Relationship Building*

Marina Halac†

Columbia University

March 2010

Abstract

The question of how to develop a relationship is central to business and management. This is especially true when the environment is characterized by informational asymmetries and subjectivity, as for example in management consulting. This paper presents a model of relationship building inspired by features of the consultant-client relationship. Consistent with the evidence, it shows that consultants and clients optimally start with low-risk, low-return assignments, and move up to high-risk, high-return assignments over time as they accumulate relationship capital. The probability of breakup is decreasing over the course of the relationship, but may jump when a higher-risk assignment is adopted.

---

*I am grateful to Ben Hermalin and Steve Tadelis for insightful discussions and guidance. I also thank Pierre Azoulay, Patrick Bolton, Wouter Dessein, Bob Gibbons, Andrei Hagiu, Botond Kőszegi, Glenn MacDonald, Michael Ryall, Olav Sorenson, seminar participants at Harvard-MIT and WashU, and participants at the 2009 HBS Strategy Research Conference and the 2009 ISNIE conference for helpful comments.

†Graduate School of Business, Columbia University, 3022 Broadway, Uris Hall 616, New York, NY 10027. E-mail: mhalac@columbia.edu.

1 Introduction

The question of how to develop a relationship or reputation is central to business and management. This is especially true when the environment is characterized by informational asymmetries and subjectivity, as for example in management consulting. Consulting guides and manuals stress the importance of creating solid consultant-client relationships, building up "relationship capital," and "developing trust."[1] Consulting firms such as McKinsey & Company and the Boston Consulting Group place building enduring relationships with clients among their core values.[2]

Yet, little is known about relationship building in economic theory. The theory of repeated games or relational contracts has traditionally focused on how ongoing relationships can sustain cooperation, rather than on how these relationships develop.[3] There are some notable exceptions, including Sobel (1985), Ghosh and Ray (1996), Kranton (1996), and Watson (1999, 2002), but these articles study the building of a relationship as resulting from the parties' inability to observe whether the other party is of the "cooperative" or "non-cooperative" type.[4] Instead, in some settings, other elements may be more important to determine how parties interact. The relationship between consultants and clients, for instance, is largely shaped by the informational asymmetries they face about the nature of the client's problem and the efforts made by each party to collaborate in the consulting assignment. In this kind of settings, there may not be inherently non-cooperative types, but parties trade in a complex environment where cooperation is hard to assess and thus incentives to behave opportunistically may arise.

This paper aims to shed light on how relationships optimally grow in such settings. I start

[1] See Biswas and Twitchell (2002), Kubr (2002), and Stroh and Johnson (2006).

[2] See the companies' websites.

[3] Recently it has been pointed out by Gibbons (2006, p. 4) that "not only are building and changing equilibria hard problems in the real world, but we have few existing tools for exploring these problems."

[4] See also Halac (2009).

by exploring, in Section 2, real-world relationships in the management consulting industry. Two facts seem prominent in this industry. First, the evolution of the consultant-client relationship is typically accompanied by changes in the nature of the assignments. In particular, consultants and clients start with small and relatively safe tasks, for which successful outcomes are close to certain but expected returns not very high. As the relationship develops, they gradually move to larger, more specialized, and riskier tasks, for which successful outcomes are less sure but expected returns higher. An example is offered by the consulting firm Barakat & Chamberlain: "When we started, we were willing to do the $50,000 studies which the bigger firms were unwilling to do. We started with more *analytical* work and then we moved up to more *management* stuff. We first did most of our work with supervisors, then managers, and now with vice presidents and presidents."[5]

The second salient fact is that disputes between consultants and clients are not rare, and in some cases they result in a breakup of the relationship. Consultants and clients tend to disagree on the magnitude of the client's problem and the amount of consulting services that are needed to solve this problem. A conflict may then arise, for instance, when the consultant asks for more time (and fees) than initially arranged in order to complete an assignment, or tries to charge the client for expenses that the client finds unnecessary. As the examples in Section 2 show, these disputes sometimes end with termination of the contract and cause the dissolution of the consultant-client relationship.

In Section 3, I attempt to theoretically uncover what might be behind these facts. I develop a formal model that explains the gradual building and occasional breakup of relationships. The assumptions of the model are based on what seem to be the most important aspects of the consultant-client relationship, as described in Section 2. Specifically, I consider two risk-neutral parties, such

[5]The quote is from the firm's cofounder, Samir Barakat. See Bhide (2000, p. 86). The italics are added.

as a consultant and a client, who can trade for infinitely many periods. Every time trade takes place, the parties make relationship-specific investments; hence, repeat business is valuable. The terms or conditions of trade, however, cannot be objectively defined. In particular, in any given period, the difficulty of the client's problem, and thus the cost of consulting services, may be the consultant's private information—the consultant may be more knowledgeable, or her evaluation of difficulty subjective—and the client's decision to exert costly, difficulty-reducing effort is the client's private information. The relationship is thus subject to both private monitoring and moral hazard.

The parties structure their relationship by choosing the type of trade or assignment in which they engage in a given period. High-risk, high-return assignments produce a larger per-period joint surplus than low-risk, low-return assignments when the difficulty of the assignment turns out to be minor (the good state). The probability that difficulty is minor (rather than serious), however, is higher for low-risk, low-return assignments. In expectation, high-risk, high-return assignments generate a larger per-period surplus. The incentive and monitoring problems, as well as the pace with which the parties accumulate relationship capital over the course of their relationship, are the same for both types of assignments.

Two benchmark settings are analyzed in Section 4. First, I consider a setting with common monitoring, where the client is always able to observe the true difficulty of the assignment. This can be interpreted as the client having enough expertise to assess the assignment's difficulty every period, or having a subjective perception of difficulty that is perfectly correlated with the consultant's perception. Second, I consider a setting with no moral hazard, where the client's effort is either observable or costless. I show that, in any of these settings, the parties optimally choose the high-risk, high-return assignment in all periods in which they trade, so the question of relationship building is rather trivial, and paths as those followed by the consultant-client relationship cannot be replicated. Furthermore, in these settings, the relationship is never terminated in equilibrium.

The results change when both private monitoring and moral hazard are present. As discussed in Section 5, in this case, an optimal contract must simultaneously induce the consultant to report the difficulty of the assignment truthfully and the client to exert difficulty-reducing effort. In such a contract, the relationship is terminated if the client finds the consultant cheating (that is, misreporting difficulty), which never occurs in equilibrium. However, this threat of termination off the equilibrium path is not sufficient to sustain trade when the value of the relationship is low and the monitoring and incentive problems are severe. Intuitively, this punishment for cheating is low if the client is unlikely to detect cheating and the relationship is not very valuable, while the consultant's temptation to cheat is high if high-powered incentives are given. Consequently, an optimal contract must involve inefficient termination in equilibrium: the client ends the relationship with positive probability when the consultant claims a serious difficulty and the client cannot assess the true difficulty of the assignment.

The optimal type of assignment is then determined by a risk-versus-return tradeoff. High-risk, high-return assignments generate a higher per-period expected return, but, because they are more likely to be seriously difficult and thus to put the parties in a situation where cooperation is ambiguous, they also entail a higher risk of inefficient termination in equilibrium. On the other hand, low-risk, low-return assignments make cooperation easier to assess and thus reduce the risk of inefficient termination, but at the cost of generating a relatively low per-period expected return.

