

Information: Hard and Soft*

José María Liberti
Kellogg School of Management
Northwestern University
and DePaul University

and

Mitchell A. Petersen
Kellogg School of Management
Northwestern University
and NBER

March 2017

Abstract

Information is an essential component of all financial markets and transactions. However, information can arrive in multiple forms. In this paper, we begin to define what is meant by hard and soft information. Hard information is quantitative, easy to store and transmit in impersonal ways, and its information content is independent of the collection process. Technology has changed and continues to change the way we collect, process, and communicate information. This has fundamentally transformed the way financial markets and institutions operate. One of these changes is a greater reliance on hard relative to soft information in financial transactions. This has altered the design of financial institutions by moving decisions outside the traditional boundaries of organization. We discuss the advantages and limits of hard and soft information and the possible consequences of hardening information on both financial markets and institutions. We provide several applications of soft information and avenues of future research.

Keywords: soft information, hard information, boundaries of firm, discretion

JEL codes: G20, G21, G30

* This is a significantly revised version of the working paper previously circulated as “Information: Hard and Soft” (2004) by Mitchell A. Petersen.

I. Introduction: The History of Soft and Hard Information

Information is an essential component of all financial markets and transactions. A major purpose of financial markets and financial institutions is to collect, process, and transmit information. Given the importance of information and communication in the study of finance, as technology changes the way we communicate, it must fundamentally change financial markets and institutions, especially financial intermediaries. However, newer technologies (i.e., those developed in the past fifty years) are more adept at transmitting and potentially processing information that is easily reduced to numbers. We call this hard information. Information that is difficult to completely summarize in a numeric score is what we call soft information. Information technology developments have changed the organizational design and scope of financial markets and, in particular, the organizational design of lending. Improvements in the information environment of borrowers have altered the landscape of lending by moving financial decisions outside the boundaries of the financial institutions.¹

The transformation of banking over the past five decades, from a local and personal to a national, competitive, and impersonal market, has been partially propelled by the different types of information on which credit decisions are made. Soft and hard information are important characteristics of the environment and of these decisions. The type of information helps explain both market developments and institutional design. Banks have historically been a repository of information about borrowers' creditworthiness and the kinds of projects available to them. This

¹ Petersen and Rajan (2002) show that the distance between small borrowers and their lenders increased and the borrower-lender communication is done in a more impersonal way. Berger, Miller, Petersen, Rajan, and Stein (2005) show that smaller financial institutions are better able to collect and use soft information relative to hard information than large ones. Specifically, they study whether the nature of the organization affects the tasks of the individuals providing certain services. They find that large banks lend at a greater distance and interact in an impersonal way with their borrowers.

information was collected over time through frequent and personal contacts between the borrower and the loan officer. Over time the banks built up a more complete picture of the borrower than was available from public records. Since banks focused on the borrowers that were most difficult to monitor and evaluate (for example, small and medium-sized private firms), much of the information they collected was soft information. This was the source of the banks' competitive advantage relative to arm's-length lending markets. As technology has increased the opportunities and competitive pressure the incentive for using hard information, the description of banks as repositories of soft information has changed.

In this paper, we first discuss what is meant by hard versus soft information, and its applications in the academic literature. Before defining the terms, we briefly discuss the use of these ideas in the finance literature in Section II. Then in Section III, we discuss what is meant by hard and soft information. We examine the potential benefits and costs of using hard instead of soft information in financial decisions in Section IV. Finally, we consider the possibility of hardening soft information in Section V. Section VI presents a series of applications for the use of soft information, and what implications this has for both financial markets and finance research. Section VII concludes.

II. Soft and Hard Information in the Finance Literature

A. The Theoretical Literature

The concept of soft and hard information has been widely developed in the organizational economics literature. The finance literature has been exploring the distinction between soft and hard information for several decade now and our understanding has evolved since the early years. The distinction is not always explicitly stated, and even when it has been, the definition is not incomplete, formally treated or consistent across applications. In the theoretical literature, the

distinction between the role of banks (or other private lenders) versus the public bond market is driven in part off the superior ability of banks to collect and process information (Diamond 1984, 1991; Ramakrishnan and Thakor 1984). However, the public debt markets, with the help of rating agencies, have the same job description. The difference is the type of information each specializes in collecting and processing. The public bond markets and the rating agencies collect financial disclosures, accounting reports, and default histories. Each of these has the feel of hard information. They can all be reduced to a series of numbers. Banks, on the other hand, especially as described by the lending relationship literature, collect information that is neither initially available in hard numbers (the ability of the managers, their honesty, the way they react under pressure), nor are they easily or accurately reducible to a numerical score. Even if reduced to a numerical score, the interpretation of the information may be judgmental and include a discretionary component. Once the relationship is established, even then this information is not hard. The firm is still unable to communicate this information to the broader lending markets and thus negotiate a lower loan rate from its bank (Petersen and Rajan 1994).

The literature on organizational form has also exploited the distinction between hard and soft information to help explain the scope of the firm (Stein 2002). In many industries, both small and large firms coexist. One might think that a dominant production technology would lead to a uniform firm size. However, if the information collection, processing, and communication is fundamentally important to the production process (e.g., banking or drug research), then firms may specialize in different sectors of the market depending upon the type of information (hard or soft) that is an input into their production process. Some firms may specialize in production processes based on soft information, and others in a production process based on hard information. Stein (2002) argues that larger, more hierarchical firms, where the decision maker is further from the

information collector, are more likely to use production technologies that rely on hard information. If innovations in communication technology have reduced the cost to access information at a distance possible for production technologies, then financial institutions have lost the unique property of accumulation and utilization of knowledge.

B. The Empirical Literature

The distinction between soft and hard information has also been used in the empirical relationship banking literature to explain organizational structure and access to capital. Due to the organizational diseconomies described in Stein (2002), large banks are expected to be less efficient at making relationship loans—that is, loans that depend upon soft information. Information in a large bank is potentially collected by one individual or group, and a decision is made by another. Thus, decisions require information that is easy to transmit across physical or organizational distances. The information must also have a uniform interpretation that does not depend upon the context under which it was collected. Large banks are also more likely to have multiple layers of management—that is, be hierarchical or centralized as opposed to flat or decentralized organizations. Thus, the oversight of loans in this context again implies that larger banks rely relatively more on hard information. Soft information is collected by lenders through personal interactions with the borrowers over time. In particular, proximity between the lender and the borrower facilitates the collection of soft information (Petersen and Rajan 2002; Berger, Miller, Petersen, Rajan, and Stein 2005; Hauswald and Marquez 2006; Mian 2006; Liberti and Mian 2009; Agarwal and Hauswald 2010). These papers addressing questions on soft information used geographical or organizational distance between lenders and borrowers and between agents within organizations as a proxy to ease the collection of soft information.

Consistent with Stein (2002), Berger, Miller, Petersen, Rajan, and Stein (2005) find that larger banks are more likely to lend to more distant customers (when there is greater physical distance between a firm and its bank) and communicate with borrowers more impersonally (by mail or phone opposed to face to face). They also find that relationships between a firm and its banks are less durable and less exclusive when the banks are larger. Most importantly, they find that firms that are forced to choose a larger bank than they would prefer (i.e., informationally opaque firms) are credit rationed. When informationally opaque firms have the choice of which size bank to borrow from, they choose to borrow from smaller banks; this matching alleviates much of the credit rationing.

Liberti and Mian (2009) use organizational distance between loan officers and their superiors to study the causal impact of hierarchical structures on the relative importance of hard and soft information in credit approval decisions inside a large financial institution. The authors find that greater hierarchical distance is associated with less reliance on soft information and more on hard information. They also find that personal interaction between loan officers and the superiors approving the loans helps mitigate the effects of hierarchical distance on information use and minimizes the loss of soft information transmission.

More generally, the empirical literature has found that firms' access to capital depends upon how informationally transparent the firms are or how much hard information the financial markets have about the firms. We expect small firms to face greater credit rationing. This is why they are more reliant on banks that are better at extracting and using soft information. However, when we look only at small firms, we still find that a firm's access to credit is a function of how much information is available to the financial markets—not just the bank. Firms that are more

informationally transparent—for example, if they maintain formalized records—find that they have a higher probability of their loans being approved (Petersen and Rajan 2002).²

For publicly traded firms, the amount of hard information available about the firm is much greater. However, even for publicly traded firms, the existence of information that is easy to access and evaluate on their likelihood of default—such as a credit rating—appears to increase their access to debt capital (Petersen and Faulkender, 2006). Controlling for traditional determinants of capital structure (e.g., taxes, asset tangibility, and growth opportunities), firms with a debt rating have 35 percent more debt than otherwise identical firms. Controlling for industry at the four-digit level or controlling for firm characteristics using firm dummies does not change the result.

III. Characteristics That Define Soft and Hard Information

We do not propose a simple definition of what information is hard and what information is soft. Instead, an objective of this paper is to describe a common framework for defining hard and soft information which is both consistent with (much) of the literature but also useful for framing finance questions. Like many labels in finance (e.g., debt and equity), there is not a clear dichotomy. Rather than two distinct classifications, we should think of a continuum along which information can be classified. Our interest is what characteristics of information, its collection, and processing makes it hard or soft, and how these characteristics influence the structure of financial markets and institutions.

