The Economics of Intangible Capital

Nicolas Crouzet, Janice C. Eberly, Andrea L. Eisfeldt, and Dimitris Papanikolaou

Intangible capital is generally defined by what it lacks—that is, as productive capital that lacks a physical presence. Familiar and important examples include patents, software and databases, trademarks, customer lists, franchise agreements, and organization capital and firm-specific human capital.

In contrast, we will focus on the properties that affirmatively characterize intangibles. Fundamentally, since intangibles lack a physical presence, they require a storage medium. The medium can be a piece of physical capital, like a computer (for software), or a document (for a patent or a design), or a person (for a method or an innovation). This need for a storage medium has important implications for the two properties that we emphasize throughout the paper. First, intangibles can be used simultaneously in production in different locations and processes, which implies some degree of non-rivalry in use. Because the same intangible can be simultaneously stored (copied) in multiple places and used simultaneously in production, intangibles allow for economies of scale and scope. However, it can also be difficult to establish and enforce exclusive property rights to an intangible; unlike a physical piece of capital, an intangible can be readily copied or imitated, simply by copying...
software or by learning information, for example. We describe this property as \textit{limited excludability}. Limited excludability makes it difficult to establish an intangible as an asset with enforceable property rights.

We begin by discussing these properties and their implications. We point out that the extent to which these properties generate a valuable intangible asset—which motivates investment—depends on the properties of the storage technology, and the resulting non-rivalry and excludability, and the institutions that enforce property rights.

\textbf{Figure 1} plots the relative value of internally generated intangible assets to tangible capital. Two things are of note in the figure. First, intangible assets grew much faster than tangible assets throughout the 1990s and early 2000s. Second, the faster growth in intangibles appears to have ceased around 2005 (at least for public firms). Regardless of the current growth rate, it is clear that intangibles represent a very large fraction of corporate capital.

Existing research has largely defined intangibles as a variant within the traditional physical capital framework: for example, the investment process for intangible capital is more uncertain; intangibles depreciate slower (or faster, or more randomly) than physical capital; intangibles have a different relative price; and so on. Intangibles then often amount to “missing” or “mismeasured” capital. Measuring intangible capital is difficult and does tend to exclude significant components (for instance, intangibles stored in employees). Starting from a more affirmative description of intangibles bypasses these boundaries and leads to novel implications for the theory and measurement of intangibles. We discuss a model for capturing the economic implications of these properties and consider its implications. We show how this approach can shed light on some important recent macroeconomic and financial trends, including declining measured productivity growth, growing inequality, rising market power, rising valuations, and declining tangible investment rates.

\section*{Characteristics of Intangibles as Assets}

\subsection*{Two Fundamental Properties}

Intangibles are capital, like machinery or structures, in the sense that creating them requires foregoing consumption today (investment) to achieve more output in the future. Unlike machinery or structures, however, intangibles lack a physical presence. At an abstract level, intangibles consist of information. As information without a physical form, intangibles must be stored in some medium in order to be used in production. The need to store intangibles creates their two fundamental properties, which we will call non-rivalry in use and limited excludability.

Storage of intangible assets may be done using different media: speech, writing, drawing, software, recordings, or other technologies. For instance, an algorithm is stored using code or software; a logo is stored using drawings; a managerial process is stored in a team of workers or in written instructions specifying the operational rules of a firm. Although intangibles are sometimes conflated with their storage
medium, the two are distinct. The value of a book is not the paper and cover, but rather the information it holds. Algorithms and data can be written down on paper or encoded in software or databases, but the value derives from the information, not the medium that encodes it.

Given a storage medium, the non-rivalry in use property arises because the same intangible can be stored simultaneously in multiple places. By duplicating the storage medium, the same intangible can be used as an input in production across multiple goods or services at the same time. For instance, the same algorithm can be copied (stored in multiple places) and used in multiple simultaneous instances to produce, say, search results. The same design for a logo can be drawn and then copied and used in multiple simultaneous instances to brand clothing products. Managerial processes can be used in multiple simultaneous instances in different parts of the same organization, or across firms around the world. We specify “non-rivalry in use,” as opposed to simply “non-rivalry,” to stress that intangibles are production inputs. By contrast, the public economics literature commonly uses the expression “non-rival” to refer to consumption goods, rather than to production inputs. The degree of non-rivalry in use depends on the technology underlying the storage medium: an algorithm stored in teams of workers, for example, may be less efficient to use across locations than one that is stored and deployed in software.

Similarly, even if an intangible is stored in a particular medium, it can be difficult to claim and enforce property rights to the surplus it might create. We refer to
this second property of intangibles as limited excludability. An extreme example of
intangibles with limited excludability are public goods that can be used as capital
inputs, such as an open-source operating system, or a method for making fire. These
inputs are non-rival in use within the firm, and also it is not possible to exclude other
firms from using them. Other examples of intangible capital—such as patents—
offer more property rights protection to their owners. In what follows, we will use
“intangible assets” to describe intangible capital inputs whose value can be captured
and privately appropriated. The extent to which this is possible depends both on the
 technological features of the storage medium and on the institutional environment.

Technology determines how intangibles are stored. Prior to the development of
writing (including images), intangible knowledge was passed down from one person to
another through speech. Writing allowed for intangibles to be stored independently of
individuals. As technology has progressed, the scope of which intangibles can be stored
has expanded. For instance, digital media can store larger amounts of information
than writing, allowing for storage of complex intangibles such as genomic sequences
or consumer databases. Recording a lecture is a more comprehensive form of storage
than distributing the notes for that lecture. These technological advances in storage
technologies may help to explain the rapid rise in measured intangibles since the 1990s.

Institutions, both informal and formal, create extrinsic value from stored intan-
gible assets by enforcing excludability, which limits the set of agents with the right
to use the asset and capture its value. For instance, limits on the disclosure of ideas,
such as trade secrets, create excludability. Excludability is often formalized and
enforced through the legal system, including the patent system, copyright enforce-
ment, and non-compete clauses.

