Relative deprivation: A theory of distributive injustice for an era of shrinking resources. Graduate School of Business, Stanford University.

Masten, S.E., 1988. A legal basis for the firm. Journal of Law, Economics, & Organization, 4(1), pp.181-198.

Mayo, E., 1945. *The Social Problems of An Industrial Civilization, Including, as an Appendix: the Political Problem of Industrial Civilization*. Division of Research, Graduate School of Business Administration, Harvard University.

Nelson, R.R. and Winter, S.G., 1982. The Schumpeterian tradeoff revisited. The American Economic Review, 72(1), pp.114-132.

Nickerson, J.A. and Zenger, T.R., 2004. A knowledge-based theory of the firm—The problem-solving perspective. *Organization science*, *15*(6), pp.617-632.

Nickerson, J.A. and Zenger, T.R., 2008. Envy, comparison costs, and the economic theory of the firm. *Strategic Management Journal*, *29*(13), pp.1429-1449.

Obloj, T. and Zenger, T., 2017. Organization design, proximity, and productivity responses to upward social comparison. *Organization Science*, *28*(1), pp.1-18.

Ouchi, W.G., 1980. Markets, bureaucracies, and clans. *Administrative science quarterly*, pp.129-141.

------------------ 1984. The M-form society. *Human Resource Management (pre-1986)*, *23*(2), p.191.

Penrose, E. and Penrose, E.T., 2009. The Theory of the Growth of the Firm. Oxford university press.

Perrow, C., 1986. Economic theories of organization. *Theory and society*, *15*(1-2), pp.11-45.

Rawls, John. *A theory of justice*. Harvard university press,1971.

Robé, Jean-Phillipe. 2011. "The Legal Structure of the Firm." Accounting, Economics, and Law 1:1–86.

Robertson, D.H., 1923. *The control of industry* (No. 4). Harcourt, Brace.

Santos, F.M. and Eisenhardt, K.M., 2005. Organizational boundaries and theories of organization. *Organization science*, *16*(5), pp.491-508.

Selznick, P., 1948. Foundations of the theory of organization. *American Sociological Review*, *13*(1), pp.25-35.

Selznick, P, 1957. Leadership in administration, New York: Harper and Row.

Simon, H.A., 2013. Administrative behavior. Simon and Schuster.

Simon, H.A., 1951. A formal theory of the employment relationship. Econometrica: Journal of the Econometric Society, pp.293-305.

Simon, H.A., 1991. Bounded rationality and organizational learning. Organization science, 2(1), pp.125-134.

Smith, A., 1758. The Theory of Moral Sentiments

Smith, A., 1776. An Inquiry into the Nature and Causes of the Wealth of Nations,

Teece, D.J., Pisano, G. and Shuen, A., 1997. Dynamic capabilities and strategic management. Strategic management journal, 18(7), pp.509-533.

Weick, K.E., 1995. Sensemaking in organizations (Vol. 3). Sage.

Williamson, O.E., 1971. The vertical integration of production: market failure considerations. *The American Economic Review*, *61*(2), pp.112-123.

------------------ 1973. Markets and hierarchies: some elementary considerations. *The American economic review*, *63*(2), pp.316-325.

------------------, 1975. Markets and hierarchies. *New York*, *2630*.

------------------ 1991. Comparative economic organization: The analysis of discrete structural alternatives. *Administrative science quarterly*, pp.269-296.

Figure 1: Economic theory of the firm (Gibbons 2005)