The model predicts that the parties typically choose the low-risk, low-return assignment in the early stages of the relationship, and move to the high-risk, high-return assignment in later stages. This path follows from the fact that the value of the relationship is low when the parties start trading, but increases over the course of the relationship as the parties accumulate relationship capital. Hence, initially, the threat of termination off the equilibrium path is low, and the probability of termination following a serious-difficulty report in equilibrium must be high. The low-risk, low-return assignment is then optimal because it minimizes the risk of a serious-difficulty realization. Over time, however, the threat of termination off the equilibrium path becomes more effective, so this risk becomes less relevant relative to the possibility of generating a large per-period expected surplus. It is then when the high-risk, high-return assignment becomes more attractive.

The model also shows that costly disputes between consultants and clients arise and parties end their relationship with positive probability in equilibrium. Furthermore, the results indicate that, unlike in models where relationship building results from asymmetric information about the parties' types, here the probability of breakup is decreasing over the course of the relationship given a type of assignment.[6] At the time the parties move up to a riskier assignment, the probability of breakup may jump.

While the model is inspired by the features and problems of the consultant-client relationship, it may also help understand how other real-world relationships are developed. In Section 6, I consider other applications that are consistent with the model's assumptions and predictions, and discuss how the framework can be extended to study richer environments. This section also provides a brief review of the related literature. Formal proofs are contained in the Appendix.

[6] Although I have not found empirical evidence on how the probability of termination evolves over the course of the consultant-client relationship, the results are in line with empirical studies on other industries. Zylbersztajn and Lazzarini (2005), for example, analyze technology licensing contracts between seed companies and a governmental R&D organization in Brazil, and find that rates of contract termination decrease as a function of past satisfactory outcomes.

2 The consultant-client relationship

IBISWorld Inc. estimates that, in 2008, there were over 465,000 establishments in the U.S. management consulting industry that generated a total revenue of around $150 billion. These firms provide advice and assistance to businesses and other organizations on management issues. The major services include process/operations management, corporate strategy, IT strategy, actuarial/benefits, and organizational design. The major clients for the top 40 companies are largely from the areas of financial services, manufacturing, consumer products, and telecommunications.

Below I describe several aspects of the relationship between management consultants and their clients. I show that this relationship is characterized by repeated interaction and subject to different forms of asymmetric information. I describe the types of contracts that are more often used in the industry and the problems that they may present. Finally, I study how the nature of consulting assignments changes over the course of the consultant-client relationship, and show that disputes between consultants and clients sometimes lead to termination of the relationship.

Repeat business. The consultant-client relationship is characterized by repeated interaction. Empirical studies show that the consultant's reputation, third-party recommendations, and previous use of the consultant are the most important choice criteria used by firms to select consultants (see Dawes, Dowling and Patterson [1992], Clark [1995], and Bennett and Smith [2004]). There is also evidence that repeat business represents a large proportion of consultants' revenues, between 60 and 80 percent (Kipping 1999).

The large amount of repeat business is in part due to the fact that finding new clients is much more costly than selling to existing clients.[7] Indeed, because of this, consultants are sometimes

[7] Karl Scholz, co-founder and principal of Virtual PR Director, reports that for a medium-sized German consulting

willing to do the first assignment at a loss, particularly if valuable projects are likely to follow it.[8] Another reason for repeat business is that not only searching for, but also switching to new clients, as well as switching to new consultants, is costly. When a consultant and a client engage in an assignment, they make relationship-specific investments—collection of data, analysis of the client's goals and strategy, efforts to effectively work with each other, inter-personal relationships, etc. Thus, ceteris paribus, it is efficient for both parties to collaborate in new assignments.

Informational asymmetries. Management consulting is an industry where the parties are inherently asymmetrically informed. The quality of the consultant's service or advice is very difficult to assess for the client, both before the service has been provided and after the service took place (Nayyar 1990; Clark 1993). Not only do the parties have different degrees of knowledge about the various aspects of the client's problem, but their evaluation of the problem is also highly subjective. In addition, the consulting process is such that, even if objective benchmarks were available ex ante, complete contracts would be impractical, if not impossible, to write. Unforeseen events may occur at any point, and the parties' perceptions of what the client must do may change as the project develops.

Due to this asymmetry of information, client firms are vulnerable to opportunistic behavior and, hence, often reluctant to work with consultants.[9] Kubr (2002, p. 61) explains that "the client may have only a vague idea of how consultants work and may be slightly suspicious—possibly he or she has heard about consultants who try to complicate every issue, require more information than

firm, some 21 to 24 days have to be spent to generate a new assignment of average size from a new client, while only three to five days are needed to generate an assignment of the same size from an existing client. See Kubr (2002, p. 671).

[8]Sweeney (2001) reports that Bain & Company initially obtained clients by offering several weeks of work at no cost until proving the results of their services.

[9]Of course, the extent to which clients are vulnerable to opportunistic behavior depends on the form of the contract. This is discussed below.

they really need, ask for more time in order to justify longer assignments, and charge exorbitant fees." In fact, more than a few have criticized the profession for launching "management fads" and "stating the obvious," and complained that consultants never want to leave and constantly try to expand the length and scope of their work.[10] O'Shea and Madigan (1997), for instance, claim that, all too often, management consulting firms fail to serve the best interest of the client.

But clients are not the only ones that are exposed to opportunistic behavior. The difficulty, cost, and completion of an assignment generally depend on the consultant's as well as the client's actions. The latter include, for example, participating actively in the consulting process, making available any data that the consultant may require in a timely fashion, assigning qualified managers to work with the consultant, learning how to implement the consultant's advice, and minimizing resistance to change. According to Kubr (2002, p. 67), "collaboration allows the consultant to refrain from undertaking tasks that the client is able and willing to do, thus saving the consultant's time and reducing the cost of the assignment;" further, "without consultant-client collaboration, there is no effective consulting." Collaborating with the consultant, however, is costly for the client. Consultants thus run the risk of high costs if clients have no incentives to collaborate.

Types of contracts. The two most widely used fee-setting methods for consulting assignments are to charge a fee per unit of time (generally, one working day) and to charge a flat fee per project. In both cases, differential fees are specified for different levels of consulting staff. Another type of arrangement that is sometimes used is a retainer, under which the client pays in advance for broadly-defined consulting services to be delivered over a typically long period of time. Performance-contingent fees are more rare; indeed, consulting firms have had policies banning

[10]See Shapiro, Eccles and Soske (1993) for a list of the most common complaints about consultants.

such type of fees for many years (see Kubr [2002], chapter 30).

The way consultants should charge for their services is a controversial issue. Under fees per unit of time, consultants have incentives to prolong the assignments. In fact, as mentioned above, consultants are often criticized for extending the assignments to run up the bill. Under flat fees, on the other hand, the consultant and the client respectively take the risk that the assignment may take more or less time than initially thought. In particular, a flat fee seems inadequate for assignments whose costs are highly dependent on the client's actions. Kubr (2002, p. 688) explains that "[the consultant] cannot accept this form of fee if completion of the job depends more on the client's than on the consultant's staff. Thus, a flat fee may be charged for a market survey, a feasibility study, a new plant design or a training course, but not for a reorganization that depends much more on decisions and actions taken by the client than by the consultant."[11]

Evolution of assignments. Consulting assignments vary in a number of dimensions. I classify assignments as being either relatively low-risk and low-return, or high-risk and high-return. Although this classification is far from exact, as would be necessary to conduct a rigorous empirical analysis, it has the advantage of being applicable to a broad set of examples. Moreover, examples suggest a common pattern in the way consultants and clients build their relationships: they begin with low-risk, low-return assignments, then gradually move up to high-risk, high-return assignments.