There is feedback between technology and institutions. Institutions affect the
incentives to store intangibles in different media. For example, the fact that software
can be copied will undermine its value, unless intellectual property rights are enforced
by institutions. Conversely, the degree of legal protection depends on the storage
technology. Reliable storage makes it easier to identify and enforce legal protections.
Moreover, as the technology to store intangibles evolves, it can displace the value of
intangibles stored in now obsolete technologies: for example, software has replaced
many of the human resource functions previously done by trained labor. Intangibles
that are harder to codify, such as higher-level management practices, can be harder
to imitate—except by hiring away key labor inputs. Differences in displacement risk
for intangibles stored in labor inputs may have played a role in trends in income and
wealth inequality.

In Table 1, we list some common examples of intangible capital. All have an
element of non-rivalry in the sense that underlying information or instructions or
contract provisions that make up the intangible capital can be used repeatedly in
different times and places, though sometimes with imperfect resolution or reproduc-
ibility. In addition, these types of intangible capital vary in how they are stored and
how property rights are generated. For example, property-rights enforcement may be
centralized (say, via the US Patent and Trademark Office), or bilateral, using contract
law.
Parallels with Physical Capital

Intangible capital shares several properties with physical capital: 1) it is an accumulated factor; 2) it depreciates; and 3) it can be firm-specific to varying degrees. We briefly discuss these parallels.

Like physical capital, intangibles require investment, which is commonly observed in functions like research and development, marketing, or human capital and skill accumulation. The mapping from investment to accumulated capital may be less certain for intangibles than for physical capital—which is one reason that conventional accounting has historically not capitalized research and development in the same way as physical capital expenditures. Of course, the mapping for physical capital may be less certain than is typically assumed. The measured physical capital stock is not a census of machines—it is an accumulation of investment. Historical expenditure on fiber optic cables, for example, did not accumulate in a simple way to the current value nor productive use of the current stock of fiber optic cable.

For physical capital, depreciation can be caused by wear and tear and by obsolescence. Intangibles do not suffer wear and tear, only obsolescence. Obsolescence of intangible capital can result from several causes: lack of continued investment (maintenance) in intangibles; the arrival of new/or better vintages of intangibles; or other reasons exogenous to the firm. For instance, brand value may be forgotten if marketing expenses are not kept up; management and production processes may become obsolete as new methods appear; knowledge can be lost when employees depart; and data that is not up-to-date becomes less useful. For intangible capital, reversing or slowing the extent of depreciation due to obsolescence requires investment that involves innovation and whose outcome may be more uncertain than replacement of physical capital. In addition, similar to the way in which physical capital may be destroyed as a result of a natural disaster, intangibles can be destroyed by other disasters: corporate scandals; violation of intellectual property
laws by private actors or expropriation by governments; employees with key
skills leaving the organization; changes in laws; or shifts in consumer tastes (for
example, when a sports team changes its brand name in response to shifting
cultural norms). The forces driving depreciation of intangibles can lead to large
and abrupt negative shocks in the form of rare disasters to the accumulation of
intangible capital.

Finally, the two key properties of intangibles highlighted above—non-rivalry
in use, and limited excludability—can also be thought of in the context of phys-
cical capital. Physical assets are, by definition, rival in use: a particular truck cannot
produce transportation services across different routes at the same time; the same
mill cannot produce steel pipes in different locations at the same time. Additionally,
property and control rights are generally easier to assign to physical assets. Trucks
must be titled, and the title identifies the owner. Ownership of the steel mill, while
it might be shared, is formalized through contracts, and disputes regarding control
generally have legal remedies.

Production with Intangible Capital

In this section, we describe a model that links the ideas discussed in the previous
section—storability, non-rivalry, and limited excludability—and derive their impli-
cations for production and investment. A full algebraic presentation of the model is
available in the online Appendix with this article at the JEP website.

A Model of Production with Intangibles

We focus on a single firm which operates for a single period and makes oper-
ating and investment choices to maximize the terminal value of profits. Although we
use the word “entrepreneur” to describe the owner of the firm (and the intangible
asset), the term is meant more broadly to encapsulate all parties that participate in
the creation, dissemination, and use of intangible assets. Examples include all skilled
personnel who are responsible for the creation of new inventions or business ideas,
entrepreneurs, managers, and startup employees. The model includes two types of
capital, physical and intangible. Both types of capital can be deployed across multiple
production streams, which can be thought of as different product lines, physical loca-
tions, or market segments. The number of production streams determines the span of
the firm. To highlight the role of intangibles, we minimize the role of physical capital:
in this model, it can be rented at a constant user cost. By contrast, intangibles in the
model need to be stored but may be non-rival in use. Moreover, limited excludability
will limit the scope of deployment of intangible capital.

For ease of exposition, we split the model into three stages shown in Figure 2:
decision about the level of intangible investment, choice of the span of the firm,
and the allocation of intangibles and physical capital across production streams.
In building intuition about the model, it is convenient to discuss these choices in
reverse order.
In the production stage, the firm’s intangible capital and its span of production are taken as given. Thus, the firm chooses the amount of physical capital and the allocation of physical and intangible capital to each stream of production. Again, the firm can rent whatever total stock of physical capital it wants to use. Within each stream, production uses the two inputs, intangible capital and physical capital, with constant returns to scale. In the deployment of physical capital, the same unit (say, a machine) cannot be simultaneously used in multiple production streams. For the profit-maximizing firm, the marginal revenue product of applying physical capital across each production stream will be the same.

In principle, intangibles are non-rival in use, and hence the same intangible can be used simultaneously in different production streams. But in practice, intangible capital need not be completely non-rival within a firm and across production streams. Instead, there can be partial non-rivalry. In our model, the degree of non-rivalry in use for intangible capital within the firm can be thought of as a parameter ranging from 0 to 1. At one extreme, intangibles are rival within the firm. In this case, just like physical capital, using an intangible in a production stream precludes its use in another stream. A luxury brand cannot be used to market household cleaning products, and data about luxury spending can’t be used to plan cleaning product inventories. At the other extreme, the same intangible can be used in every production stream—that is, the firm could use its entire stock of intangible capital in each of the production streams. Payroll software can be used across the entire firm, as can a healthy corporate culture. In that case, investing in intangibles (in Stages 1 or 2) may generate economies of scale or scope.