Hard information is almost always recorded as numbers. Thus in finance we think of financial statements, the history of payments which were made on time, stock returns, and the

² We can think of geographical distance as proximity that facilitates monitoring activities not only in banking but in other activities as well. For example, venture capitalists are more likely to serve on the boards of local firms as opposed to more distant firms since monitoring becomes easier due to personal interactions (Lerner 1995).

quantity of output as being hard information. Soft information is often communicated in text.³ It includes opinions, ideas, rumors, economic projections, statements of management's future plans, and market commentary. The fact that hard information is quantitative means that it can easily be collected, stored, and transmitted electronically. This is why the advent of computers, large database programs, and networking has been such a boon to production technologies that rely on hard information (e.g. quantitative lending and quantitative trading).

A second dimension of hard information is the way in which it is collected. The collection method of hard information need not be personal. Instead, information can be entered into a form without the assistance of or significant guidance from a human data collector. Alternatively, the data collector does not need to understand the questions that the information will be applied. This has the advantage of expanding the geographic and time dimensions across which data can be collected. With computers, web-based forms, and networks, the information can be collected at any time and almost any place. The only required input is access to the person with the required information. This collection method lowers the cost of collecting the data, but limits what data can be collected.

This characteristic of hard information means the collection of the information can be separate from the use of the information. The meaning of the information depends only upon the information that is sent. When I code the information and transmit it to someone else, they know all that I know, or at least the portion of what I know that is useful. With soft information, the

³ Can't text be converted to numbers and then stored and transmitted electronically? Text files can obviously be translated into numbers; this is how it is stored. Can't text files be processed electronically? Again, the answer has to be yes, conditional on what one means by processed. The ability of computers algorithms to process and generate speech (text) has improved dramatically since we first discussed soft and hard information. Whether it can be interpreted and coded into a numeric score (or scores) is a more difficult question. The question is how much information is lost in the process. This process is what we call the hardening of information and will discuss it below in Section V.

context under which the information is collected and the collector of the information are part of the information. It is not possible to separate the two. This constrains the environments in which the data is collected and used. It has to be predictable. In some environments, prior to entering the situation and collecting the data (e.g. talking to a potential borrower), we know what variables we will collect, what possible values each variable can take (i.e., a signal will be either good or bad; one, two, or three), and why it will be valuable in making a specific decision.⁴ In other cases, prior to commencing our search for information, we are not sure what we may find or why it is valuable until we are collecting the data and sometimes not until much later. Think of this as experience.⁵ Later, when we are confronted with a decision, we recall the collected information (e.g. the experience), and it is only then that it is apparent (hopefully) how the information we collected is useful. This is another characteristic of soft information. If we don't know what the information will be used for, or which parts of the information are relevant or useful, it is difficult to code and catalog it for future use.

This characteristic of hard information means it is possible to delegate the collection of the data and have it be distinct from the person making the decision. It may also be possible to delegate the decision-making based on hard information to another person or an algorithm. Soft information must be collected in person, and historically the decision maker was the same person as the information collector.⁶ This is the intuition in Stein (2002) for why smaller, less hierarchical firms

⁴ A firm's sales revenue is an example of such hard information. There is wide agreement as to what it means for a firm to have had sales of \$10 million last year. However, if we say the owner of the firm is honest, there is less agreement about what this means and why it is important. My definition of honest may be different from yours.

⁵ This distinction also arises in the way the first econometrics course is taught and what actually happens when PhD students start their research. In the econometric class, we start by assuming we know what the dependent variable and independent variables should be. We know the functional form; the only piece is the value of the slope coefficient. The econometrics is set up to test the hypothesis that β is zero (or some other value). When we start doing research we realize we don't actually know which independent variables (or dependent variable) is correct or what the correct functional form should be.

⁶ A typical example is that of a relationship-based loan officer. The loan officer has a long history with the borrower and, based on a multitude of personal contacts, has built up an impression of the borrower's honesty, creditworthiness,

are better able to use soft information in their decisions. It is also why we see relationship lending is associated with soft information.

Knowing what information you are looking for and why it is valuable (i.e., for what it will be used) is essential if information collection and possibly decision making will be delegated. If we knew what information we were looking for, we could specify instructions for collecting it and then delegate the collection to a subordinate or a web page. Part of the reason hard information is more efficient (consumes less labor or less expensive labor) is because the collection and processing of the information can be systematized, automated, and delegated. It is necessary to have an expert write out the rules or procedures by which the information is collected and processed. However, once the rules are specified, it isn't necessary to have the expert be a part of the actual data collection. It must be feasible, however, to specify which pieces of information (numbers) must be collected and coded prior to the actual data or actual environment being seen by the expert. In this way, the expert's knowledge has been embodied into the data collection rules. If the information is hard and you know all possible variables and values, then automating the decision process is also feasible. In this case, the expertise of how to make the decision given the possible inputs is also automated by embodying the decision rules into a computer program, for example. Automated loan approval is one such example. This type of automation works, however, only when the information is hard, i.e. easily concentrated into a fixed set of numbers that uniformly communicate the relevant information.

Hard information may or may not be public and may or may not be verifiable by a third party. One can always create a numerical score with soft information. One can create an index of

and likelihood of defaulting. Based on this view of the borrower and the loan officer's experience, the loan is approved or denied. Uzzi and Lancaster (2003) provide descriptions of the interaction between borrowers and loan officers.

honesty from 1 to 10 or an index of the transparency of financial markets across countries. This in and of itself doesn't make the information hard. My interpretation of a 3 on this index must be the same as yours.⁷ Financial statements of a publicly traded firm are public and verifiable, but the payment history of a borrower is information that is private to a lender and not directly verified by a third party. By contrast, soft information is private and not verifiable as it involves a personal assessment and depends upon the context, which is not easily captured and communicated. I can produce records that a borrower has paid their bills on time (hard information), but I cannot fully document that a borrower is honest as this relies on multi-dimensional observations and my personal assessment and standards.

Our description of hard and soft information raises the question of whether hard information is less precise than soft information. Is hard information just soft information with some of the data thrown away? Think of a digital version of an analog signal. One of the plausible reasons to collect soft information is that it is more valuable than automated hard information since we do not separate the collection from the use. Understanding how precise soft or private information is important since it has implications on the social benefits of public information (Morris and Shin 2002).

IV. Advantages and Disadvantages of Hard Information

A. Lower Costs of Production

The evolution of financial markets over the past forty years has been in part a replacement of soft information with hard information as the basis for financial transactions. The full

⁷ A distinction between hard and soft information is related to the distinction in the contract literature of whether a signal that is observable by outsiders is also verifiable by outsiders (Hart 1995). For a signal to be verifiable, the interpretation of the signal by the two contracting parties—and any third party who may be required to enforce the contract—must be the same. This is a characteristic of hard information. Following the organizational economics literature, hard and soft information can also be referred to as objective (hard) or subjective information (soft).

ramifications of this transformation are not yet fully apparent, and there seem to be both advantages and disadvantages of the change. One of the major advantages of using hard information is lower transactions costs. These come from several sources. First, by its nature, production technologies (such as loan origination) that depend upon hard information are easier to automate. The job of collection and in some cases processing of information can be delegated to lower-skilled workers or computers. Thus, expensive labor can be replaced by cheaper labor or cheap capital. This has been the source of productivity gains in the manufacturing sector. The substitution of soft information for hard information may offer similar productivity gains. The size of these is an empirical question (see Petersen and Rajan 2002 for one example).

Hard information is also more standardized. By construction it arrives in the same format and is processed in the same way, for each application or transaction. This standardization introduces savings into the production process due to economies of scale. Once the computer system is designed and built to retrieve credit scores from the credit bureau and then make an approval decision on a credit application, adding additional applications to the system has a small incremental cost. This is one reason why lending based on hard information (credit cards) has come to be dominated by large lenders much more so than traditional relationship lending.

The greater reliance on hard information may also increase the competitiveness of these markets.⁸ First, the standardization of information and the resulting lower transactions costs can expand the size of the market. For small business loans and consumer loans, the size of the loans isn't large. Thus, large fixed costs can make such loans prohibitively expensive.⁹ One of the

⁸ Greater competition, arising from deregulation for example, also increases the pressure to lower costs and thus transform the production process to depend more on hard information.