The case of Barakat & Chamberlain cited in the introduction illustrates this pattern. This case is described by Bhide (2000) in his study of how businesses start and grow; it is, according to the

[11]The comparison between fees per unit of time and flat fees is related to Bajari and Tadelis (2001)'s analysis of procurement contracts. They provide a model that explains why complex projects in the construction industry are procured using cost-plus contracts while simple projects are procured using fixed-price contracts. Like in management consulting, these two extreme contractual forms are the most commonly used in construction. Bajari and Tadelis discuss different "nonconvexities" that can explain why extreme contracts are preferred over intermediate ones.

author, an example of the gradual or "incremental commitments" approach. Another example of this approach offered by Bhide is that of Russell Personnel Services. Russell started by helping clients find permanent secretaries, which was a low-risk task because the clients would decide themselves who to employ. With time, this allowed the company to move to the riskier and more profitable business of providing temporary secretaries: "The temp business required clients to trust Russell to screen the temps carefully; clients also saw themselves at some risk if Russell defaulted on its payments to temporary workers." (Bhide 2000, p. 86.)

More generally, the evolution of many top consulting firms seems to agree with this pattern of relationship building—they start with tasks whose outcomes are relatively certain or easy to assess, and then move to tasks that are riskier but also more profitable. For example, Deloitte started by delivering tax and audit services; in the 1970s, the company incorporated business consulting to its portfolio: "Accountants began to emphasize their abilities as business consultants—offering the full range of accounting services and actively seeking additional ways to help their clients."[12] Accenture was established primarily as a technology consultant and systems integrator; over time, as it achieved a credible track record, it began to offer management strategy solutions to its clients.[13] James O. McKinsey first developed a reputation by providing finance and budgeting services; he then founded the management consulting company that bears his name.[14] McKinsey initially served small firms and then moved to larger clients who could pay higher fees; today, McKinsey

[12] See the company's website. The quote is from the history page.

[13] See the company's website

[14] See the company's website. Although Marvin Bower, a founder of McKinsey, wanted to give up the accounting practice and focus solely on management consulting, the company was initially quite dependent on James O. McKinsey for clients (Bhide 1992). Over time, as it built a reputation in consulting, the company gradually moved away from accounting and other practices that were not part of its long-term goals. Bower describes these changes: "In the early years, as part of our upgrading of clients and types of studies, we withdrew completely from one filed (department stores) and from work measurement and job evaluation studies. In 1951, despite the good income it provided, we withdrew from the executive recruiting field, believing it to be out of character for us." (Bhide 1992, p. 16).

avoids small firms which cannot afford its fees.[15]

Small consulting firms also appear to shift to riskier, more specialized, and more profitable projects when the relationship with clients grows. This is supported by evidence on the selection criteria used by clients for different types of projects. Using a survey covering 454 small consulting firms in Britain, Bennett and Smith (2004) show that there is a significant and positive correlation between the probability that clients select previously used consultants, instead of new consultants, and the consulting assignment having relatively high fee rates (controlling for total cost, duration, and other variables). High fee rates, the authors explain, are used to differentiate "highly specialized and intensive assignments" from the rest.

Another interesting piece of evidence is provided by the accounts of how consulting firms entered foreign markets. Wright (2002) studies their expansion to Australia in the 1970s and 1980s, and stresses the considerable resistance that consultants encountered from managers, employees, and trade unions. Even large consulting firms such as McKinsey faced the problem of developing legitimacy and building a client base in an environment that was hostile to the use of consultants. But the story was not the same for all consultants. Wright (2002, p. 195) notes that "while Australian business acceptance of elite strategy consulting took time to develop, a more favourable corporate reaction greeted the emergence of the other major segment in the Australian consulting market: the Big Eight accounting firms." The revenues generated by the consulting divisions of these accounting firms, which in some cases became separate business units, increased dramatically in the 1980s. As one of the main reasons, the author points to the building of relationships with clients: "...the accounting firms had the advantage of pre-existing client relationships built upon years of tax and audit work. All the Big Eight (...) were well known to large Australian

[15] Bhide (2000, p. xi).

corporations and already had a well-developed market presence."[16]

Conflict. Given the intangible nature of consulting services and the difficulty in measuring the parties' collaboration efforts, disputes between consultants and clients are not rare. In some cases, differences of opinion end in costly conflict and breakup of the relationship. One example that was publicized in the business press is that of UOP, a company that develops petroleum and gas process technologies, and Andersen Consulting. In 1994, UOP terminated its contract with Andersen for a project they had started in 1991. The company decided to fire Andersen shortly after the consulting firm asked for more time to finish the assignment. In 1995, UOP took the case to court, accusing Andersen of bungling the project and delivering systems that "were materially defective, failed to comport with Andersen's prior representations and promises, greatly exceeded budget costs, and were delivered far beyond original target dates and in violation of project schedules."[17]

Another case that ended in litigation concerns the Boston Consulting Group and the management consulting arm of Deloitte. Both firms were hired by the industrial conglomerate Figgie International in the early 1990s to work in a major corporate transformation. In 1994, Figgie sued them for not delivering the services they had promised. According to the lawsuit, BCG had told Figgie that it would help the company realize "above average growth and maximum profitability within a three-year period of time." Figgie claimed that BCG breached its contract by providing "erroneous market studies and business reports" and billing Figgie for "unnecessary, excessive and/or inflated time and expenses." In the case of Deloitte, the consulting firm sued Figgie first,

[16]Although more tangentially related, Kipping (1999)'s study of the strategies used by American consulting firms entering Western European markets in the twentieth century is also of interest. Kipping finds that consulting firms used multinational clients as "bridges" to these foreign markets. Consultants started working with companies with which they had developed a relationship in the home country, thereby building a reputation and attracting new clients in the local markets.

[17]See *Chemical Week* (March 29, 1995, p. 9) and *International Accounting Bulletin* (April 19, 1995, p. 5).

charging that the company had not paid for Deloitte's services; Figgie countered by suing Deloitte for "failing to design and implement the conversion" of the company.[18]

3 The model

Consider a market with infinitely many sellers and buyers of a service. I call them consultants and clients, although, as discussed in Section 6, the model can be used to describe other relationships as well. A consultant (she) and a client (he) can trade at dates $t = 0, 1, ...$ The parties have the same discount factor $\delta \in (0, 1)$.[19] They can engage in two types of trade or assignments, $i \in \{\ell, h\}$. Below I characterize these as low-risk, low-return and high-risk, high-return assignments; for brevity, I refer to them as low and high assignments.

The sequence of events, shown schematically in Figure I, is as follows. At the beginning of each date t, the parties decide whether to trade, choose the type of assignment i_t, and negotiate a fee for the consultant. It will be irrelevant for the results which party chooses i; it may be reasonable to think of it as a mutually agreed decision. For the bargaining protocol, I adopt Nash bargaining with bargaining power $\lambda \in (0, 1)$ for the consultant and $1 - \lambda$ for the client. That is, the parties set the consultant's compensation to split the gains from trading with each other with shares λ and $1 - \lambda$. The disagreement point or outside option for both parties is to trade with a new party from the market or, if such trade does not yield a positive expected surplus, to engage in no trade and receive a zero expected payoff. The form of the contract is further discussed in the next sections.

Every time a consultant and a client trade, they make relationship-specific investments. They

[18] See *The Boston Consulting Group* (City Edition, August 30, 1994, p. 39) and *International Accounting Bulletin* (November 11, 1994, p. 3).