The degree of non-rivalry in use may only be partial—that is, the parameter mentioned above may fall strictly between 0 and 1. This could arise as a result of imperfections in the storage technology. For instance, when intangible capital is stored within key employees, it may be difficult to communicate knowledge perfectly across different parts of the firm—information is often lost in translation. Similarly, software and brands may need to be customized to fit different production locations or product lines.
It is useful to think about partial non-rivalry in use of intangible capital within the firm in terms of the marginal rate of substitution between intangible capital allocated in any two production streams. Remember that each production stream is constant-returns-to-scale in this model, but some production streams can have a higher relative use of intangible capital than others. For a given total stock of intangibles, when there is perfect non-rivalry in use, increasing the intangible asset in one production stream does not require reducing its use in any other production stream. When non-rivalry is partial, increasing the use of intangible capital in one production stream does require reducing its use in another production stream—but less than one-for-one. In other words, increasing the intangible input in one production stream is not entirely costless, but it does not necessarily eliminate its availability for other production streams, either.

A firm will make its Stage 1 and Stage 2 decisions—about intangible capital and the span of the firm across production streams—with the knowledge that at Stage 3 it will be able to rent physical capital as needed. Thus, when thinking about the quantity of intangible capital, the firm will also consider the span of production streams over which that intangible capital could be applied, along with the extent of non-rivalry in use of intangibles across these production streams. In this sense, the quantity of intangible capital and the span of the firm across production streams are complements. In other words, a higher span of production processes further increases the returns to intangible investment. We next discuss how the firm chooses firm span (at Stage 2) and intangible investment (at Stage 1).

Stage 2: Storage Choice and Costs of Expanding Firm Span

At this stage, the firm will take as given the level of intangible capital. In the absence of any cost of increasing firm span, the intangible would be allocated to every possible production stream in order to take advantage of non-rivalry. This section introduces a tradeoff between the value generated by increasing the span of the firm and potential costs of doing so. This cost is not a physical cost, but rather arises endogenously from the limited excludability of intangibles.

We can think of an entrepreneur initially facing a storage choice: whether to retain the intangible as closely held or to codify the intangible and store it in an external medium, such as general labor or capital. This storage decision involves an important tradeoff. By storing more of the intangible externally, an entrepreneur or manager can increase the firm’s span of production and better exploit the non-rival nature of intangibles. On the other hand, by codifying the intangible and storing it externally, the entrepreneur reduces the cash flows, control rights, and rents that can be obtained from the intangible. In other words, codifying intangibles increases the size of the firm but reduces the share that is appropriable solely by the entrepreneur. More generally, the choice of storage medium could also affect the degree of non-rivalry. We discuss two specific examples in which greater span undermines excludability: imitation and incomplete contracts.

The issue of imitation arises because as a firm adds new production streams it enables the use of those intangibles across more production facilities, markets,
or geographies. More workers and more consumers are exposed to the firm’s intangibles, and so competitors may find it easier to replicate the intangible and to appropriate some of the returns it generates. This effect could vary with the type of intangible and its storage technology. For instance, the imitation problem will tend to be higher for intangibles that are easy to store in capital versus labor (and thus easier to copy) and those that are not well-protected by specific legal institutions, like software or research ideas. This potential for imitation limits the firm’s ability to exclude others from appropriating the benefits of its intangibles—and more so as the firm’s span rises. This generates a cost of increasing firm span that balances the benefits of non-rivalry.

A cost due to incomplete contracts arises because, if the entrepreneur keeps the intangible closely held, outside investors are subject to potential hold-up problems or information asymmetries. As a result, for the intangible to serve as a basis for external financing, it must first be stored and codified outside of the entrepreneur, so that future cash flow or control rights can be assigned to the firm’s investors. Once stored, returns to the intangible can then partially accrue to outside investors as the return on their investment in the firm. The costs of external finance will tend to be higher if the nature of the storage technology for the intangible makes pledgeability difficult. Intangibles stored in key talent, organization capital, and technical advances that are difficult for outside parties to evaluate (so that only the entrepreneur can assess its true value) present challenges to pledgeability. In turn, these challenges to pledgeability mean that the entrepreneur must give up a larger share of the intangible in order to obtain financing. Note that improvements to storage technologies can imply that the creators of intangibles accrue substantial value if the span of those intangibles increases substantially as external storage increases.

Under either approach to limited excludability (imitation or incomplete contracts), the entrepreneur loses the ability to claim fully the benefits of the intangible; this cost is captured by a limited excludability parameter, which can be thought of as ranging from 0 to 1. This parameter captures the decline in the share of the total surplus generated by the intangible that actually accrues to the entrepreneur as span rises, owing to the inability to exclude outside agents from claiming the benefits of the intangible. In the case of imitation, this may generate a gap between the private and social choice of span, where the firm may choose a lower span than is socially optimal, by ignoring the potential external benefits of spillovers. In the case of incomplete contracts, the entrepreneur who only controls a fraction of the firm’s intangibles will also choose span based on how much of their value she can capture, rather than on the full value of intangibles. Even if external financing brings in funding, losses in translation from codification or mispricing reduces the value of the intangible to the entrepreneur.

The limited excludability parameter can be thought of as capturing limits to property rights or information asymmetries. In the case in which the cost of storing intangibles comes from imitation, stronger and more well-defined property rights can lead to a larger increase in span per unit of codified intangibles. For the case in which codification improves pledgeability and external financing opportunities,
better property rights or more perfect information improves the tradeoff between the increase in span and the rents captured by the creator of the intangible.