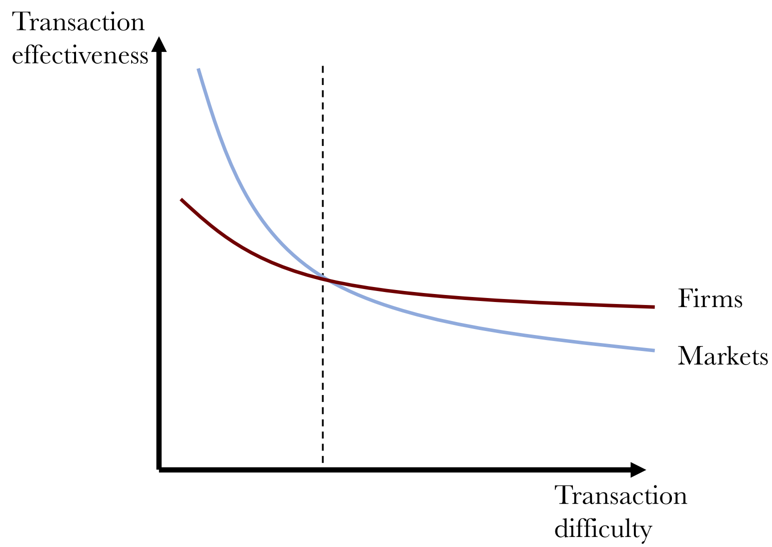

Figure 2: Firms versus markets as function of social consistency

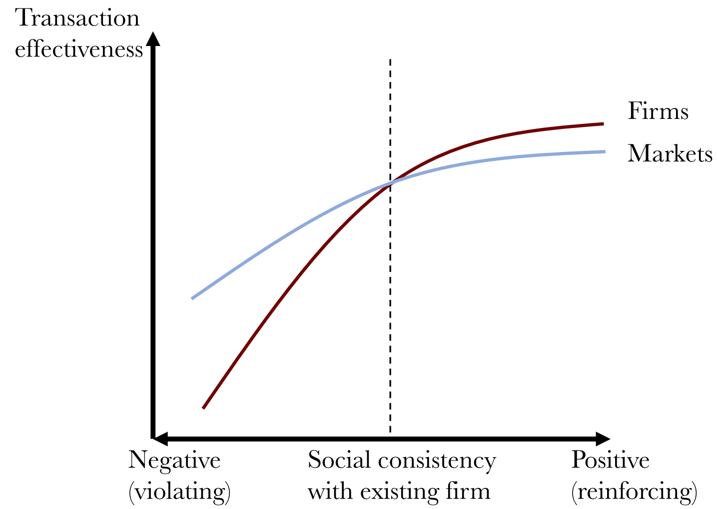

Figure 3: Firm boundaries as function of economic and social logic

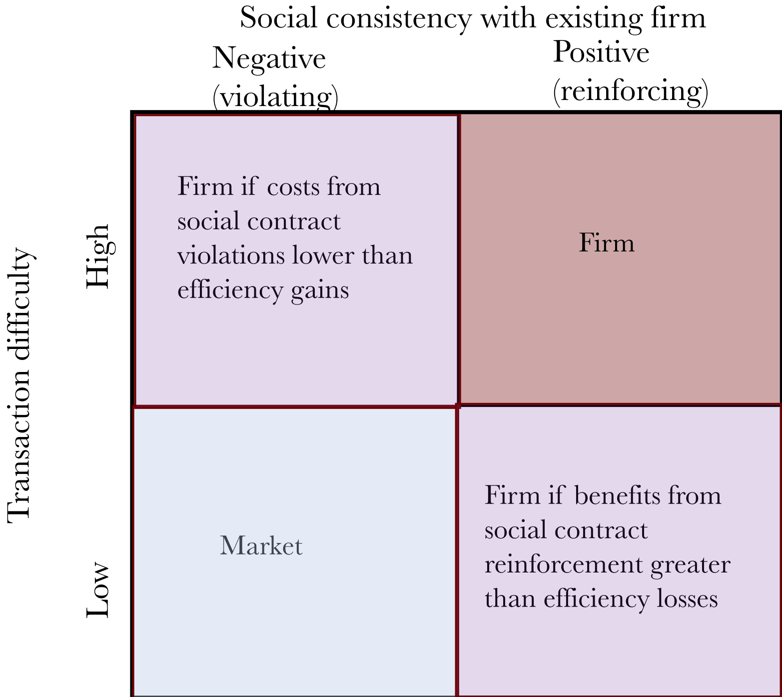

Table 1: Summary of company actions and employee response in illustrations

Illustration	Company	Corporate action	Employee response	Category of response	Company response (if reported)
1	Google	AI partnership with US Dept Defense	Employee protests over violations of company purpose	Purpose	Cancellation of contract, publication of AI principles
2	Energy companies	Establishment of internal trading operations	Employee protests within core business about inequitable compensation policy	Justice	Shuttering internal operations and outsourcing to financial firms
3	Pharmaceutical companies	Acquisitions of biotech firms	Protests from scientist of parent business about inequitable compensation policy	Justice	Various, including out-licensing, pay harmonization, scientist turnover
4	Law firms	Investment in pro-bono services	Increased attorney job satisfaction and lower turnover	Purpose	
5	Skydance Productions	Hiring John Lasseter as head of new animation division	Employee protests over hiring decision	Justice	Maintained hiring decision
6	Amazon	Facial recognition partnership with police departments	Employee protests over violations of company purpose	Purpose	Suspension of partnership
7	Disney	Launch of Disney+ streaming service	Reported internal challenges with definition of metrics, compensation across divisions, and cannibalization of revenue by new service	Justice	

Appendix A

Illustration 6

In an October 2018 letter to CEO and founder Jeff Bezos, Amazon employees asked the CEO to discontinue the company's sale of facial recognition technology to police departments. The employees wrote, "Technology like ours is playing an increasingly critical role across many sectors of society. What is clear to us is that our development and sales practices have yet to acknowledge the obligation that comes with this. Focusing solely on shareholder value is a race to the bottom and one that we will not participate in." In June of 2020, Amazon stopped its sale of the technology to all government entities including police.

During this same time period, Microsoft, IBM, and Google all faced similar pressures involving their own related technologies. Google was pressured to walk back its efforts to provide surveillance technology to the Chinese Government. IBM announced an intent to discontinue entirely its work on facial recognition technology. Microsoft faced pressure from employees to discontinue their AI work with the US Immigration and Customs Enforcement Agency and the DOD, but has so far resisted efforts to discontinue this work. (June 10, Electronic Foundation Frontier)

Illustration 7

In 2017, Disney announced that it would launch Disney+, an online streaming service to provide content directly to consumers, competing with other streaming providers such as Netflix, Hulu, and Amazon Prime Video. Two years after the announcement, the implementation of this shift became, as one industry observed, "[CEO Bob] Iger's biggest challenge." One specific issue was the compensation structure for the new division. Since streaming services were not expected to be profitable for several years, Disney could not use the standard compensation model focused on profitability and ratings to reward and motivate the leadership. Instead, Disney considered linking compensation to subscriber growth. But of course, subscriber growth would likely come at the expense of the traditional networks. Because of the combination of lack of comparability and cannibalization of the legacy businesses, senior leaders at Disney were concerned they could not implement or, more specifically, create incentives for the

new service in a way that would be considered fair by both the new and existing divisions. Ultimately, Disney executives pushed ahead, viewing these incentive dilemmas as a cost of their entry into streaming. At the same time, they were aware that employee rebellions over perceived unfairness had deeply impaired other large media companies during similar transitions, including Time Warner during its merger with both AOL as well as more recently with AT&T. Comcast, another peer, also announced that it would forego entry into streaming altogether, in part for implementation challenges of this type.