[19] More realistically, this infinitely repeated game can be interpreted as a game that ends at a random date: in any period, the probability that parties can trade the following period is exogenously given by δ.

gather information, discuss the short- and long-term goals of the client and how to achieve them, and learn how to communicate and work with each other effectively. These investments are valuable for future assignments in which the parties may collaborate; they form the "relationship capital." The stock of relationship capital at time t is denoted by k_t. For concreteness, I assume that the parties accumulate one unit of relationship capital every time they trade, so $k_t = t$ if the parties always traded with each other since $t = 0$.

The difficulty of an assignment can be minor or serious, $d_t \in \{m, s\}$. More difficult assignments require more costly consulting services. The client can exert effort $e_t \in \{0, 1\}$ at private cost $c(e_t)$ to reduce the expected difficulty of the assignment (or time needed to complete it), where $c(e) = ce$ for $c > 0$. The client's effort choice is his private information. The probability that assignment i is of minor difficulty given effort e is p_e^i, where $1 > p_1^i > p_0^i > 0$. I assume that $p_1^i - p_0^i = \Delta p$ for $i \in \{\ell, h\}$, so the incentive problem is the same for the low and high assignments. That is, the punishments and rewards necessary to induce effort by the client are the same for both types of assignments. This assumption is introduced to simplify the analysis and to consider assignments that are similar in dimensions other than their risk and return.

Over the course of the assignment, the consultant learns its difficulty d_t, and reports a difficulty $\hat{d}_t \in \{m, s\}$. Upon receiving the consultant's report, the client can observe the actual difficulty of the assignment with probability $v \in (0, 1)$ (where v stands for "*v*alidate"). The difficulty of the assignment may be the consultant's private information because she is more knowledgeable about the client's problem, or because the evaluation of the problem is subjective. In the former case, v represents the probability that the client also has the expertise and capacity to assess difficulty. In the latter, v can be interpreted as the correlation between the two parties' subjective perceptions of difficulty.[20] The difficulty of an assignment cannot be observed by third parties, so it is not verifiable.

Consistent with the above interpretation of the consultant's private information, I assume that the consultant does not observe whether the client learns the true difficulty of the assignment or not. That is, the consultant does not know whether the client has the knowledge to validate the consultant's evaluation of difficulty, or whether he agrees with this evaluation. This assumption, however, is not at all necessary for the results of the paper. In fact, as explained in Section 5, letting the consultant observe whether the client learns d_t after receiving the consultant's report has no effects on contracting.

I assume that, regardless of d and $\hat{d}$, the parties always complete the assignment they started (so the assignment's current-period output is always realized). The motivation for this assumption is twofold. First, difficulty cannot be fully assessed ex ante, before starting an assignment; it is learnt and re-evaluated during the consulting process. Thus, it may be prohibitively costly or simply unfeasible for the parties to interrupt an assignment, and part of the assignment's output may be already realized when difficulty is re-assessed. Second, relationship-specific investments made during the assignment may be lost if the assignment is not completed.[21] Hence, even if its difficulty is serious, the parties will find it beneficial to proceed with the assignment. These features can be introduced formally into the model; I refrain from doing so to keep the model simple.

The consultant incurs a cost of providing her services $q(i_t, d_t, k_t)$. The cost of consulting services is increasing in difficulty and decreasing and convex in relationship capital: $q(i, m, k) <$

[20] See MacLeod (2003) for a model of subjective evaluations with this structure.

[21] For example, the analysis and design of a marketing strategy for the client is more valuable for future assignments if implemented and tested.

$q(i, s, k)$, $\partial q(i, d, k)/\partial k \leqslant 0$, and $\partial^2 q(i, d, k)/\partial k^2 \geqslant 0$ for all i, d, k.[22,23] I define the expected cost as $\bar{q}_e(i, k) \equiv p_e^i q(i, m, k) + (1 - p_e^i) q(i, s, k)$, and assume that the decline in the expected cost caused by an increase in relationship capital is the same for both types of assignments: $\partial \bar{q}_e(\ell, k)/\partial k = \partial \bar{q}_e(h, k)/\partial k$ for all k. For example, this can be thought of as the cost of consulting services having a variable component that depends on the tasks involved in the assignment and their level of difficulty, and a fixed component that is independent of these tasks and can be reduced as relationship capital is accumulated: $q(i, d, k) = \tilde{q}(i, d) + g(k)$ for some function g with $g'(\cdot) \leqslant 0$, $g''(\cdot) \geqslant 0$. I assume that the consultant always delivers the consulting services, and he does so at the rate negotiated at the beginning of the period. The idea here is that the choice of effort by the client, the realization of the assignment's difficulty, and the provision of consulting services by the consultant, all occur not sequentially but rather simultaneously during the project.[24]

Finally, consulting services generate an output y_t for the client, and the client makes a payment to the consultant W_t. Output is $y_t = y^\ell$ if the low assignment was chosen, and $y_t = y^h$ if the high assignment was chosen. I assume that output is non-stochastic for simplicity. This assumption is not restrictive; as described in the previous section, in reality, even though the outcomes of consulting activities are stochastic, consultants' fees are generally independent of such outcomes.

As already mentioned, in any given period t, rather than jointly collaborating in an assignment, the parties can decide to trade with a new party from the market or not to trade. In addition, I

[22]The fact that the cost of consulting services declines with the stock of relationship capital relates to the idea of asset specificity as discussed in Williamson (1971, 1985). That is, while the market is initially competitive, when a consultant and a client trade with each other over time, the consultant acquires a cost advantage.

[23]One could imagine that as the parties accumulate relationship capital, the cost of effort for the client may also go down. I refrain from incorporating this effect into the main model as it would cause not only the value of trade between the parties to increase but also the moral hazard problem to decline over time. This would then make it harder to understand what drives the relationship dynamics, and to assess whether different types of assignments can be optimal at different stages even when the incentive and monitoring problems are unchanged. I briefly discuss these effects as extensions of the model in Section 5.2.

[24]Technically, this assumption rules out the possibility of hold-up, but has no effects on the qualitative results of the paper.

assume that a consultant and client can also choose to exclusively invest in relationship capital (without engaging in an assignment). As above, the parties can accumulate one unit of relationship capital per period. Investing in relationship capital in this case entails costs χ_C and χ_F for the consultant and client firm respectively, where $\chi \equiv \chi_C + \chi_F$. These costs may be interpreted as opportunity costs (the consultant and the client must spend time together to build up relationship capital) or as the costs of efforts made towards these relationship-specific investments. As will become clear in the next sections, a consultant and a client may want to exclusively engage in relationship capital accumulation in a given period only if they will collaborate in assignments in future periods. Indeed, the parties may want to do this only at the beginning of their relationship, during a period of time which I will refer to as the "observation period." A possible interpretation of this observation period is then that the parties do engage in an assignment, but need to spend some time learning how to work with each other before they can generate an output. For simplicity and to rule out the possibility of hold-up, I assume that χ_C and χ_F are contractible.

Assumption A1 below characterizes, for any assignment i, when trade is profitable. It states that the per-period expected surplus generated by the relationship is positive if and only if the client exerts effort $e = 1$ in that period. Additionally, the per-period expected surplus generated when the client exerts no effort is lower than the surplus generated when the parties decide not to engage in an assignment and only invest in relationship capital. This latter assumption implies that trading without providing incentives for effort is never beneficial.

Assumption A1. *For* $i \in \{\ell, h\}$ *and all* k*,*

A1a. $y^i - \bar{q}_1(i,k) - c > 0 > y^i - \bar{q}_0(i,k)$,

A1b. $-\chi > y^i - \bar{q}_0(i,k)$.