The fact that increasing the span of the firm reduces excludability creates a tradeoff relative to the previous subsection. There, increasing span could generate returns to scale or scope—because of the non-rival use of intangibles—which is now constrained by limited excludability. How does the limited excludability parameter interact with the degree of non-rivalry in use parameter discussed in the previous subsection? If an entrepreneur faces a high degree of non-rivalry and also a high level of excludability, there will be an incentive to adopt an intangible-intensive production technology and to operate at high scale. After all, the span of a firm tends to increase with the degree of non-rivalry. However, the entrepreneur will only value the associated intangible asset to the extent that the benefits can be claimed and property rights are strong. If non-rivalry is high but excludability is low, the entrepreneur might instead steer away from investing in intangible capital, and instead pick a technology that emphasizes physical capital inputs and focuses on a single production stream. When it comes to investment in intangible capital, non-rivalry and excludability are complements.

Finally, this framework illustrates that an entrepreneur’s scale choices may be socially inefficient. If excludability is low, whether for imitation or incomplete contract reasons, the entrepreneur will be able to receive only a portion of the social returns to intangible capital. As a result, the entrepreneur chooses a lower level of intangible investment and/or a smaller span of production streams. For society as a whole, however, it would be preferable to have a higher level of investment and its positive spillovers.

**Stage 1: Intangible Investment/Creation**

The first step in the timeline of our model is to determine the initial investment in the intangible asset. Here, we can think of the entrepreneur as exerting effort in search of a profitable new idea, which will be for the new intangible. Higher effort is more likely to yield more profitable ideas, but a substantial level of risk is involved. For present purposes, we are especially interested in how the entrepreneur’s choice of effort depends on the parameters governing limited excludability and the degree of non-rivalry in use.

If excludability is low, then an entrepreneur will have less incentive to exert effort in generating intangibles, because their ideas can be expropriated. By contrast, the effects of the non-rivalry parameter on effort are more subtle, in a way related to the complementarity we have emphasized between non-rivalry and excludability. The scalability of creating intangible assets may generate value, but the entrepreneur will only value the associated intangible asset to the extent that the benefits are sufficiently excludable. Also as noted earlier, the model features underinvestment in intangibles since the entrepreneur’s effort choice depends on the parameters governing limited excludability and the degree of non-rivalry in use.

Perhaps surprisingly, the degree of under-investment can be greater for intangibles that are highly scalable if excludability is low enough. The intuition follows from the
complementarity argument above: since highly scalable intangibles only generate value to the entrepreneur if they are appropriable, the entrepreneur will especially undervalue highly scalable intangibles when they lack appropriability.

Finally, it is important to emphasize the distinction between expected and realized returns to the entrepreneur. If there is selection on which entrepreneurs enter the market (or equivalently if failure in creating an intangible asset is a feasible outcome despite the amount of effort involved) then focusing on compensation received by entrepreneurs will be misleading. With free entry of entrepreneurs, it is possible to have a situation in which the expected rents from creating intangible capital are zero, but the realized rents from doing so—looking only at those who succeeded—can be positive. This is often emphasized in analysis of patents, for example, where the observed payoffs may be high but may not fully capture the risk and the failures that are known before the patent is created but may be unobserved by looking at patent success stories.

Implications and Relation to Other Approaches

Our approach to intangibles, incorporating non-rivalry and limited appropriability, leads to unique implications compared to the standard neoclassical model. We now discuss the relationship between our model and the key properties of intangibles used in existing work, especially with regard to non-rivalry and limits to excludability.

A common premise in the literature on intangible capital is that it can contribute to “higher returns to scale,” or more generally, that intangibles are “scalable” (Haskel and Westlake 2018). A standard rationale is that to assume that intangible investment involves high fixed costs, but leads to lower marginal cost of production for the firm. As a result, production at intangible-intensive firms may be characterized by increasing returns to scale, at least locally. Of course, locally increasing returns are not specific to intangibles; for instance, the production of power from nuclear plants relies heavily on physical inputs but has the same profile of high fixed and low marginal costs.

Instead of assuming a particular structure of fixed and marginal costs that lead to increasing returns, our model starts from the idea of non-rivalry in use. As a result, our notion of “scalability” is somewhat different from the existing literature. In our model, “scalability” derives endogenously from the complementarity between intangible capital and firm span, which arises so long there is some degree of non-rivalry in use within the firm. In this case, the value of the intangible asset increases when it can be employed in multiple segments. Similarly, the higher the value of the firm’s intangible asset, the greater is the benefit of expanding the span of operations.

This scalability property is modulated by limits to the excludability of returns from investment in intangible capital. Here, we emphasize that greater excludability is not always desirable. The flipside of limits to excludability is that intangibles can generate spillovers outside of the firm. Ideas that are stored and widely disseminated can be used effectively in production—or even spur the development of better ideas. The use of a specific intangible asset by one firm may indirectly increase
productivity in other firms who can potentially adopt the same intangible. Negative spillovers are also possible. The same forces that lead to wide dissemination and adoption of new ideas imply that older ideas become more easily obsolete. A new and more efficient method of production can be licensed to many firms, leading to a drop in the value of the intangible asset (say, a patent) representing the old production method. The assumption that there are limits to excludability could also capture this process of “external” depreciation resulting from a firm’s investment in intangible assets (Jovanovic and Rousseau 2002).

**Intangible Capital and Economic Trends**

Economists typically estimate output and the stock of physical capital with greater precision than the stock of intangible assets. Indeed, measuring the stock of intangibles as an input to production is quite challenging, since they can be embodied in a variety of media, including human capital. However, accounting for intangible assets and understanding their unique characteristics can shed light on some key economic trends. In particular, the period since the 1990s has been characterized by relatively low growth in total factor productivity, a decline in the labor share, weak tangible investment and rising valuations, an increase in economic rents, and rising inequality.

**The Productivity Slowdown**

After a productivity boost in the 1990s, the United States (and other) economies have seen a widespread productivity slowdown during the last two decades. Based on a standard aggregate production function, growth in output can be decomposed into growth of each of the inputs, like capital and (quality-adjusted) labor. The unexplained “Solow residual” term is then taken as a measure of the change in total factor productivity. But if a substantial part of intangible capital is not captured in the statistics on “capital” inputs—and indeed, some of intangible capital investment is embodied in wages paid to, say, those creating intangible capital—can this help to explain the decline in measured productivity?