⁹ For small business loans, the size of the fees is independent of the size of the loan. Thus, the percentage fee declines with loan size (Petersen and Rajan 2001).

advantages of hard information, and thus automation, is that it can lower the transactions costs sufficiently to expand the number of suppliers who can profitably offer such loans or services. In addition to expanding the number of suppliers in a given market, a reliance on hard information can also increase the geographic reach and competitive impact on existing suppliers. Availability of high quality, quantifiable information and improvements in credit scoring technology facilitates standardization. Standardization is based on attributes that inherently use verifiable hard information as opposed to soft information. This same standardization has contributed to the development of different markets. The evolution of the mortgage and signature loan (now called the credit card) market is an example.¹⁰ In the 1950s, the market was local and based on soft information obtained through personal contact. It is now national and based on hard information often obtained through impersonal contact. This has led to a wider availability of and arguably cheaper capital for the middle class (Nocera 1995). The nature of the information may also increase the competitiveness of the markets. Once the information is systematized and easy to communicate (hard), it also becomes more difficult to contain. In the early years of the credit reporting agencies (e.g., Bradstreets or R. G. Dun), only a summary of the information the agencies had on borrowers was published in their quarterly books (Carruthers and Cohen 2001). This disclosed information was quantitative and easy to compare. Merchants in need of more detailed and in-depth reports would purchase detailed reports on specific customers. By keeping the information difficult to replicate and transmit, by maintaining its softness, the credit reporting agencies hoped to maintain their control over the information and thus extract greater rents from the information they had collected. Once information is hard, passing it along and capturing its full value is easier. Information in electronic form can easily be passed along. Information that is hard can be

¹⁰ Subprime mortgage loans are less standardized and more informationally sensitive than normal mortgages because sometimes borrowers are not able to provide full disclosure of their income.

understood independent of the collector and the context under which it was collected. If the collector is not necessary, once the user has the data, this makes charging high rates for the data more difficult.

The lower costs of producing hard information can depend on more than just the nature of the information. It may also depend upon the organizational design of financial institutions. Lenders who are larger in size and hierarchically organized benefit from economies of scale in using hard information but can find it more costly to transmit soft information. This will cause them to place a greater reliance on hard information and can improve their ability to compete. Degryse, Laeven, and Ongena (2009) find that a banks geographic reach is smaller when they compete against such banks which thus focus their lending on hard information loans.¹¹

B. Durability of Information

The durability of information is also greater when it is hard. The fact that it is easily stored means that the cost of maintaining it for future decisions is low. The fact that the information can be interpreted without context means that it is possible to pass it along to individuals in different parts of an organization (Stein 2002). Firms no longer need to be part of the data collection process to be part of the decision-making process. This is especially important if the people involved in data collection are not expected to be around later. Given the increased turnover in many finance professions (loan officers or investment bankers), the movement toward hard information seems inevitable. As described in Crane and Eccles (1998), junior investment bankers used to rise through the bank as junior employees of their clients simultaneously rose through the ranks at their own firms. By the time junior bankers became senior bankers, they had developed a relationship with

¹¹ Reducing the costs of communicating soft information, for example when reporting lines in a bank coincide with long term personal relationships, the cost of communicating soft information decreases and this increases its use in the lending decision (Qian, Strahan, and Yang, 2014).

the people who were now in senior positions in the bank's clients. There is no need to rely on formal records (hard information) in the presence of these long-term relationships. However, if bankers turn over more frequently, new bankers must rely on the records left behind by the previous bankers. This creates a greater reliance on hard information.

C. Lost Information

Part of the reason that hard information is less costly to communicate is that there is less information. The replacement of soft with hard information inevitably results in a loss of information. This is why it is possible to use a smaller bandwidth to transmit the information. As an example, compare two methods of making a loan approval decision. First is the stereotypic credit scoring decision, in which a finite number of quantitative variables are weighted and summed to obtain a credit score. Based on the score, a decision is made to approve or deny the loan. Compare this to the traditional lending relationship story of how a loan is made. After spending several hours discussing the borrower's investment plans and using the loan officer's years of experience with the borrower, a decision is rendered. Both decision-making methods lead to a loan decision, but the first requires less information as inputs to the decision.

The reduction of information is never good, as long as processing costs are zero. However, in reality the loan approval committee or risk management committee of a bank has a limited amount of time and attention to devote to each decision. Thus, to prevent information overload, they need the information to be stripped down to what is important.¹² The larger the organization and the higher we go in the organization, the more the information needs to be concentrated (or

¹² Friedman (1990) argues this is one advantage of a market economy. All of the information that is relevant to a consumer or producer about the relative supply of a good is contained in the price. He argues it is not necessary for a supplier to know whether the price has risen because demand has risen or supply has fallen. The supplier only needs to know that the price has risen, and this will dictate his decision of how much to increase production.

decision making delegated). This concentration occurs when soft information is replaced with hard information. Granted, information is lost—in some sense this is unavoidable (as when an analog signal is converted to a digital signal). The question is how much and what type of information is lost.

The concern about small firms' and individuals' access to capital in the presence of bank consolidation and the growing use of credit scoring-type lending decisions is driven by this question. If there are borrowers that are really good, but look bad on paper (i.e., when we look at only the hard information), then such borrowers would be incorrectly denied credit and thus credit rationed. The empirical evidence thus far is mixed. It is clear that some small borrowers are dislocated by their banks when the banks merge, but there is also evidence that existing and new small banks may be filling the gap. The underlying empirical question is, how accurate can the hard information be. We return to this question when we discuss the hardening of information below.

D. Gaming the System

Thus far we have focused our discussion on the decision maker (e.g., the bank making a loan decision), not the target of the decisions (e.g., a loan applicant). By choosing to use hard versus soft information, the decision maker is trading off the transactions costs of collection and processing the information with potential losses in accuracy of the information upon which they are basing their decisions. However, the way a decision is made and the type of information upon which the decision is made will also influence the actions of the target of the decision.

Accounting numbers, a firm's income and balance sheet, are a classic example of hard information. The information is all quantitative, easy to store and transmit electronically, and there

is uniform agreement about what numbers like revenues and costs mean. This is why quantitative decisions from asset allocation to credit approval rely on these numbers. However, the newspaper accounts of accounting manipulation and the manuals put out by the credit rating agencies make it clear that these decisions are not simply a function of the numbers the firms disclose. Neither do most academic economists expect that mechanical trading rules based on the disclosed financial numbers can earn risk-adjusted positive returns. The implicit assumption is that not all useful information is summarized by the numbers. Some of the relevant data is qualitative and requires a judgment call. It cannot accurately be reduced to a number.

There is also an incentive for not requiring decisions such as credit rating to depend only upon the numbers or at least not disclose the decision rule. Having a decision depend only upon the numbers and a transparent decision rule can work only if the cost of manipulating the numbers is sufficiently costly relative to the benefits.¹³ If a firm can raise its reported assets or sales by a small amount for a small cost, and this would raise its credit rating and lower its cost of capital, it would have an incentive to inflate its reported assets or sales.¹⁴ The rules cannot be a direct and transparent functions of the hard numbers if the hard numbers are under the discretionary control of the market participants. In this case, the decision maker has an incentive to make the decision a

¹³ In the financial crisis of 2008, a large number of investment grade securities defaulted. The magnitude of the defaults suggested there was a problem with the rating process. Observers in industry, academics, and government suggested possible sources of the problem and potential solutions. What is intriguing is the defaults were concentrated in securitized lending market (e.g. RMBS) opposed to the bonds of operating companies. Thus, the problem with the rating process must uniquely reside in this part of the market. For an operating company, a low cost of capital is an advantage but not its only or predominant source of competitive advantage. For a securitization, a lower cost of capital is one of its few source of “competitive advantage.” A bank might change which mortgages are placed in a securitization if this increased the fraction of the securitization rated AAA and thus lowered its cost of capital. Whereas an auto-manufacturing firm is unlikely to close plants or close down a division solely to get a higher credit rating. The costs of altering the business to improve a credit score are higher and the benefits are (relatively) lower for an operating firm. This may be why we saw fewer defaults in this sector of the market.

¹⁴ The importance of nonlinearities in the return manager’s misstatement of information is discussed in Jensen (2001). In his examples, the incentives to misstate one’s information go away if the payoff function is linear. Thus, small changes in the reported information have only small changes in the manager’s payoff.

fuzzy function of the inputs. The line between an AA and an A rating can be kept secret, or additional sources of soft information can be included.¹⁵ In practice, this occurs as models that try to fit the ratings when a function of the firm's financial numbers don't have R^2 of 100 percent. This leads to a discussion on the limits of hard information.

E. The Role of Discretion

A cost of using hard information is it throws away any information that is not captured by the variables used in the decision algorithm. It can also remove the discretion of the agent who collects the information to the extent the information to be collected is specified and the decision automated. We have discussed how this loss of information could lead to worse decisions, but the reduction in discretion can cut both ways. Relationships are useful as a way to elicit information that is not available in the numbers (i.e. lending relationships). Relationships have additional dimensions. Relationships generate a sense of mutual obligation (reciprocity). You help me out and I want to help you out. Thus, when a loan officer is evaluating a potential loan from a long-term borrower, they can use their discretion to more accurately evaluate the borrower's current credit quality. The sense of reciprocity can also make repayment more likely. A borrower is not defaulting on an obligation to an unknown faceless financial institution, but to someone with whom they have worked for years. This is the positive side of the relationship. Loan officers can also use their discretion to put a thumb on the scale and influence a loan decision for the sake of their relationship, instead of a strictly financial decision. A number of papers have documented that loan officers do use their discretion, and in the documented cases, the discretion does not improve the

¹⁵ There also may be strategic reasons to avoid a transparent mapping between the numbers and the credit rating. The business of credit rating agencies relies on market participants being unable to replicate the ratings at lower cost than the agency. If the mapping were a direct function of the inputs (income and balance sheet) and nothing else, some clever assistant finance professor would figure out the function. This is one reason that the early credit reporting agencies released only a fraction of their information publicly in the form of a credit score. For additional fees, users could review a more complete report (Carruthers and Cohen 2001).

quality the decision.¹⁶ The challenge is one of incentives. The loan officers are not lending their own capital, but the bank's. The bank manager or shareholder must trade-off the value of the loan officer using their soft information (better quality decision) against the misaligned incentives between the loan officer and the bank. The advantage of hard information is can remove the loan officer's discretion. The relevant variables and the mapping from the variables to the decision is beyond the control of the loan officer.¹⁷

¹⁶ Brown, Schaller, Westerfeld, and Heusler (2012) find that loan officers use discretion to smooth credit, but there is limited information in discretionary changes. Desgrye, Liberti, Mosk, and Ongena (2013) provide evidence that soft information helps predict defaults over private and public information, but discretionary actions do not predict default. Gropp, Gruendl, and Guettler (2012) show that the use of discretion by loan officers does not affect the performance of the bank portfolio. Puri, Rocholl, and Steffen (2011) document the use of discretion as a widespread phenomenon inside a German savings bank but also find no evidence that loans approved based on discretion perform differently than those that do not use discretion. Cerqueiro, Degryse, and Ongena (2011) find that discretion seems to be important in the pricing of loans, but plays only a minor role in the decision to lend.