Next, Assumption A2 specifies how the low and high assignments differ from each other. If difficulty is serious (the bad state), the surplus generated by the relationship is the same regardless of the type of assignment. Intuitively, in this case, an assignment generates zero surplus, and the parties bear the cost of engagement. If difficulty is minor (the good state), the surplus generated with the high assignment is larger; however, the probability that this assignment is of minor difficulty is lower than for the low assignment. In expectation, given effort $e = 1$, the high assignment is more profitable. Therefore, the high assignment is a high-risk, high-return bet, and the low assignment is a low-risk, low-return bet.

Assumption A2. *For* $e \in \{0, 1\}$ *and all* k*,*

A2a. $y^\ell - q(\ell, s, k) = y^h - q(h, s, k)$,

A2b. $p_e^\ell > p_e^h$,

A2c. $y^h - \bar{q}_1(h, k) > y^\ell - \bar{q}_1(\ell, k)$.

I multiply expected lifetime payoffs by $(1 - \delta)$ to express them as a per-period average; I call these normalized payoffs or simply payoffs. Suppose that the parties engage in an assignment (either with each other or with a new party from the market) in every period $t = 0, 1, ...$ In such a case, the consultant and client's expected payoffs at time t are respectively

$$
\begin{aligned}
\pi_t &= (1 - \delta)\mathbb{E}\sum_{\tau=t}^{\infty} \delta^{\tau-t}(W_\tau - q(i_\tau, d_\tau, k_\tau)), \\
U_t &= (1 - \delta)\mathbb{E}\sum_{\tau=t}^{\infty} \delta^{\tau-t}(y_\tau - W_\tau - c(e_\tau)),
\end{aligned}
$$

and the expected joint surplus is $S_t \equiv \pi_t + U_t$.

Lastly, for the equilibrium concept, note that if a flat fee is specified, or if a reported-difficulty-contingent fee is specified but the consultant does not report difficulty honestly, then the client will not be willing to exert costly effort. But then, trade is not profitable. Therefore, for trade to occur, the contract must set differential fees for the minor and serious difficulty levels and induce truthful reporting by the consultant. Following Levin (2003), I focus on perfect public equilibria. In this class of equilibria, the consultant reports the difficulty of the assignment truthfully in every period t in which the parties trade, and hence does not keep any private information from a period to the next.[25,26] As usual, equilibrium also requires that the consultant and manager maximize their expected lifetime payoffs.

4 Benchmarks

I consider two benchmarks, one with common monitoring and another with no moral hazard. The aim of this section is to show that, in any of these cases, the parties optimally choose the high assignment in every period in which they trade, and they never end their relationship in equilibrium. Hence, the question of relationship building is rather trivial, and the path followed by relationships such as the consultant-client relationship cannot be replicated.

[25] Also, in a public equilibrium, the client's strategy is independent of the unobserved history of effort decisions. It is intuitive that the client will not condition his actions on past effort choices, as these do not affect the parties' payoffs nor continuation strategies in any way.

[26] The assumption that the consultant reports difficulty every period is not without loss of generality. As shown by Fuchs (2007), under private monitoring, a contract where evaluations are made only every $T > 0$ periods may be optimal. Such a contract, however, may not be practical in the real world, where difficulty evaluations may be necessary for the client to learn how to collaborate in the assignments more effectively. Additionally, note that, in this model, the client observes the true difficulty of the assignment with positive probability in any given period. Thus, it is not hard to show that a contract that specifies difficulty reports only every $T > 0$ periods cannot induce effort by the client with probability one in all periods.

4.1 Common monitoring

A setting with common monitoring is one in which the difficulty of the assignment is always observed by both parties; that is, $v = 1$. This difficulty may also be verifiable (observed by a third party), or not. Suppose first that difficulty is non-verifiable. Then the parties cannot write a formal contract contingent on difficulty; however, they may be able to use a simple relational contract. Under this contract, the consultant's compensation is composed of a formally-enforced fixed wage $w_t(i_t, k_t)$ and discretionary difficulty-contingent fees $f_t(i_t, d_t, k_t)$, where, without loss of generality, $f_t(i_t, s, k_t) \geqslant 0$, $f_t(i_t, m, k_t) \leqslant 0$. If $f_t > 0$, the client decides whether to honor or renege on the fee payment at the end of period t; if $f_t < 0$, the consultant makes this decision. If any of the parties reneges on a payment, it is assumed that their relationship ends with probability one. Note that in equilibrium, no party reneges, so there is no loss in assuming that default is followed by termination, which is the worst outcome (Abreu 1988).

It follows from Levin (2003) that, in this setting, conditional on the parties engaging in assignment i and having relationship capital k, the optimal contract is independent of time: in every period on the equilibrium path, $e_t = e(i, k)$, $f_t = f(i, d, k)$, and $w_t = w(i, k)$. Further, it is then straightforward to see that, if engaging in an assignment is optimal, conditional on k, the optimal type of assignment is also independent of time: $i_t = i(k) \equiv i_k$. Hence, suppose that the parties always choose to engage in an assignment and that relationship capital at time t is k. Making the choice of assignment explicit, the parties' expected payoffs are

$$\begin{aligned}
\pi(i_k, k) &= (1-\delta)\sum_{\tau=0}^{\infty}\delta^{\tau}(\bar{W}_e(i_{k+\tau}, k+\tau) - \bar{q}_e(i_{k+\tau}, k+\tau)), \\
U(i_k, k) &= (1-\delta)\sum_{\tau=0}^{\infty}\delta^{\tau}(y^{i_{k+\tau}} - \bar{W}_e(i_{k+\tau}, k+\tau) - c(e(i_{k+\tau}, k+\tau))),
\end{aligned}$$

where $\bar{W}_e(i_k,k) \equiv w(i_k,k) + p_e^{i_k} f(i_k,m,k) + (1-p_e^{i_k}) f(i_k,s,k)$ is the consultant's expected payment.

For the contract to be self-enforcing, neither party can whish to walk away:

$$\begin{aligned} -(1-\delta) f(i_k,m,k) &\leqslant \delta\big(\pi(i_{k+1},k+1) - \pi(i_0,0)\big), \\ (1-\delta) f(i_k,s,k) &\leqslant \delta\big(U(i_{k+1},k+1) - U(i_0,0)\big). \end{aligned}$$

Depending on the bargaining power distribution, the fixed wage is adjusted and slack transferred from one constraint to the other. Thus, the two conditions above can be combined into a single enforcement constraint:

$$(1-\delta)(f(i_k,s,k) - f(i_k,m,k)) \leqslant \delta\big(S(i_{k+1},k+1) - S(i_0,0)\big). \tag{E}$$

Given (E), two conditions determine the consultant's compensation: Nash bargaining and the client's incentive compatibility (IC) constraint for effort. Nash bargaining determines the expected payment, $\bar{W}_e(i_k,k)$. I consider that for new relationships (that is, when $k_t = 0$), competition drives consultants' pay down to cost levels. Thus, for a relationship capital k and an effort level $e = 1$, the consultant's expected payment is

$$\bar{W}_1(i_k,k) = \bar{q}_1(i_k,k) + \lambda\big[(y^{i_k} - \bar{q}_1(i_k,k)) - (y^{i_0} - \bar{q}_1(i_0,0))\big]. \tag{1}$$

The client's IC constraint determines the difference between the serious- and minor-difficulty fees.