One can take two approaches in re-interpreting the official measures of productivity with intangible capital in mind: reinterpreting the Solow residual and re-estimating intangible capital.

The first approach would treat the entirety of the Solow residual as driven by incomplete measurement of intangible capital. In this view, intangibles and their properties are the “dark matter” that explain a wedge in between measured output and inputs. This is essentially the view adopted by some models of endogenous growth (for a recent survey, see Jones 2021). In this view, slower productivity growth could result from lower investment in intangibles. It could also result from changes in the degree of non-rivalry or appropriability, if the spillovers from intangible capital to the rest of the economy decreased. From this approach, the slowdown in the growth rate of measured productivity can be rationalized in three possible ways.
First, the benefits of intangible capital may be delayed by substantial time-to-build: as noted earlier, its effects may resemble types of tangible capital with high fixed costs of installation, but with negligible variable costs. For example, logistics-optimization software may require substantial development time, but once operational, it can improve delivery times for all of a firm’s production units. New production methods may take time to be adopted and for learning-by-doing to take effect. To the extent that these up-front costs are not recognized as investment, they can generate a slowdown in measured productivity over the short- and even the medium-run even as long-run productivity is higher (Brynjolfsson, Rock, and Syverson 2021).

Second, obsolescence of intangible capital may obscure measured total factor productivity growth. Because intangible capital can be superseded by innovation, it can become obsolete quickly. Combined long time-to-build lags and displacement of existing intangible capital during periods of rapid innovation can exacerbate the slow rise of measured output in the short run (Greenwood and Jovanovic 1999).

Third, the fact that the investing firm or entrepreneur cannot capture all of the value of intangible investment may reduce the incentive to create intangible capital internally. As we discussed earlier, the degree of non-rivalry or degree of appropriability not only affect output directly, but they also indirectly affect the incentive to exert effort to create new intangibles. Variations in the effective degree of appropriability can also lead to fluctuations in measured total factor productivity (as in the model of Kondo, Li, and Papanikolaou 2021).

The second overall approach to re-interpreting the official measures of productivity with intangible capital in mind re-estimates the stock of intangibles, rather than reinterpreting the existing Solow residual. In this approach, researchers aim to construct more “complete” measures of capital, in part by estimating intangible capital directly, and then to see how the total factor productivity residual adjusts. This approach proceeds by identifying expenditures and prices of excluded intangible investment, and then using them to construct estimates of the intangibles stock, sometimes with the help of an equilibrium model. For example, this approach is followed in Basu, Fernald, Oulton, and Srinivasan (2003) and Corrado, Hulten, and Sichel (2005), and also explored in McGrattan and Prescott (2010a, 2010b), and Crouzet and Eberly (2021a).

One lesson from this literature is that including unmeasured intangibles has opposing effects on the Solow residual measure of total factor productivity growth. For example, McGrattan and Prescott (2010b) argue that intangible capital is typically expensed (that is, treated as a production cost rather than as an investment), or else is financed by employees’ “sweat equity.” They show how the resulting underestimation of output and income, due to not measuring the production of investment in the form of intangible capital, mechanically leads to lower measured labor productivity growth in the 1990s. The effect on total factor productivity growth is less clear: underestimated intangible capital becomes both an additional output (which raises actual total factor productivity) and an additional input to production (lowering actual total factor productivity). These two opposing effects imply that
the exact timing of the additional input and the additional output affects the path of estimated total factor productivity.

This literature, as well as national statistical agencies, also faces open questions on how to measure intangible capital: for example, how to construct appropriate price indices to deflate past investment expenditures in intangibles. A further complication is that output, intangibles, and measured productivity in the United States may be mismeasured for other reasons; for example, due to the fact that US corporations have a tax incentive to book income from intangible assets abroad (Guvenen et al. 2021). The income that is booked offshore lowers gross operating surplus in the United States, thereby reducing measured value added. The impact on measured total factor productivity depends on the countervailing effects of missing intangible inputs and missing income. This is less of an issue in accounts that consolidate firms’ activities across countries, but it is a question for national-boundary-based measures.

From the perspective of our model, a main difference with this literature is that it treats intangible investment as an input with similar economic characteristics to physical capital; in effect, it assumes that intangible and physical capital are substitutes along with other factors of production. However, a fundamental property of intangibles is non-rivalry, which together with limited appropriability, can lead to positive spillovers and thus also affect measured productivity. To measure the full contribution of intangibles to economic output, researchers need to measure not only the intangible stock but also to account for the value of spillovers, which involves estimation of the parameters governing non-rivalry in use and limited appropriability.\(^1\)

### Factor Shares

The labor share of income in national accounting data has been declining, both in the United States and globally (for example, Elsby, Hobijn, and Şahin 2013; Karabarbounis, Loukas, and Neiman 2014). When interpreting these trends, however, the existence of intangibles implies that factor shares are also mismeasured. Depending on the implicit assumptions researchers make, the share of output that would accrue to intangible inputs could be allocated to either physical capital, labor, or “rents,” where the latter is defined as monopoly profits.

Our model helps shed some light on the underlying issues. Recall that the limits to excludability can be motivated by incomplete markets, in which the entrepreneur gives up some rents by finding a way to store some of the intangible externally, in exchange for funds for expansion from outside investors. In addition, at an optimum, the greater the degree of non-rivalry, the smaller the share that accrues to the entrepreneur. The entrepreneur chooses to give up a larger fraction of rents

\(^1\)To the extent that aggregate market values are used to measure the price of intangibles included in national accounts, some of these spillovers may be included in existing estimates of the intangible capital stock. But aggregate market values for intangibles are scarce, and even when these measures exist, it is not always clear how to allocate spillover returns across sectors (Moynan and Okubo 2020). Further, market values may underestimate the contribution of intangibles if these are partly stored in key labor inputs (Eisfeldt and Papanikolaou 2014).
to achieve a higher span of product streams (and scale). Should the entrepreneur’s remaining share be treated as returns to capital or to labor? If human capital is the key input in the production of new intangibles, it is labor income and hence part of labor share. The residual to which outside investors have a claim could be (though it need not be) part of capital income. Given that many intangible assets like patents, copyrights, and trademarks confer exclusivity and hence monopoly power, the conceptual distinction between monopoly rents and the factor share of intangibles may be hard to disentangle.