¹⁷ This turns out to be an imperfect solution when the loan officers has an incentive to manipulate the inputs, just as the borrower might. The loan officer in Berg, Puri, and Rocholl (2016) work for a bank that uses an internal credit score to evaluate loans. They show that loan officers repeatedly enter new values of the variables into the system until a loan is approved. Not only are they able to get loans approved that were originally rejected, but they also learn what the model's cut offs are and thus what is required to get a loan approved. These results suggest that even hard information decision-making algorithms which depend upon data subject to the control of either participant (local decision maker or the target of the decision) are subject to the Lucas critique

V. Evolution of Information: Hardening Soft Information

We have been implicitly assuming in the discussion thus far that information type is static or immutable. Information is either hard or soft, but it is not malleable. This has been a simplification to allow us to focus on the definition and advantages of each type of information. How immutable the information type is? We think of hard information as a numeric index. Soft information can and is converted into an index, but not without losing some of the information or context. Markets and individuals are constantly taking in soft (and hard) information and condensing it into binary decisions: whether to fund a project, sell a stock, or make a loan. Soft information is eventually condensed to a binary decision. This does not create a meaningful loss of information if the decision is the final step and does not feed into later decisions.¹⁸ What happens when information is condensed as an intermediate step, not as the final decision? The next step is to discuss the process of converting soft information into hard information: the hardening of soft information. By way of illustration, we first examine two historical examples: the development of credit ratings and how the creation of the Center for Research in Security Prices (CRSP) influenced our understanding of how equity markets work. We then return to the incentives this creates, the problems that arise, and how institutions and markets have responded.¹⁹

¹⁸ In Bikhchandani, Hirshleifer, and Welch's (1992) study of informational cascades, they model sequential decisions where agents see the (binary) decisions of prior agents but not the information upon which the information is made. This reduction (hardening) of information leads to agents ignoring their own (soft) information and following the crowd.

¹⁹ Loss of information is not only due to the effect of hardening the information. A change in the compensation structure of agents may also affect the use of information. In a controlled experiment, Agarwal and Ben-David (2016) study the impact that changing the incentive structure of loan officers to prospect new applications have on the volume of approved loans and default rates. They find that after the change, loan officers started relying more on favorable hard information and ignoring unfavorable soft information. The results highlight how specific activities, such as loan prospecting, may transform the use of hard and soft information. Another form of loss of information is due to the portability of soft information. For example, Drexler and Schoar (2016) show that when loan officers go on leave, they generate a cost to the bank since it affects the borrower-lender relationship. Since the loan officers have no incentives to voluntarily transfer the information, they find that borrowers are less likely to receive new loans from the bank.

Credit ratings originated in the United States during the nineteenth century.²⁰ Prior to this time, most trade among merchants was local. The extension of trade credit was common and merchants traditionally relied on soft information accumulated over time and through repeated personal interactions to make their credit decisions (Carruthers and Cohen 2001, 2009).²¹ The development of communication and transportation technologies made it possible to sell one's goods to a geographically much larger market. These were customers with whom merchants had no prior personal experience, and thus their traditional approach to trade credit lending was not possible. These technological shocks created demand for new sources of information about creditworthiness that did not rely on direct personal connections. This led to the formation of such firms as the Mercantile Agency, R.G. Dun, and Bradstreets in the 1840s. These firms promised precise, standardized ratings that would allow merchants to avoid extending credit to customers who were not credit worthy.

The credit rating bureaus established local offices in major cities and relied on local merchants, lawyers, or bankers as the sources of their information. The input to the process was the same soft information that had previously been the basis of credit decisions.²² The credit agencies used this information to create two credit scores which were sold to merchants: pecuniary strength (essentially net worth) and general credit (ability and willingness to repay).²³ In this way, the agencies were able to harden the soft information available to local merchants and provide it

²⁰ A precursor to Dun and Bradstreet was founded in 1841 (The Mercantile Agency) and a precursor to Standard and Poor (The History of Railroads and Canals in the United States by Henry Poor) was founded in 1860.

²¹ The authors' description of trade credit markets during this period is strikingly similar to Nocera's (1995) description of the consumer lending market of the 1950s.

²² "What went into the reports was a variable and unsystematic combination of facts and rumors about the firm, its owners, his personality, and family" (Cohen 1998).

²³ Not all of the information that the credit reporting bureaus had was released in the form of the credit ratings. For additional fees, subscribers could visit the office of the agencies to view a detailed report on a potential customer. The soft information that the credit rating bureaus had, they were either unable or unwilling to quantify and include in their reported credit scores. Interestingly, information in these reports was better at predicting bad outcomes (business failures) than the published credit ratings (Carruthers and Cohen 2001).

in a form that was useful to distant merchants. They could make lending decisions based on this number, even though they had no contact with the potential customer or the data collector. The standardization of the information in the form of the credit reports was a very early form of hard information and allowed the geographic reach of trade credit lenders to expand.

The Center for Research in Securities Prices began as a database of monthly and then daily returns on all NYSE stocks in the early sixties, stereotypic hard information. There is rarely disagreement about what a return of 4 percent means. Prior to the construction of the CRSP databases, however, there was limited knowledge about what the returns on equities were let alone what the determinants of equity returns were.²⁴ The existence of a comprehensive database containing the returns on all stocks unleashed a torrent of research into the determinants of both expected returns (i.e. factor returns) and realized returns (e.g. event studies). It was now possible to carefully document what announcements or events determined stock prices. The dependent variable is a unidimensional index of value: the stock price. The independent variables are also coded into numeric values. Initially the coding was rudimentary: dividends increased, decreased, or did not change. Over time, the independent variable used to explain stock returns in the event study became more elaborate. However, they were always quantitative simplifications of the underlying events.

²⁴ CRSP began with a question from bankers at Merrill Lynch, Pierce, Fenner and Smith. They wanted to know what the long-run return on equities was. They turned to Professor Jim Lorie at the University of Chicago, who didn't know either but was willing to find out for them (for a \$50,000 grant). The process of finding out led to the creation of the CRSP stock return database. The fact that neither the investment banks nor academic finance knew the answer to this question illustrates how far we have come in depending upon hard information such as stock returns. Professor Lorie described the state of research prior to CRSP in the Philadelphia talk in 1965 "Until recently almost all of this work was by persons who knew a great deal about the stock market and very little about statistics. While this combination of knowledge and ignorance is not so likely to be sterile as the reverse-that is, statistical sophistication coupled with ignorance of the field of application-it nevertheless failed to produce much of value." In addition to CRSP, he talks about another new dataset: Compustat (sold by the Standard Statistics Corporation) which had 60 variables from firm's income statement and balance sheet.

Although the event studies often found important determinants of stock prices, even when they focused on the individual days when seemingly large announcements were made, the fraction of cross-sectional variability that the models could explain was small. This omission could be due to daily movements driven by the trading process (market micro-structure effects) or by the inadequacy of the right-hand-side variables. There are many forces that move stock price (rumors, news accounts, different interpretation of public releases) that are not easily and accurately converted to a numeric score. The market does convert this soft information into the hard information of stock prices, but the academic models have had difficulty replicating the process.

Part of the problem is that much of the information that drives stock prices, even when publicly available, is in text form not numbers. A typical firm's 10K filing can run into hundreds of pages. Its income and balance sheet take up only half a dozen of those pages at most. However, the vast majority of studies that tried to explain the changes in equity values with firm data relied only on the accounting and macroeconomic numbers. This changed when academics started including textual information in regressions by coding the text into numeric scores. A very early first was Das and Chen (2007). They examined the effect of message board postings on the stock prices of Amazon and Yahoo. Although the algorithm is crude, it showed a potential way to incorporating the vast amount of textual data into our research. With the digital availability of text and gains in automated methods of analyzing text, there has been an increase in this kind of research. The next iteration, and arguable the paper that kicked off the revolution was Tetlock (2007) who "...quantitatively measure(d) the interactions between media and the stock market using daily content from a Wall Street Journal column." As the datasets have grown and finance researchers have become more adept at translating the text into numbers, the literature has grown significantly (Tetlock, 2010, Da, Engelberg, and Gao (2011), Engelberg and Parsons,

(2011), Dougal, Engelberg, Garcia, and Parsons, 2012). The literature has expanded our understanding how information reported in the media (news stories, columns, as user initiated searches) is impounded into prices. In each of these papers, the text is condensed into numerical indexes, which capture relevant information (given the results) but likely capture only a portion of what humans may capture.²⁵ This is the context that is a fundamental characteristic of soft information.