For the client to have incentives to choose effort $e = 1$, this difference must be

$$f(i_k, s, k) - f(i_k, m, k) \geqslant \frac{c}{\Delta p}. \tag{IC}$$

Since this constraint puts a lower bound on $f(i_k, s, k) - f(i_k, m, k)$, it is required that (E) be satisfied when (IC) holds with equality. Therefore, a self-enforcing contract that implements effort $e = 1$ exists if and only if[27]

$$\frac{c}{\Delta p} \leqslant \frac{\delta}{1-\delta}\big(S(i_{k+1}, k+1) - S(i_0, 0)\big). \tag{2}$$

It is immediate to see that if (2) holds for some type of assignment, it holds for the high assignment. Furthermore, recall that, by Assumption A2c, the per-period expected surplus is largest under the high assignment. Hence, since it then generates a higher lifetime expected surplus, the high assignment is optimal for any relationship capital k for which engaging in an assignment is optimal.[28]

So far, I have assumed that the parties always engage in an assignment. If condition (2) holds for $k = 0$, engaging in an assignment is indeed always optimal—the parties can provide effort incentives and thus generate a positive expected surplus by collaborating in an assignment given

[27] Setting the (IC) constraint with equality and combining it with (1) gives an optimal contract with payments

$$\begin{aligned} W(i_k, m, k) &= \bar{q}_1(i_k, k) + \lambda\big[(y^{i_k} - \bar{q}_1(i_k, k)) - (y^{i_0} - \bar{q}_1(i_0, 0))\big] - \frac{(1 - p_1^{i_k})c}{\Delta p}, \\ W(i_k, s, k) &= \bar{q}_1(i_k, k) + \lambda\big[(y^{i_k} - \bar{q}_1(i_k, k)) - (y^{i_0} - \bar{q}_1(i_0, 0))\big] + \frac{p_1^{i_k} c}{\Delta p}. \end{aligned}$$

[28] This result is easily obtained due to the fact that the lower bound for $f(i_k, s, k) - f(i_k, m, k)$ given by the client's IC constraint for effort is independent of i. Clearly, the result holds under less restrictive assumptions, but the derivation is not as clean as here.

any stock of relationship capital $k \geqslant 0$. Suppose instead that condition (2) holds only if $k > \bar{k}$, for some $\bar{k} > 0$. Then for $k \leqslant \bar{k}$, the parties cannot provide effort incentives and, given Assumption A1, will prefer not to trade. In this case, the parties may want to either spend the first $\bar{k}$ periods of their relationship investing in relationship capital and then collaborate in assignments in the remaining periods, or not to trade in any period. The former will be preferred if and only if the cost of the observation period is low relative to the expected surplus that can be generated in the trade period; that is,

$$(1 - \delta^{\bar{k}+1})\chi < \delta^{\bar{k}+1} S(i_{\bar{k}+1}, \bar{k} + 1).$$

Yet, even if the relationship initially goes through an observation period, it follows directly from the analysis above that the high assignment is optimal in all periods in which the parties engage in an assignment.

Now suppose that difficulty is verifiable. Then the difficulty-contingent fees $f_t(i, s, k)$, $f_t(i, m, k)$ can be enforced formally and, thus, inducing effort $e = 1$ is always feasible. (Essentially, in this case, the available punishments for defection are unbounded, so enforceability of payments is not an issue.) Then, by Assumption A2c, engaging in the high assignment in every period is optimal.

Summarizing,

Proposition 1. *Under common monitoring, if trade is feasible, the parties choose the high assignment in every period t in which they engage in trade. Further, a difficulty-contingent (or per hour) fee is used, and a consultant and a client never end their relationship.*

4.2 No moral hazard

Consider next a setting with no moral hazard. Such a setting may be one in which the client's effort choice is observable, and either verifiable or non-verifiable, or one in which the cost of effort is zero. Suppose first that the client's effort choice is observable but non-verifiable. Then the parties may be able to use a relational contract as the one described above, but where fees are contingent on effort instead of difficulty. That is, the consultant's compensation is composed of a formally-enforced fixed wage $w_t(i_t, k_t)$ and discretionary effort-contingent fees $f_t(i_t, e_t, k_t)$, where $f_t(i_t, 0, k_t) \geqslant 0$, $f_t(i_t, 1, k_t) \leqslant 0$. Any failure to make a promised payment ends the relationship. Following the same steps as above, for a relationship capital k, trade is feasible if and only if

$$c \leqslant \frac{\delta}{1-\delta}\big(S(i_{k+1}, k+1) - S(i_0, 0)\big).$$

It is straightforward that, as above, the high assignment is feasible whenever trade is feasible, and is optimal in all periods in which trade is optimal. Additionally, if effort is observable and verifiable, then trade is always feasible and the high assignment always selected.

Finally, suppose that effort is not observable but exerting $e = 1$ is costless for the client; that is, $c = 0$. Then the parties optimally use a flat contract, where the consultant's fixed wage $w_t(i_t, k_t)$ is determined by Nash bargaining. Since $c = 0$, the client is indifferent between choosing $e = 0$ and $e = 1$ and, thus, as is standard in the literature, chooses the effort level desired by the consultant, $e = 1$. Clearly, in this case, the high assignment is also always optimal.

Summarizing,

Proposition 2. *Under no moral hazard, if trade is feasible, the parties choose the high assignment* *in every period t in which they engage in trade. Further, a flat (or per project) fee is used if $c = 0$, and a consultant and a client never end their relationship.*

5 Relationship building

5.1 Increasing the value of the relationship

Consider now the case where there is a dual informational-asymmetry problem: there is private monitoring because the difficulty of the assignment may be the consultant's private information, and there is moral hazard because the client's effort is costly and privately chosen by the client.

In this setting, if the required incentives to induce costly effort and the probability that difficulty is not observed by the client are relatively high, and the value of the relationship (the surplus generated by the relationship relative to the relationship's outside option) is relatively low, contracts as those specified in the previous section will not work. To see this, suppose first that the parties promise to pay each other fees that are contingent on the consultant's reported difficulty, and continue with their relationship as long as such fees are honored. For the contract to provide effort incentives, a serious-difficulty report must be associated with a high payment for the consultant. But then, the consultant will never report a minor difficulty—as cheating is costless, she maximizes her expected payoff by reporting a serious difficulty every period. In turn, the client will not be willing to exert costly effort, and trade will not be feasible.

Consider next a similar contract but specifying that, in any period t, if the client observes difficulty and finds that the consultant has cheated (that is, $\hat{d}_t \neq d_t$), then the relationship ends with probability one. Now cheating is not costless; however, if the private monitoring and moral hazard problems are severe enough, and the parties have not yet accumulated much relationship capital, cheating will still be profitable for the consultant. The temptation to cheat—the difference between the serious- and minor-reported-difficulty fees—will be high if c is high and Δp is low, while the punishment for cheating—the expected loss due to the possibility that the client will find the consultant cheating and end the relationship—will be low if v and $[S(i_{k+1}, k+1) - S(i_0, 0)]$ are low. Hence, the consultant will again choose to report a serious difficulty every period and, anticipating this, the client will not be willing to exert effort.

Finally, consider a flat-fee contract. If the client pays a fixed fee to the consultant, independent of reported (or actual) difficulty, then monitoring is not an issue. However, the client will have no incentives to choose effort $e = 1$ at cost c, and will thus choose $e = 0$ at zero cost every period.