Cash flows from intangibles likely go to both labor (key talent) as well as owners of tangible capital (shareholders). The appropriate allocation likely varies across different types of intangibles, depending on how the intangible capital is stored, and then on its property rights and excludability. For example, managers may accrue income from intangible assets such as organization processes or corporate culture. Similarly, capital owners (shareholders) may accrue income from software, patents, or brands. In general, the full value of intangible capital will be observed partly on the market value balance sheet of firms, and partly on the market value balance sheet (wealth) of key talent. Variation in bargaining power between the two parties can lead to insiders appropriating a larger or smaller share. As a result, imputing the stock of intangibles based on firms’ market valuation ratios will likely underestimate the value of intangibles (Eisfeldt and Papanikolaou 2014).

Any calculation of labor and capital factor shares should make these distinctions, but many do not. For instance, Barkai (2020) measures monopoly rents as output minus labor expenses minus the stock of physical capital times its user cost. As a result, this measure of monopoly rents will include the share of output due to intangibles. Alternatively, if one estimates the capital share as one minus the wage payments to labor (Elsby, Hobijn, and Sahin 2013; Karabarbounis and Neiman 2014), this factor share for capital will include income from intangibles. If labor is a key input in the production of intangibles, part of that income should in fact be classified as labor income. To the extent that intangibles are stored in key employees, this choice can understate the labor share. In particular, some capital income and profits are actually equity compensation for high-skilled labor inputs and are thus partially misclassified. Eisfeldt, Falato, and Xiaolan (2021) document the large fraction of labor compensation in the form of equity-based pay in recent decades. Ownership of private firms can lead to compensation of labor inputs with capital income (Smith et al. 2019; Bhandari and McGrattan 2021). Note also that factor shares have an important impact on measuring total factor productivity, which we discussed in the previous section. In addition to the direct effect of measuring factor growth, misclassification of intangibles as intermediate inputs can bias factor shares and reduce estimates of total factor productivity (Crouzet and Eberly 2021a).

A further complication in interpreting trends in the labor share is that total output is also mismeasured if expenditures on intangibles are recorded as a cost of production, rather than investment, as we noted in our discussion of productivity measurement. Indeed, expenditures on intangibles tend to be recorded either as payments to labor or as purchases of intermediate inputs, such as consulting
services or lab equipment. Recent revisions to the US national income and product accounts capitalized certain intangible assets, especially intellectual property products, and added these “produced” assets to gross operating surplus and as income to capital. Indeed, this allocation of intellectual property products to the capital account in the US national income and product accounts generated most of the measured decline in the labor share (Koh, Santaeulàlia-Llopis, and Zheng 2020).

Because current practice tends to either omit intangibles from national accounts or to allocate their payments primarily to capital, the actual labor share is likely higher than the share computed using national income and product data (in particular after the statistical revisions in 1999 and 2013 that treated all of intellectual property as capital).

Inequality of Income and Wealth

Just as the treatment of intangibles can bias our view of capital and labor shares, the rising importance of intangible assets can lead to inequality between those who benefit from intangibles and those who do not. Here we discuss three mechanisms by which this could occur: i) rents may accrue to inventors and entrepreneurs, ii) rents may accrue to key employees, and iii) intangibles may exacerbate capital-skill complementarity.

Under the first mechanism, when inventors or entrepreneurs conceive and develop a new intangible, they can appropriate a fraction of the value generated. The rest of the value generated accrues to outsiders, including outside investors, or other firms and consumers in the economy more broadly. The key difference between these two parts is that the entrepreneur’s share is concentrated and not easily tradeable. Concentrated exposure to intangibles can lead to inequality through both the drift in the owner’s wealth (if the intangible is exposed to substantial systematic risk) and from the idiosyncratic shocks to the intangible’s value. The entrepreneur cannot pre-sell claims to future intangibles that have yet to be produced (otherwise, the incentive to exert effort in creating such intangibles after already receiving payment would be low). These early-stage intangibles will have very concentrated ownership. By contrast, by codifying and storing the intangible, the entrepreneur creates an asset to which outsiders can lay a claim. Moreover, these claims can be by diffuse investors who can build diversified portfolios.

This key distinction between the two shares—what is stored with the entrepreneur versus what is owned by outside stakeholders—can help to explain rising inequality, as Kogan, Papanikolaou, and Stoffman (2020) explore in a general equilibrium model. The key feature of their model is incomplete markets: during each period, a small measure of agents—the “inventors”—are randomly endowed with a

---

2 A substantial fraction of investment in research and investment is labor compensation. For example, Adobe’s research and development expenditures in 2019, and Sanderson Farm’s breakdown of Selling and General Administrative Expenses in 2019, as shown by the respective 10-Ks of these firms, show that the majority of their research and Selling and General Administrative expenditures are in fact labor compensation.
blueprint for a new project. They interpret these inventors broadly as encapsulating all parties that share the rents from new investment opportunities, other than the owners of the firm’s publicly-traded securities. This model suggests that firm owners as a group reap only part of the benefits of innovation, but bear all the costs of creative destruction. As a result, the arrival and churn in new technologies is associated with greater income and wealth inequality, together with a motive to insure against states of the world with rapid technological innovation.

The second mechanism begins with the insight that while part of investment in an intangible asset is codified in media, another part could be stored with key employees. An example would be management practices: a textbook or a business school can prescribe “best practices,” but it does not immediately follow that everyone who takes the class or reads the book becomes an effective manager. Further, because the value of the intangibles stored with key employees are likely to be partly specific to the firm, the cash flows generated by this form of intangible capital is often shared between shareholders and key talent. As intangibles grow in importance in terms of firms’ capital stocks and value, the importance of key labor inputs in contributing to and operating these assets may lead to inequality between key labor and other workers (McGrattan and Prescott 2010b; Smith et al. 2019; Eisfeldt, Falato, and Xiaolan 2021; Bhandari and McGrattan 2021). Further, the share of rents that accrues to key talent need not be constant, as it depends on their outside option, which introduces a further source of inequality (Eisfeldt and Papanikolaou 2013).