IV. Applications of Soft Information

We now document three applications of soft information given the stand of the literature and provide possible insights for future research.

A. Financial Crisis: Role of Incentives

Recently, the literature on soft information has been applied to study the role that incentives play in the production of risk estimates and thus risk taking during the financial crisis. The literature documents the role incentives played in distorting estimated risks in the mortgage securitization market and notes that the nature of information, whether hard or soft, is relevant to understand the subprime mortgage crisis.

Rajan, Seru, and Vig (2010, 2015) argue that the accuracy of forecasts from statistical models failed during the financial crisis due to the excessive reliance on hard information variables. These models ignored the role of incentives of lenders to collect soft information about the

²⁵ The literature began by simply counting positive and negative words, which proved to be more complicated than one would have initially guessed. The language of finance is not as simple as we think (Longhran and McDonald, 2011). For example, the sentence “The Dell Company has 100 million shares outstanding” would have been classified as an extremely positive sentence by the early dictionaries, since “company”, “share”, and “outstanding” are coded as positive words (Engelberg, 2008). Hoberg and Phillips (2010) are similar in method but are interested in a very different question. They use text-based analysis of firm’s 10-K to measure how similar firms involved in mergers are and thus predict the success of the mergers.

borrowers and, this adversely effected their ability to forecast default. Such failures are in the spirit of the Lucas critique (Lucas 1976) since these statistical models ignore the change in incentives of the individuals. As securitization increased the distance between the originator and the ultimate investor that bears the default risk, they argue that the increase in distance resulted in lenders choosing not to devote time to collect soft information about borrowers. Distance in the securitization market changed the incentives of lenders to carefully screen borrowers and their behavior.²⁶ Along this line of research, Keys, Mukherjee, Seru, and Vig (2009, 2010, and 2012) explore whether the securitization process reduced the incentives of financial intermediaries to screen borrowers. Using an ad-hoc rule of thumb in the FICO credit score had an adverse effect on the screening incentives of lenders to collect and use soft information. They argue that for those borrowers where it is relatively easier to securitize the loan, the lender has weaker incentives to carefully screen the borrower and use soft information in lending decisions. Keys, Seru, and Vig (2012) show that the incentive for lenders to process soft information critically depends on whether they have to bear the risk of the loan they originate.

Purnanandam (2011) explores the costs of an originate-to-distribute model of lending by comparing default rates on banks that originated loans to sell with banks that originated loans to keep in their portfolio. As the originated lender offloads the credit risk through securitization, the screening incentives of lenders to collect soft information decrease. Purnanandam provides

²⁶ Loutskina and Strahan (2011) explore the dark side of diversification. They argue that mortgage lenders that concentrate in few markets invest more in information collection, specifically private soft information, than diversified lenders, who concentrated their data collection efforts and based their decisions on public hard information. They conclude that geographic diversification adversely affected the ability to collect information about borrowers. Note that incentive-based theories within the firm have similar predictions. Lenders concentrating in few markets have the incentives to gather, collect, and use soft information, as opposed to more diversified lenders (as in Stein 2002; Berger, Miller, Petersen, Rajan, and Stein 2005; Mian 2006; Liberti and Mian 2009; and Agarwal and Hauswald 2010).

evidence that those lenders with high involvement in the securitization market had lower screening incentives, resulting in the origination of loans with poor soft information.²⁷

Overall, the findings of this literature complement one another and have important implications for the nature of banking regulation and securitization practices.

The use of hard and soft information provides useful inputs to regulators in regulating financial markets and institutions and provides a framework to eliminate inconsistencies of internal ratings across banks, which is crucial under Basel II and III (Firestone and Rezende 2013; Plosser and Santos 2016). Results should also make regulators pay special attention to policies that depend excessively on default models based on hard information. These models ignore the strategic behavior of lenders when it comes to reporting due to the potential loss of rents through the acquisition of private soft information (Giannetti, Liberti, and Sturgess 2016).

This also highlights the progressive shift in intermediation, as lending activities move outside the boundaries of the financial intermediaries, changing the organizational design of lending: intermediation through brokers in the mortgage market. These patterns are reported in Chernenko, Hanson, and Sunderam (2016). Part of this change can be attributed to improvements in the information environment of borrowers, such as the availability of credit registries and credit scores reducing the cost of information acquisition (Liberti, Seru, and Vig 2016).

Two potential avenues of future research may incorporate some of these ideas. For example, incorporating strategic behavior to default models may be a potential and challenging direction of future research. Similarly, exploring the information structure of loans may shed light on the

²⁷ Saengchote (2013) studies whether geographical distance, an indirect proxy for soft information, matters in the subprime mortgage market. This author finds that mortgage default rates are higher the further away the borrower is from the broker.

debate of which types of loans (hard- or soft-based loans) tend to benefit the most from securitization.

B. Organizational Design of Lending

One of the initial applications of soft information has been on the organizational design of lending and its implications for the boundaries of activities performed inside and outside financial institutions. Since Coase (1937), the idea of allocating control and decision making within organizations has been at the center in discussions of the theory of the firm. The allocation of control shapes the incentives of agents working in the organizations (i.e., in our case, loan officers in financial institutions). Grossman and Hart (1986), Hart and Moore (1990), and Hart (1995) define allocation of control from the ownership of a tangible asset. In the case of financial institutions, the critical resource or asset is intangible in nature: the access to information, especially soft information.²⁸

As financial institutions have become larger, more globalized, and more complex they face a tradeoff between benefits from economies of scale and costs from inferior organizational designs when it comes to granting loans. This led to a debate in the banking literature over whether decentralized organizational structures are better or worse than centralized structures in terms of providing the right incentives to loan officers to produce, transmit, and use soft information. The discussion has centered in understanding how informational distance between the decision-maker and loan officers shapes the nature of information acquisition and, therefore, the types of activities performed inside or outside the financial institution.

²⁸ Rajan and Zingales (1998) argue that ownership is not the only way to allocate power in an organization. Another and in some cases a better way is through access. Access is the ability to work with or use a critical resource, not necessarily a physical resource that can be owned.

Academic work in the organizational design of lending application has concentrated in answering three distinctive questions:

1. How and where do participants produce and use soft information?
2. How does soft information influence lending decisions?
3. How do incentives for the production of information and its use vary with the organizational structure of the financial intermediary?

An important branch of the banking literature shows that bank distance affects lending decisions (Petersen and Rajan 2002; Degryse and Ongena 2005; Mian 2006; DeYoung, Glennon, and Nigro 2008; Agarwal and Hauswald 2010).²⁹ The literature has interpreted this finding largely in terms of the difficulty of transmitting soft information. Despite its prominence, this interpretation is largely based on the observed correlation between loan characteristics and distance.

If the difficulty of soft information transmission is indeed behind the empirical relation between geographical or hierarchical distance and lending decisions, we would expect greater distance to lead to more delegation to the loan officers. This forces the bank organization to adapt, presumably relying more on hard information and also by delegating decisions less. Given the decreased delegation, we would then expect loan officers to produce and use less soft information in their decisions. Assessing these two predictions is typically hard, because data on delegation and information production is hard to get, however, a series of papers have tested these predictions.

²⁹ Although these papers are on geographical distance, they are different in nature. Petersen and Rajan (2002) document that distance between lenders and borrowers increased due to improvements of lender productivity. Degryse and Ongena (2005) study the effect on loan conditions of distance between borrowers, the lending bank, and all other banks in the vicinity. Mian (2006) suggests that greater distance not only decreases the incentives of a loan officer to collect soft information, but also makes it more costly to produce and communicate soft information. DeYoung, Glennon, and Nigro (2008) document the relationship between the use of hard information using credit scoring technologies and increases in borrower-lender distances. Finally, Agarwal and Hauswald (2010) study the effects of distance on the acquisition and use of private information in informationally opaque credit markets. They show that borrower proximity facilitates the collection of soft information, which is reflected in the bank's internal credit assessment.

Aghion and Tirole (1997) and Stein (2002) argue that large hierarchical organizational structures inhibit the ex-ante incentives to collect and use soft information. This decrease in incentives occurs because those in charge of collecting non-verifiable soft information cannot act on it, and instead have to send the information to their superiors. Given the soft nature of the information, there is a chance that the information may be overruled or ignored, thus generating a negative impact on the incentives to collect soft information. Along these lines, Liberti (2016) provides support for the loan officers' incentives view by showing that loan officers who receive relatively more formal authority as opposed to real authority put more effort into producing and using soft information. The results of this paper contrast with the implications of bank consolidation. Consolidation of financial institutions may have a negative impact on small business lending due to the potential loss of information. The results highlight how a large bank may be able to replicate the organizational structure of a small bank by delegating decision-making authority to the lower layers of the organizations. Agarwal and Hauswald (2016) provide direct evidence that the findings on distance and loan characteristics in the sizable existing literature are really due to the difficulty of transmitting soft information. In other words, they provide evidence that a bank endogenously responds to information transmission problems by effectively delegating more authority to loan officers. Skrastins and Vig (2016) also find evidence that increasing the hierarchical structure of a branch decreases the ability of the loan officers to produce soft information, leading to an increased standardization of the information collected for each loan.