In sum, in this setting, an optimal contract must make the consultant's cheating sufficiently costly to induce truthful reporting, while providing sufficiently high incentives for the client to induce costly effort. Because there is only one profitable level of effort in the model, the latter requirement is completely pinned down by the client's incentive compatibility constraint. As for the former, it is clear from the discussion above that inducing truthful reporting may entail punishing the consultant not only when the client finds her cheating, but also when the consultant reports a serious difficulty and the client does not observe the actual difficulty of the assignment. While different forms of punishments may be used, it is possible to show that if an optimal contract that induces truthful reporting in every period exists, then a termination contract—a contract that ends the relationship with positive probability following a serious-difficulty report—is optimal (see Levin [2003]). Thus, without loss of generality, I focus on termination contracts.

An optimal termination contract is as follows. The parties agree on a fixed wage $w_t(i_t, k_t)$

and reported-difficulty-contingent fees $f_t(i_t, \hat{d}_t, k_t)$ (where f_t may be greater or less than zero). If the consultant reports $\hat{d}_t = m$, or if she reports $\hat{d}_t = s$ and the client observes that $d_t = s$, the parties continue with their relationship with probability one. If the consultant reports $\hat{d}_t = s$ and the client cannot observe d_t, the parties end their relationship with probability $(1 - \phi(i_t, k_t))$ and continue with their relationship with probability $\phi(i_t, k_t)$, for some $\phi(i_t, k_t) \in [0, 1]$. Finally, if the consultant reports $\hat{d}_t = s$ but the client observes $d_t = m$, the parties end their relationship with probability one.[29]

The smaller $\phi(i_t, k_t)$, the stronger the punishment for misreporting. However, inducing truthful reporting is costly, as it requires inefficient termination of the relationship in equilibrium in periods in which difficulty is not observed. An optimal termination contract then sets the minimum probability of termination $(1 - \phi(i_t, k_t))$ such that the consultant reports difficulty truthfully. This is the probability that leaves the consultant indifferent between reporting a serious and minor difficulty when the actual difficulty of the assignment is minor (and that, therefore, makes the consultant prefer reporting a serious difficulty when the actual difficulty is serious). Note that termination of the relationship when the client observes difficulty and can validate that the consultant has cheated never occurs in equilibrium. For this reason, it is optimal to specify that the relationship ends with probability one (the worst punishment) following such event.

As in the case of common monitoring, it is possible to show that the optimal contract is independent of time conditional on the stock of relationship capital. For a relationship capital k at time t, the consultant is indifferent between reporting a serious and minor difficulty when difficulty is

[29] It is straightforward to show that specifying fees that are contingent not only on reported difficulty but also on whether the client observes d cannot improve upon this contract where fees depend only on reported difficulty (and the probability of termination depends on whether the client observes d).

minor if and only if for some $\phi(i_k, k) \in [0, 1]$,[30]

$$(1-v)\phi(i_k,k)f(i_k,s,k)-f(i_k,m,k) = \frac{\delta}{1-\delta}(1-(1-v)\phi(i_k,k))\big(\pi(i_{k+1},k+1)-\pi(i_0,0)\big), \quad (3)$$

where $\pi(i_k, k) - \pi(i_0, 0) = \lambda(S(i_k, k) - S(i_0, 0))$ and, under the proposed contract,

$$\begin{aligned} S(i_k,k) &= (1-\delta)(y^{i_k} - \bar{q}_1(i_k,k) - c) + \delta\big[S(i_{k+1},k+1) \\ &\quad -(1-p_1^{i_k})(1-v)(1-\phi(i_k,k))(S(i_{k+1},k+1) - S(i_0,0))\big]. \end{aligned} \quad (4)$$

The client's IC constraint for effort now implies that $e = 1$ is chosen if and only if

$$\begin{aligned} &[v + (1-v)\phi(i_k,k)]f(i_k,s,k) - f(i_k,m,k) \geqslant \\ &\qquad \frac{c}{\Delta p} - \frac{\delta}{1-\delta}(1-v)(1-\phi(i_k,k))\big(U(i_{k+1},k+1) - U(i_0,0)\big), \end{aligned} \quad (5)$$

where $U(i_k, k) - U(i_0, 0) = (1 - \lambda)(S(i_k, k) - S(i_0, 0))$.

Condition (3) shows that an increase in the serious-difficulty fee $f(i_k, s, k)$ reduces the consultant's incentives to report difficulty truthfully, while condition (5) shows that it increases the client's incentives to exert effort. Furthermore, these conditions show that, of these effects, the latter is larger. Intuitively, a higher fee $f(i_k, s, k)$ reduces the consultant's truthful reporting incentives by increasing her payoff from cheating when difficulty is not observed and the relationship continues. But a higher fee $f(i_k, s, k)$ increases the client's effort incentives by increasing the contingent

[30]Note that (3) implies

$$[v + (1-v)\phi(i_k,k)]f(i_k,s,k_t) - f(i_k,m,k) \geqslant \frac{\delta}{1-\delta}(1-v)(1-\phi(i_k,k))\big(\pi(i_{k+1},k+1) - \pi(i_0,0)\big),$$

so the consultant reports a serious difficulty when difficulty is serious.

payment he must make not only when a serious difficulty is not observed and the relationship continues, but also when a serious difficulty is observed. Hence, an optimal contract sets the highest fee $f(i_k, s, k)$ that is self-enforcing, and specifies a probability of termination $(1 - \phi(i_k, k))$ such that conditions (3) and (5) hold given such fee. To see this more clearly, combine (3) and (5):

$$\begin{aligned}\frac{c}{\Delta p} \;&\leqslant\; \frac{\delta}{1-\delta}(1-(1-v)\phi(i_k,k))\big(S(i_{k+1},k+1)-S(i_0,0)\big)\\ &\quad -v\left[\frac{\delta}{1-\delta}(U(i_k,k)-U(i_0,0))-f(i_k,s,k)\right]. \qquad (6)\end{aligned}$$

Condition (6) shows that to induce truthful reporting and high effort with the lowest possible probability of termination, $f(i_k, s, k)$ must be set such that the client's self-enforcement constraint binds (that is, the expression in square brackets is equal to zero). Therefore, trade is feasible only if there exists $\phi(i_k, k) \in [0, 1]$ such that

$$\frac{c}{\Delta p} \;\leqslant\; \frac{\delta}{1-\delta}(1-(1-v)\phi(i_k,k))\big(S(i_{k+1},k+1)-S(i_0,0)\big). \qquad (7)$$

Condition (7) is not only necessary but also sufficient to sustain trade (see the proof of Proposition 3 in the Appendix). That is, if this condition holds, there exists a contract that induces the consultant to report difficulty truthfully, the client to exert effort and to be truthful about whether he can observe difficulty, and both parties to honor the promised payments. Also, it is worth mentioning at this point that, given this contract, the client has no incentives to lie about his observation of difficulty if he has no incentives to renege on the payments. Hence, as noted above, it is irrelevant for the results whether the client's observation is common knowledge or not.

If condition (7) can be satisfied with $\phi(i_k, k) = 1$, then no inefficient termination in equilibrium is necessary. However, note that for $c/\Delta p$ sufficiently large and v and $[S(i_{k+1}, k+1) - S(i_0, 0)]$ sufficiently small, (7) holds only if $\phi(i_k, k) < 1$. In this case, and by the reasoning above, an optimal termination contract sets the probability $\phi(i_k, k)$ that satisfies (7) with equality.

Propositions 3 and 4 below state the main results of the paper. The first result concerns the choice of assignments over time:

Proposition 3. *Under private monitoring and moral hazard, if trade is feasible, the low assignment is typically chosen in the early stages of the relationship and the high assignment in later stages.*

The intuition for this result stems from a tension between generating a high per-period expected surplus and minimizing inefficient termination. High assignments, by assumption, offer a higher per-period expected surplus. However, because they are more likely to be seriously difficult and thus to put the parties in a situation where cooperation is ambiguous, high assignments also require a higher probability of inefficient termination in equilibrium. On the other hand, low assignments make cooperation easier to assess and thus reduce inefficient termination, but at the cost of generating a relatively low per-period expected surplus.