The third mechanism we consider is how intangibles can affect income inequality even when intangibles are all stored externally, if they exacerbate capital-skill complementary. A common view of technological progress since the mid-twentieth century is that it is primarily skill-biased—that is, technology is generally a complement to high-skill workers, but was a substitute for low-skill workers (Goldin and Katz 2008). If intangibles increase the marginal productivity of high-skilled labor inputs, then the rise in the importance of intangibles over the last few decades may have contributed to rising inequality. In support of this view, Eisfeldt, Falato, and Xiaolan (2021) find that pay to high-skilled labor, and in particular equity pay, grew fastest in recent decades within industries which were most exposed to declining investment goods prices, as a proxy for the growth in intangibles stored in capital goods. Although the model we developed earlier in this paper has no explicit labor inputs, it is relatively straightforward to include different types of labor into the model, possibly using a constant-elasticity-of-substitution framework similar to Krusell et al. (2000) and Eisfeldt, Falato, and Xiaolan (2021).

While the above three mechanisms imply that rising intangibles lead to higher inequality, the relation between intangibles and the level of income inequality can be ambiguous if new intangibles are also associated with skill displacement. Put differently, an increase in between-group inequality need not translate into an increase in between-worker inequality if workers transition across groups. Using detailed data on patent inventions and occupation task descriptions, Kogan et al. (2021) document that workers in occupations most exposed to technology improvements tend to experience declines in wage earnings. They find that the workers most adversely
affected are the highest-paid workers, which can be consistent with technology-skill complementarity only as long as workers can also lose part of their human capital when technology improves.

**Tangible Investment and Tobin’s $q$**

In recent years, investment in physical capital has been declining, while measures of the return to physical capital have been rising. Figure 3 illustrates this trend with data from US public non-financial firms. The blue line is the aggregate investment rate in physical capital. The orange line measures the rate of return to investment, measured as Tobin’s $q$—the ratio of total enterprise value (the sum of the value of equity plus debt, adjusted for liquid asset holdings), to the stock of physical assets. The two lines show a positive correlation at business-cycle frequencies (as highlighted in Andrei, Mann, and Moyen 2019). However, over the longer run, they diverge: the investment rate fell by about 4 percentage points, while Tobin’s $q$ increased by a factor of about 3.

For models of investment where firms only rely on physical capital (Hayashi 1982; Abel and Eberly 1996), explaining this long-run divergence is a challenge, as first noted by Gutiérrez and Philippon (2016). According to these models, Tobin’s $q$ should proxy for the marginal benefits of physical investment, because firm value scales proportionally to the stock of physical capital. Higher Tobin’s $q$ should signal higher returns to physical capital and encourage physical investment, at odds with the trend in Figure 3.

Intangible capital, using the model developed earlier, can shed light on this puzzle in at least two ways. First, with intangibles, the numerator of Tobin’s $q$ captures the market value of the whole firm, including benefits of intangible capital and any spillovers. The denominator, on the other hand, is only physical capital. Hence, the ratio overstates the true return to physical capital to the extent that value is also generated by intangibles. Put differently, with intangibles, average returns to physical capital will overstate the true incentive to invest in physical capital. The bias will grow as firms increase intangible inputs relative to physical capital, as the evidence in the introduction suggests they have in recent decades. Consistent with this intuition, Crouzet and Eberly (2019) show that, controlling for the ratio of intangible to physical capital, the difference in trends between measured returns and investment rates for physical capital highlighted in Figure 3 shrinks by about 30 percent.

---

3Tobin’s $q$ is an empirical measure of the average return to physical capital: total enterprise value per unit of physical capital. Moreover, in traditional models, the average and the marginal return to physical capital should be the same, because the firm is assumed to be able to scale up revenues one-for-one with physical capital. Finally, the marginal return to physical capital—how much enterprise value rises for each incremental unit of physical capital—measures the marginal benefit of investing in physical assets. As a result, investment in physical capital should be tightly linked to Tobin’s $q$.

4The online Appendix to the paper, as well as the companion paper (Crouzet et al. 2022b), provide a more formal discussion of these two points.

5This argument was first outlined by Hayashi and Inoue (1991).
Second, for a given ratio of intangible to physical capital, changes in the degree of non-rivalry and excludability of intangibles will also affect the relationship between measured average returns and investment rates for physical capital. With a higher degree of non-rivalry within the firm, marginal returns to intangible investment will rise, so that more of total enterprise value will be accounted for by intangibles. An example could be the growing availability (and declining price of) of digital media, which facilitate the replication and scaling of intangibles stored in software. Similarly, higher excludability (say, from strengthening enforcement of property rights institutions, like patents) would increase the share of cash flows that can be retained by firm owners, and therefore the contribution of intangibles to total enterprise value. In both cases, the gap between measured returns to physical capital and the true marginal product will rise (or conversely would fall with weaker non-rivalry and excludability). These forces could also explain the trend presented in Figure 3.

The wedge between average and marginal returns to physical capital can be affected by a rising share of intangibles in other ways, as well. As we noted in our discussion of productivity measurement, the arrival of new intangible assets (say new designs) may reduce the value of existing physical capital, so that average returns to investment could fall even as marginal returns to investment rise. As an illustration, Greenwood and Jovanovic (1999) found that the arrival of information technology led to a fall in the value of existing firms. In a model with technology...
shocks embodied in new capital goods (for example, Papanikolaou 2011), improvements in new capital goods can lead to a decline in the average rate of return if the replacement value of the installed capital stock (the denominator in calculating Tobin’s $q$) does not adjust fully to offset the decline in the market value of incumbent firms. In this case, one could observe rising marginal returns and new investment together with weak Tobin’s $q$.