One way to categorize these theories is under the umbrella of the incentive-based theories (Aghion and Tirole 1997; Crawford and Sobel 1982; Dessein 2002), as opposed to information cost-processing theories (Garicano 2000). An interesting venue of future research is to empirically disentangle both views. The problem resides on the fact that both views may have similar

predictions. Let's take the example of more information available to a loan officer. Under the incentive view, this may dampen the cost of losing control from a superior, thus increasing delegation of tasks to the loan officers. Under the information cost-processing theories, agents endogenously arrange in a hierarchy in increasing order of their ability. If more information makes the tasks of loans officers easier, then one could also observe more delegation in the lower layers of the organization.³⁰

Another potential line of future research is to study the tradeoffs of using complex numerical algorithms summarizing all the information in a single credit score. Although it makes analysis and interpretation easier, relevant information may be lost in translation, especially soft information.³¹

The conclusions could generalize to other settings and activities that use soft information intensively. Differences in organization structures may impact the allocation of resources and the activities ultimately performed inside or outside the boundaries of the firm. It would be stimulating to study these concepts more generally in other activities where soft information is a critical input in the production process.

C. Other Applications: Implications on Resource Allocation

Based on the research in the bank lending market, distance is related to information type. Hard information can be transmitted across distance without loss of content; whereas soft information cannot. This raises the question as to which financial markets are geographically close and which are not. This analysis helps us understand what kind of information undergirds each

³⁰ Liberti, Seru, and Vig (2016) explore the effect of a change in informational environment of borrowers on the organizational design of banking. Their results can be rationalized under both views.

³¹ Using a randomized control trial, Paravisini and Schoar (2015) evaluate the adoption of credit scores in a small business lending setting. They find that credit scores have a positive impact on the productivity of credit committees.

market. The finance literature has studied distance in a variety of other economic settings and financial markets including: the municipal bond market, the venture capital market, the real estate market (Garmaise and Moskowitz, 2004), the allocation of divisions within a firm, and the impact of the organization design of conglomerates on their productivity.³² In part due to credit ratings, the corporate bond market is national or international. Even though municipal bonds (tax-exempt bonds issued by state and local government entities) are also rated, Butler (2008) finds that the underwriting market is local (80% of municipal bonds are underwritten by investment banks with a local office). Unlike commercial banks, these investment banks do not hold the securities and so do not have incentives to monitor borrowers ex-post. The local underwriters have been able to credibly communicate to investors that their soft information is valuable and thus certify the bond's quality. Local underwriters are able to sell municipal bonds for higher prices (lower yields) and this results in the strongest when the ratings signal is weakest (i.e., bonds with low ratings and unrated bonds).³³

The discussions thus far has focused on external distances, but distance inside a firm may be relevant as well. Landier, Nair, and Wulf (2007) explore how the distance between divisions and headquarters may have an impact on corporate decision-making. Managers are more likely to layoff employees or divest divisions that are more distant from headquarters. Although the authors argue this could be driven by a greater affinity for the people management interacts with most

³² Even in markets that we think are dominated by hard information and thus where distance should not matter, research has found a preference for local investments. Mutual fund managers tend to hold on to average shares of local firms since access to soft information of local firms is cheaper (Coval and Moskowitz 1999, 2001). The effect is strongest in small and highly levered firms.

³³ If the local underwriters have soft information that non-local underwriters do not have and they can thus sell the bonds at higher prices, they should be able to extract larger fees. They do not. Local underwriters charge lower fees relative to non-local underwriting suggesting local competition is limiting their pricing power.

often (close) they also find the effect is stronger in environments that rely on soft information.³⁴ The problem in this type of exercise is that the choice of locations is likely to be endogenous, making it difficult to establish causality. Giroud (2013) has a clever way of solving the endogeneity issue of the distance measure. He studies the impact on plant-level investment and productivity of headquarters located close to (or distant from) plants, arguing that travel time is a better proxy for monitoring than geographical distance.³⁵ He exploits the introduction of new airline routes as a source of exogenous variation to proximity and measures the causal impact of distance on plant-level investment and productivity. In a similar type of study, Graham, Harvey, and Puri (2015) examine how CEOs and CFOs around the world delegate financial decisions and explore the drivers behind delegation. Unsurprisingly, they find that they delegate corporate policy decisions that require additional input when they are overloaded, when they are distracted by acquisitions, and when they know less.

Different organizational structures may affect the allocation of resources across conglomerates with implications for productivity. For example, allowing divisions within a conglomerate to have decentralized R&D budgets results in more creative R&D output (Seru 2014). In particular, this paper speaks to whether research is done inside or outside the firm, and how this depends on the type of research activity and information production. While divisions may be in a better position to collect soft information than headquarters, innovation that requires more

³⁴ They use the measure of distance between banks and borrowers from Petersen and Rajan (2002) to classify whether industries are hard- or soft-information intensive. Industries where distance between borrowers and lenders is large are classified as hard information environments.

³⁵ A plant may be located far away in terms of geographical distance, but monitoring may be easy since there are direct flights between the cities where the headquarters and plants are located.

decentralization and independence may be completed only outside the boundaries of the firm (i.e., strategic alliances, joint ventures).

D. New Financial Markets

Finally, we provide an example of a new market where the extent and nature of information seems relevant. Peer-to-peer lending is a natural market that provides an ideal setting to quantify the relative importance of hard relative to soft information since lenders have access to both types of information. In this market, lenders are able to infer the probability of the borrower from less quantifiable information, such as a picture of the borrower, personal characteristics, and the actual reason for a loan application (Ravina 2012). Although, on average, hard and soft information matters, Iyer, Khwaja, Luttmer, and Shue (2015) provide evidence that soft information is relatively more important when evaluating the credit decisions of lower quality borrowers. These authors are able to quantify the magnitude of different sources of information by market participants that are generally not experts in the field. One interesting aspect of peer-to-peer lending is that borrowers, who generally know each other off-line, can group themselves. These potential social network effects allow groups of people to endorse one another to potential bidders providing soft information about borrowers.³⁶ Furthermore, when any member of the group defaults, they all suffer the reputational consequences. This is in itself an interesting market since the concept of soft information emerges independently of the geographical distance between borrowers and lenders. The results from this work emphasize the screening ability of alternative markets for small borrowers, highlighting the importance of the use of soft information.

VII. Concluding Remarks

³⁶ Freedman and Jin (2010) and Lin, Viswanathan, and Prabhala (2013) find evidence that social networks alleviate informational frictions in online lending markets.

Production of information lies at the heart of theory of the firm. In this paper, we start by defining the main characteristics of soft information and hard information and explore the possible consequences of hardening information on individual decision-making in financial markets.

The production and use of soft information has been extensively applied to financial intermediation by studying the importance of building bank-borrower relationships. Theoretical and empirical work highlighted the problems in communicating and using soft information across hierarchical layers in an organization since decisions are seldom based on soft information. Accordingly, soft information can be used to influence decisions in flat organizations, therefore, impact the allocation of resources. With this in mind, in recent years, the soft information literature has progressed beyond the exclusive application on financial intermediation to other settings as we elaborate in the paper.

The nature of the activity determines the nature of the organization and its organizational form. The type of organizational structure plays a crucial role in allocating tasks and activities across individuals in an organization making more efficient the process of collecting soft information. In activities where the critical resource is information which is subjective in nature, such as consulting, financial advisory services, investment banking, lawyers and research and development divisions, among others, small divisions or decentralized structures will have a comparative advantage in collecting soft information relative to large divisions or centralized structure. The key is the access to this information. Access may be described as access to other people in the organization, access to other valuable resources or access to outside information. The principle is to organize information production such that those making decisions are close to the source of information. At the extreme, the scope and magnitude of this information as a critical

resource in the production process will impact whether activities are performed inside or outside the boundaries of an organization.

We hope that the literature on soft information finds more applications in other markets and generalize to a variety of other settings. Exploring how large organizations adapt their internal organization structure in order to mimic small organizations will deepen the understanding of ways of organizing production and clarify how rent- or power-seeking activities generate organizational inefficiencies. We leave this for future work.