**Rents and Market Structure**

We define rents as returns generated by an asset in excess of its marginal (user) cost. Like physical capital, intangible capital can generate rents for its owners. Rents relate to market structure because they often arise in situations where capital is used to produce a good for which there are poor substitutes. For instance, a drug formula produces rents to the extent that it is difficult to produce a generic substitute. These conditions can lead to imperfect competition and, in some cases, to greater market concentration.

In situations where intangible capital generates rents, the rent is the stream of appropriable returns generated by the intangible in excess of its user cost. In a business franchise, for example, the intangible asset is the combination of the brand with the logistical and organizational instructions provided to the franchisee. To generate rents, the asset must produce returns that exceed cost of using it (paying the franchiser, implementing organizational instructions, and possibly further promoting the brand). It may be possible for the franchiser and franchisee to appropriate these rents—for instance, through enforceable franchise agreements.

Conceptually, the rents generated by an intangible or tangible asset should be measured separately from the intangible itself. In this approach, the value of a firm can be divided into four categories: the value of physical capital, value of rents from physical capital, value of intangible capital, and value of rents from intangible capital. Crouzet and Eberly (2021b) show that this decomposition holds in a broad class of dynamic investment models in which capital inputs, both physical and intangible, can generate rents. Moreover, they describe how to estimate the components of the decomposition using a set of statistical moments, including Tobin’s (average) $q$ for physical assets, flow returns to physical capital, and an estimate of the ratio of intangible to physical capital. The key finding is that rents associated with intangible assets have contributed to a sharply rising share in the growth of total enterprise value of US businesses since the early 1990s, accounting for approximately 15 percent in the mid-1980s and to up to 40 percent in 2015, depending on how broadly intangibles are measured.6

---

6Crouzet and Eberly (2021b) allow for intangibles and market power within a neoclassical framework which does not incorporate non-rivalry of intangibles nor for limits to excludability. Allowing for the features we introduce here, non-rivalry of intangibles would tend to increase rents, while limits to excludability would reduce them. Thus, one should expect industries in which intangibles are highly non-rival but easy to exclude to have particularly high rents from intangibles. A potential example is the health care industry, where intangible assets (like drug patents) are easy to replicate within the firm but well
As discussed earlier, when creating intangibles involves fixed costs, their non-rivalry within the firm can generate scale economies. Scale economies in turn may lead to higher market concentration. Note that this does not necessarily require that firms earn rents in the first place, so that higher concentration need not go hand in hand with more rents. Instead, the effects of non-rivalry on concentration are closely related to the theory of natural monopolies, which also emphasizes that markets featuring some firms with high-fixed, low-marginal cost structures will have high equilibrium concentration (Baumol 1977). Thus, more intangible-intensive industries may naturally be more concentrated, more intangible-intensive firms should command higher market shares, and intangible intensity within firms should be correlated with market share. Crouzet and Eberly (2019) provide reduced-form evidence consistent with these predictions, using data on publicly traded US firms.

Intangibles might also be conducive to higher concentration by encouraging consolidations across firms. Non-rivalry implies that, rather than having two firms each bear the fixed cost of creating the same intangible asset, it may be efficient for them to merge and share the cost. For instance, two retailers might merge and operate under the same brand, rather than creating and promoting their brands separately. More broadly, the non-rival nature of intangibles may play into determining the boundaries of the firm.

**Implications for Corporate Finance and the Cost of Capital**

Even when intangible assets can be stored, limited excludability implies that intangibles are less likely to be pledgeable to outside investors than physical capital. Lack of pledgeability undermines the viability of debt contracts, so that debt is less likely to be the preferred form of financing (Falato et al. 2020). An alternative possibility, Sun and Xiaolan (2019) argue, is that intangibles are primarily financed by employees, who then have an implicit claim on them. Related to this point, the capital structure of new firms with substantial intangibles appears to include more concentrated control in the form of dual-class shares whereby initial owners, founders, and employees retain more voting rights than outside equity holders (Ritter 2022).

The presence of intangibles can also affect the cost of capital. Given that intangibles may be stored in non-capital inputs, the sharing rule for cash flows generated by intangible capital may also expose investors in firms relying on intangible assets to unique risk factors (Eisfeldt and Papanikolaou 2013; Eisfeldt, Kim, and Papanikolaou 2020). Further, if the economic value that is generated by new ideas cannot be fully pledged and hence diversified to outside investors, then states of the world with rapid technological innovation will be associated with higher inequality, creating a
demand for insurance against these states and leading to a lower cost of capital for fast-growing firms (Kogan, Papanikolaou, and Stoffman 2020).

Concluding Thoughts

We have sought to describe intangibles affirmatively—rather than simply a lack of physical form—as a way of illustrating their role in production more clearly. This topic offers many opportunities for future research. Understanding how measured productivity can remain so weak in the face of seemingly continuous innovation remains puzzling. Further research on measurement and the interaction between intangibles and other factors, especially through non-rivalry, may shed light on this apparent contradiction. The factor income earned by intangibles is particularly difficult to classify as capital or labor income, or to allocate within types of labor. Further refinement of the allocation of rents accruing to intangibles may illuminate sources of market power, especially arising via the excludability conferred on intangibles through patents, copyrights, and other institutions. The distribution of these rights may also shed light on the sources of rising inequality. Further exploration of potential connections between intangibles, market power, and industry concentration is a rich area for future research.

Moreover, we suspect that the rise of intangible assets may spur institutional developments. As we noted, firms are relying more on intangible capital for production, yet many types of intangible capital do not yet have an institutional framework to use it as collateral for financing. Some intangibles have lively secondary markets: as one example, in licensing of patents and copyrights. Other intangibles seem less separable from the rest of firm, and the institutional framework is less clear: for example, some states have widespread non-compete agreements, based on a belief that such agreements provide greater excludability for certain kinds of intangible capital, while others do not. Economists may have much to contribute to developing and implementing the potential tradeoffs between different institutional frameworks that affect incentives to invest in intangible capital.

We thank Andrew Atkeson and Mindy Xiaolan for helpful discussions and the editors for their guidance and helpful comments. Also, we thank Edward Kim (UCLA Anderson PhD student) for helpful research assistance.
References