References

- Agarwal, S., and Itzhak Ben-David, 2016, “Loan Prospecting and the Loss of the Soft Information,” Working Paper, Fisher College of Business, Ohio State University.
- Agarwal, S., and Robert Hauswald, 2016, “Authority and Information,” Working Paper, National University of Singapore.
- Agarwal, S., and Robert Hauswald, 2010, “Distance and Private Information in Lending,” *Review of Financial Studies* 23:2757–2788.
- Agarwal, Sumit, Sudip Gupta, and Ryan Israelsen, 2016 “Public and Private Information: Firm Disclosure, SEC Letters, and the JOBS Act” Working paper, Georgetown University.
- Aghion, P., and Jean Tirole, 1997, “Formal and Real Authority in Organizations,” *Journal of Political Economy* 105(1):1–29.
- Ben-David, Itzhak, 2011, “Financial Constraints and Inflated Home Prices during the Real-Estate Boom”, *American Economic Journal: Applied Economics* 3(3): 55–78.
- Benmelech, Efraim and Jennifer Dlugosz, 2009, “The Credit Rating Crisis” *NBER Macro Annual 2009*, 161-207.
- Benmelech, Efraim and Jennifer Dlugosz, 2009, “The Alchemy of CDO Credit Ratings,” *Journal of Monetary Economics*, 56(5): 617-634.
- Berg, T., Manju Puri, and Jörg Rocholl, 2015, “Loan Officer Incentives and the Limits of Hard Information,” Working Paper, Duke University.
- Berger, Allen, Nathan Miller, Mitchell Petersen, Raghuram Rajan, and Jeremy Stein, 2005, “Does Function Follow Organizational Form? Evidence from the Lending Practices of Large and Small Banks,” *Journal of Financial Economics* 76(2):237–269.
- Berger, Allan and Lamont Black, 2011, “Bank size, lending technologies, and small business finance,” *Journal of Banking and Finance* 35(3): 724-735.
- Bikhchandani, S, David Hirshleifer, and Ivo Welch, 1992, “A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades” *Journal of Political Economy* 100(5):992-1026
- Brown, M., Matthias Schaller, Simone Westerfeld, and Markus Heusler, 2012, “Information or Insurance? On the Role of Loan Officer Discretion in Credit Assessment,” Working Paper, University of St. Gallen.
- Butler, Alex, 2008, “Distance Still Matters: Evidence from Municipal Bond Underwriting,” *Review of Financial Studies* 21(2):763–784.

- Canales, R., and Ramana Nanda, 2012, "A Darker Side to Decentralized Banks: Market Power and Credit Rationing in SME Lending," *Journal of Financial Economics* 105(2):353–366.
- Carruthers, B., and Barry Cohen, 2001, "Predicting Failure but Failing to Predict: A Sociology of Knowledge of Credit Rating in Post-bellum America," Working Paper, Department of Sociology, Northwestern University.
- Carruthers, B., and Barry Cohen, 2009, "Credit, Classification and Cognition: Credit Raters in 19th-Century America," Working Paper, Department of Sociology, Northwestern University.
- Cerqueiro, G., Hans Degryse, and Steven Ongena, 2011, "Rule versus Discretion in Loan Rate Setting," *Journal of Financial Intermediation* 20(4):503–529.
- Chernenko, S., Sam Hanson, and Adi Sunderam, 2016, "Who Neglects Risk? Investor Experience and the Credit Boom," *Journal of Financial Economics*, forthcoming.
- Coase, R., 1937, "The Nature of the Firm," *Economica* 4(16):386–405.
- Cohen, B., 1998, "Marketing Trust: Credit Reporting and Credit Rating in the 19th Century," Working Paper, Northwestern University.
- Coval, J. D., and Tobias J. Moskowitz, 1999, "Home Bias at Home: Local Equity Preference in Domestic Portfolios," *Journal of Finance* 54(6):2045–2074.
- Coval, J. D., and Tobias J. Moskowitz, 2001, "The Geography of Investment: Informed Trading and Asset Prices," *Journal of Political Economy* 109(4):811–841.
- Crane, D., and Robert Eccles, 1988, *Doing Deals: Investment Banks at Work*, Boston, MA: Harvard Business School Press.
- Crawford, V. P., and Joel Sobel, 1982, "Strategic Information Transmission," *Econometrica* 50(6):1431–1451.
- Da, Zhi, Joseph Engelberg, and Pengjie Gao, 2011, "In Search of Attention" *Journal of Finance* 66(5): 1461–1499.
- Das, S., and Mike Y. Chen, 2007, "Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web," *Management Science* 53(9):1375–1388.
- Degryse, H., Luc Laeven, and Steven Ongena, 2009, "The Impact of Organizational Structure and Lending Technology on Banking Competition," *Review of Finance* 13(2):225–259.
- Degryse, H., and Steven Ongena, 2005, "Distance, Lending Relationships, and Competition," *Journal of Finance* 60(1):231–266.

- Degryse, H., José M. Liberti, Thomas Mosk, and Steven Ongena, 2013, “Is Loan Officer Discretion Advised When Viewing Soft Information?” Working Paper, Northwestern University.
- Dessein, W., 2002, “Authority and Communication in Organizations,” *Review of Economic Studies* 69:811–838.
- Diamond, D., 1984, “Financial intermediation and Delegated Monitoring,” *Review of Economic Studies* 51(3):393–414.
- Diamond, D., 1991, “Monitoring and Reputation: The Choice between Bank Loans and Directly Placed Debt,” *Journal of Political Economy* 99(4):689–721.
- Dougal, C., Joseph Engelberg, Diego García, and Christopher Parsons, 2012, “Journalists and the Stock Market,” *Review of Financial Studies* 25(3):639–679.
- DeYoung, R., Dennis Glennon, and Peter Nigro, 2008, “Borrower-Lender Distance, Credit Scoring, and Loan Performance: Evidence from Informational-Opaque Small Business Borrowers,” *Journal of Financial Intermediation* 17(1):113–143.
- Drexler, A., and Antoinette Schoar, “Do Relationships Matter? Evidence from Loan Officer Turnover,” *Management Science*, forthcoming.
- Engelberg, Joseph, and Christopher Parsons, 2011, “The Causal Impact of Media in Financial Markets” *Journal of Finance* 66(1): 67–97.
- Firestone, S., and Marcelo Rezende, 2013, “Are Banks’ Internal Risk Parameters Consistent? Evidence from Syndicated Loans,” Federal Reserve Board, Working Paper 2013-84.
- Freedman, S., and Ginger Zhe Jin, 2011, “Learning by Doing with Asymmetric Information: Evidence from Prosper.com,” NBER Working Paper No. 16855.
- Friedman, Milton, 1990, *Free to Choose*, New York: Harcourt, Inc.
- Garicano, L., 2000, “Hierarchies and the Organization of Knowledge in Production,” *Journal of Political Economy* 108(5):874–904.
- Garmaise Mark, and Toby Moskowitz, 2004 “Confronting Information Asymmetries: Evidence from Real Estate Markets”, *Review of Financial Studies*, 17(2) 405-37.
- Giannetti, M., Jose Liberti, and Jason Sturgess, 2016, “Information Sharing and Rating Manipulation,” Working Paper, Stockholm School of Economics.
- Giroud, X., 2013, “Proximity and Investment: Evidence for Plant-Level Data,” *Quarterly Journal of Economics* 128(2):861–915.

Graham, J. R., Campbell R. Harvey, and Manju Puri, 2015, “Capital Allocation and Delegation of Decision-Making Authority within Firms,” *Journal of Financial Economics* 115(3):449–470.

Grinblatt, Mark and M. Keloharju, 2001, Distance, Language, and Culture Bias: The Role of Investor Sophistication, *Journal of Finance* 56, 1053-1073.

Gropp, R., Christian Gruendl, and Andrew Guettler, 2012, “Does Discretion in Lending Increase Bank Risk? Borrower Self Selection and Loan Officer Capture Effects,” Working Paper, EBS Business School.

Grossman, S. J., and Oliver D. Hart, 1986, “The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration,” *Journal of Political Economy* 94(4):691–719.

Hart, Oliver, 1995, *Firms, Contracts, and Financial Structure*, Oxford: Clarendon Press.

Hart, O. and John Moore, 1990, “Property Rights and the Nature of the Firm,” *Journal of Political Economy* 98(6):1119–1158.

Hauswald, R., and Robert Marquez, 2006, “Competition and Strategic Information Acquisition in Credit Markets,” *Review of Financial Studies* 19(3):967–1000.

Hoberg, Gerard, and Gordon Phillips, 2010, “Product Market Synergies and Competition in Mergers and Acquisitions: A Text-Based Analysis” *Review of Financial Studies* 23(10): 3773-3811.

Iyer, R., Asim Ijaz Khwaja, Erzo F. P. Luttmer, and Kelly Shue, 2015, “Screening Peers Softly: Inferring the Quality of Small Borrowers,” *Management Science* 62(6):1554–1577.

Jensen, M., 2003, “Paying People to Lie: The Truth about the Budgeting Process,” *European Financial Management* 9(3):379–406.

Keys, B., Tanmoy Mukherjee, Amit Seru, and Vikrant Vig, 2009, “Financial Regulation and Securitization: Evidence from Subprime Loans,” *Journal of Monetary Economics* 56(5):700–720.

Keys, B., Tanmoy Mukherjee, Amit Seru, and Vikrant Vig, 2010, “Did Securitization Lead to Lax Screening? Evidence from Subprime Loans,” *Quarterly Journal of Economics* 125(1):307–362.

Keys, B., Tanmoy Mukherjee, Amit Seru, and Vikrant Vig, 2012, “620 FICO, Take II: Securitization and Screening in the Subprime Mortgage Market,” Working Paper, Booth School of Business, University of Chicago.

Keys, B., Amit Seru, and Vikrant Vig, 2012, “Lender Screening and Role of Securitization: Evidence from Prime and Subprime Mortgage Markets,” *Review of Financial Studies* 25(8):2071–2108.

- Landier, A., Vinay B. Nair, and Julie Wulf, 2009, “Trade-offs in Staying Close: Corporate Decision Making and Geographic Dispersion,” *Review of Financial Studies*, 22(3):1119–1148.
- Lerner, J., 1995, “Venture Capitals and the Oversight of Private Firms,” *Journal of Finance* 50(1):301–318.
- Liberti, J., 2016, “Initiative, Incentives and Soft Information,” *Management Science*, forthcoming.
- Liberti, J., and Atif Mian, 2009, “Estimating the Effect of Hierarchies on Information Use,” *Review of Financial Studies* 22(10):4057–4090.
- Liberti, J., Amit Seru, and Vikrant Vig, 2016, “Information, Credit and Organization,” Working Paper, University of Chicago Booth School of Business.
- Lin, Mingfeng, Siva Viswanathan, and N. R. Prabhala, 2013, “Judging Borrowers by the Company They Keep: Friendship Networks and Information Asymmetry in Online Peer-to-Peer Lending,” *Management Science* 59(1):17–35.
- Longhran, Tim and Bill McDonald, 2011, “When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks” *Journal of Finance* 66(1): 35-65.
- Loutskina, E., and Philip E. Strahan, 2011, “Informed and Uninformed Investment in Housing: The Downside of Diversification,” *Review of Financial Studies* 24(5):1447–1480.
- Lucas, R., 1976, “Econometric Policy Evaluation: A Critique,” in K. Bruner and A. A. H. Meltzer, eds., *The Phillips Curve and Labor Markets, Carnegie-Rochester Conferences on Public Policy*, Amsterdam: North Holland Press.
- Mian, A., 2006, “Distance Constraints: The Limits of Foreign Lending in Poor Economies,” *Journal of Finance* 61(3):1465–1505.
- Morris, S., and Hyun Song Shin, 2002, “Social Value of Public Information,” *American Economic Review* 92(5):1521–1534.
- Nocera, Joseph, 2013, *A Piece of the Action: How the Middle Class Joined the Money Class*, Simon & Schuster, Reissue Edition.
- Paravisini, D., and Antoinette Schoar, 2015, “The Incentive Effects of Scores: Randomized Evidence from Credit Committees,” Working Paper, London School of Economics.
- Petersen, M. A., and Raghuram Rajan, 1994, “The Benefits of Firm-Creditor Relationships: Evidence from Small-Business Data,” *Journal of Finance* 49(1):3–37.
- Petersen, M. A., and Raghuram Rajan, 2002, “Does Distance Still Matter? The Information Revolution in Small Business Lending,” *Journal of Finance* 57(6):2533–2570.

- Puri, M., Jörg Rocholl, and Sascha Steffen, 2011, "Rules versus discretion in Bank Lending Decisions," Working Paper, Duke University.
- Purnanandam, A., 2011, "Originate-to-Distribute Model and the Subprime Mortgage Crisis," *Review of Financial Studies* 24(6):1881–1915.
- Qian, J., Philip E. Strahan, and Zhishu Yang, 2015, "The Impact of Incentives and Communication Costs on Information Production and Use: Evidence from Bank Lending," *Journal of Finance* 70(4):1457–1493.
- Rajan, Raghuram and Luigi Zingales, 1998 "Power in a Theory of the Firm," *Quarterly Journal of Economics*, 113: 387-432.
- Rajan, U., Amit Seru, and Vikrant Vig, 2010, "Statistical Default Models and Incentives," *American Economic Review (Papers and Proceedings)* 100(2):506–510.
- Rajan, U., Amit Seru, and Vikrant Vig, 2015, "The Failure of Models That Predict Failure: Distance, Incentives and Defaults," *Journal of Financial Economics* 155(2):237–260.
- Ramakrishnan, R. T. S., and Anjan Thakor, 1984, "Information Reliability and a Theory of Financial Intermediation," *Review of Economic Studies* 51(3):415–432.
- Ravina, E., 2012, "Love & Loans: The Effect of Beauty and Personal Characteristics in Credit Markets," Working Paper, Columbia Business School.
- Saengchote, Kanis, 2013, "Soft Information in the Subprime Mortgage Market," Working Paper, Kellogg School of Management, Northwestern University.
- Seru, A., 2014, "Firm Boundaries Matter: Evidence from Conglomerates and R&D Activity," *Journal of Financial Economics* 111(2):381–405.
- Skrastins, J., and Vikrant Vig, 2016, "How Organizational Hierarchy Affects Information Production?" Working Paper, London Business School.
- Stein, J., 2002, "Information Production and Capital Allocation: Decentralized versus Hierarchical Firms," *Journal of Finance* 57(5):1891–1921.
- Tetlock, P. C., 2007, "Giving Content to Investor Sentiment: The Role of Media in the Stock Market," *Journal of Finance* 62(3):1139–1168.
- Tetlock, P. C., 2010, "Does Public Information News Resolve Asymmetric Information," *Review of Financial Studies* 23(9):3520–3557.
- Uzzi, B., and Ryon Lancaster, 2003, "Relational Embeddedness and Learning: The Case of Bank Loan Managers and Their Clients," *Management Science* 49(4):383–399.

Exhibit A: Assessment Criteria of Subjective Information Measures

BUSINESS RISK ASSESSMENT							
1 Industry		RR1-RR2	RR3	RR4	RR5	RR6	RR7
Trend in Output		Very Strong Growth	Strong Growth	Growth	Stable	Uncertain / Declining	Declining
Trend in Earnings		Very Strong Growth	Strong Growth	Growth	Stable	Uncertain / Declining	Declining
Cyclicality (Fluctuations)		Very Stable	Very Limited	Small	Moderate	Large	Large & Unpredictable
External Risks		No Risks	Few Risks, Non Cyclical	Few Critical Risks	Varios Critical Risks	Numerous Critical Risks	Widespread Risks
2 Competitive Position		RR1-RR2	RR3	RR4	RR5	RR6	RR7
Market Position		Over 50% / Clearly Dominant	Over 20% / Dominant	Over 10% / Major Player or Strong Niche	Over 5% / Known Player or Established Niche	2 to 3% / Minos Player	Below 2% / Minor Player; Declining Share
Product Line Diversity		Over 3 Growing Lines	Over 3 Lines	At least 2 Growing Lines	At least 2 Stable Lines	Only 1 Stable Line	Only 1 Declining Line
Operating Cost Advantage		Global Leader	Achieves Low Global Costs	Has Lowest Local Costs	Some Cost Advantages	No Cost Advantages	High Cost Producer
Technology Advantage		Global Leader in Many Areas	Global Player in Some Areas	Leader in Local Market	Mostly New; Upgrading Old	Technology Follower	Predominantly Outdated
Key Success Factors		Global Capabilities in All Factors	Global Capabilities in Most Factors	Strong Locally in All Factors	Strong Locally in Some Factors	Strong in Some; Weak in Others	None
3 Management		RR1-RR2	RR3	RR4	RR5	RR6	RR7
Professionalism		At all Levels With Extensive Experience	At all Levels in Operations & Management	At all Key Positions in Operations & Management	At Most Key Positions & Most Levels	At Some Key Positions	In Few Positions
Systems and Controls		Meets Highest Global Standards	Meets Highest Local Standards	Very Reliable and Strong	Acceptable	Unreliable	Largely Absent
Financial Disclosure		Meets Highest Global Standards	Always Timely and Accurate	Usually Timely and Accurate	Satisfactory Reporting	Delayed, Inaccu-rate or Incomplete	Unreliable
Ability to Act Decisively		Proven to be Very Strong	Proven to be Strong	Good, but Untested	Good, but Untested	Weak	Hopeless
Risk Management Policies		RR1-RR2	RR3	RR4	RR5	RR6	RR7
Leverage Policy		Extremely Conservative	Very Conservative	Low Tolerance	Some Tolerance	High Tolerance	Unlimited Appetite
Liquidity Policy		Extremely Conservative Cushion	Conservative Cushion & Contingency Plan	Some Cushion & Sound Contingency Plan	Maintains Some Cushion	Low Liquidity Acceptable	No Policy
Hedging Policy		All Risks Understood; No Open Positions	Most Risks Understood; No Open Positions	Most Risks Understood; Few Open Positions	Risks Understood but Not Always Covered	Risks Understood but Most Not Covered	No Hedging Policy / Speculative Policy
4 Access to Capital		RR1-RR2	RR3	RR4	RR5	RR6	RR7
Capital Markets		Wide Access; Domestic & International	Wide Access; Domestic & International	Primarily Domestic; Some International	Primarily Domes-tic Banking; Some Capital Markets	Limited Largely to Domestic Banking	No access to Capital markets
Banks		Established Relationships; Strong Commitments	Established Relationships; Strong Commitments	At Least One Bank Strongly Committed	At Least One Bank Strongly Committed	No Bank Strongly Committed or Some Banks Getting Out	Bank Cutting Lines; Some Locked-in
Overall Business Rating							
(Do not use +/- in the final Business Rating)							