The optimal path for the relationship then follows from the fact that the value of the relationship is low when the parties start trading, but increases over time as the parties accumulate relationship capital. This implies that the threat of termination off the equilibrium path—the threat that the client will end the relationship if he finds the consultant cheating—is relatively low at the beginning of the relationship, but increases as the relationship develops. As a result, the risk of termination following a serious-difficulty realization in equilibrium—the probability that the client will end the relationship when not able to validate a serious-difficulty report—must be relatively high at the beginning of the relationship, but can be reduced as the relationship develops. Therefore, in the early stages, the low assignment is optimal because it minimizes the risk of a serious-difficulty realization. Over time, however, as the threat of termination off the equilibrium path becomes more effective, this risk becomes less relevant relative to the possibility of generating a large per-period expected surplus; the high assignment thus becomes more attractive relative to the low assignment. This can be seen formally by combining equation (7) with equality with the expression for the expected surplus given in equation (4), as is shown in the proof of Proposition 3 in the Appendix.

Note that this result holds regardless of whether the relationship initially goes through an observation period, during which the parties engage only in relationship capital accumulation, or not. Such an observation period may be beneficial for the relationship because it allows the parties to have a larger stock of relationship capital, and thus be exposed to a lower risk of inefficient termination, once the trade period starts. In general, though, reducing the risk of inefficient termination to zero before the trade period starts will not be optimal; that is, $(1-\delta)(-\chi)+\delta S(i_{k+1},k+1) < S(i_k,k)$ for some $\phi(i_k,k)<1$ and $\chi>0$. So, when the parties move to the trade period, they will typically start with the low assignment and then switch to the high assignment.

The second result concerns the probability with which the relationship is terminated in equilibrium:

Proposition 4. *Under private monitoring and moral hazard, the probability that a consultant and a client end their relationship is positive and decreasing over the course of the relationship given a fixed type of assignment. At the point that the relationship switches from the low to the high assignment, given a relationship capital k, the probability of termination increases if $(1-p_1^\ell)/(1-p_1^h) \leqslant (1-\phi(h,k+1))/(1-\phi(\ell,k))$, and decreases otherwise.*

As already mentioned, a positive probability of termination on the equilibrium path is necessary to sustain trade when the monitoring and incentive problems are severe and the value of the relationship is low. As the value of the relationship increases, the threat that the relationship will be terminated if the client finds the consultant cheating becomes more effective, so the probability of inefficient termination following a serious-difficulty report that is not validated, $(1 - \phi(i_k, k))$, can be reduced. This is is shown by equation (7) with equality, and implies that, given a fixed type of assignment i, the probability of termination in equilibrium, $(1 - p_1^i)(1 - v)(1 - \phi(i_k, k))$, is also reduced. On the other hand, as the risk of inefficient termination falls, the high assignment becomes increasingly more attractive, up to the point that the parties find it optimal to switch to this type of assignment. At that moment, given relationship capital k, the probability that the relationship ends following an unobserved serious-difficulty realization falls from $(1 - \phi(\ell, k))$ to $(1 - \phi(h, k + 1))$, but the probability of a serious-difficulty realization increases from $(1 - p_1^\ell)$ to $(1 - p_1^h)$. If the effects of relationship capital on the relationship's value are relatively small, then $(1 - \phi(\ell, k)) - (1 - \phi(h, k + 1))$ is low and the latter effect dominates. In such a case, the probability of termination in equilibrium jumps at the time that the high assignment is adopted.

Although the model features only two levels of effort and difficulty (for expositional convenience), it is not hard to show that, under certain conditions, the results presented above extend to a more general setting with a continuum of effort and difficulty. That is, even if the parties can adjust the terms of the contract to change the levels of risk and return over time given a fixed type of assignment, different types of assignments may be optimal at different stages of the relationship. In particular, building relationship capital not only allows the parties to provide stronger incentives and generate a larger surplus given a type of assignment, but also to switch to riskier and more profitable types of projects.

Lastly, I consider a numerical example to illustrate the results. I take low and high assignments and costs of accumulating relationship capital such that assumptions A1 and A2 are satisfied. I assume that the costs of consulting services decrease over the first 14 periods of the relationship as relationship capital is accumulated, and remain constant thereafter.[31] I study how the optimal choice of assignment and the probability of breakup change over the course of the relationship.

I find that the optimal path for the relationship is to implement the low assignment in each of the first nine periods, and switch to the high assignment in the tenth period. (I assume costs of accumulating relationship capital without engaging in an assignment that are large enough that no observation period is optimal.) In Figure II, I compare the (lifetime) expected surplus generated by the relationship on the optimal path with that generated on two alternative paths, namely an $h-$path where the high assignment is implemented in each period and an $\ell-$path where the low assignment is implemented in each period. For both of these paths, the figure shows that the consultant and the client generate a strictly lower expected surplus $S(i_0, 0)$ than on the optimal path.

The bottom panel of Figure II describes how the probability of termination changes over the course of the relationship. As above, I look at the probability of termination (as given by $(1 - p_1^i)(1 - v)(1 - \phi(i_k, k)))$ along the optimal path and along the $h-$ and $\ell-$paths. The figure shows that, given a fixed type of assignment, the probability of termination decreases as relationship capital is accumulated. When the consultant and the client switch to the high assignment, however, the probability of termination increases discretely. It is also clear in this example that implementing the high assignment in the initial periods of the relationship would entail a very high risk of inefficient termination (not shown in the graph given the scale), which explains why the low

[31] I consider the following values: $y^h = 27$, $y^\ell = 6$, $q(i, d, k) = \tilde{q}(i, d) + g(k)$, $\tilde{q}(h, s) = 36$, $\tilde{q}(h, m) = 0$, $\tilde{q}(\ell, s) = 15$, $\tilde{q}(\ell, m) = 0$, $g(k) = 4 - 0.3k$ for $k \leqslant 13$, and $g(k) = 0$ for $k > 13$, $c = 1$, $p_1^h = 0.4$, $p_0^h = 0$, $p_1^\ell = 0.95$, $p_0^\ell = 0.55$, $\delta = 0.97$, $v = 0.095$, and $\chi = 0.5$.

assignment is implemented in those periods along the optimal path.

5.2 Improving monitoring and decreasing effort costs

The relationship dynamics described in the previous subsection could also result from changes in monitoring or effort costs.

Over the course of the relationship, the consultant and the client may become able to communicate more effectively and assess difficulty more objectively. In the model, this would be reflected as an increase in the probability that difficulty is observed by both parties, v. As the probability that the client can validate the consultant's difficulty report increases, the threat that the relationship will be terminated if the client finds the consultant cheating increases, and the consultant's incentives to cheat fall. Consequently, as the relationship grows, the need for inefficient termination in equilibrium falls. The resulting path for the relationship is as above: initially, when the client is unlikely to observe difficulty, the parties choose the low assignment to minimize inefficient termination. As they learn how to communicate and measure difficulty, the parties switch to the high assignment.

Over the course of the relationship, the cost of collaborating with the consultant may become smaller for the client. In the model, this would be reflected as a decrease in the cost of effort, c. As the cost of effort falls, the moral hazard problem is reduced—the client is willing to exert effort $e = 1$ with less-powered incentives. That is, the difference between the serious- and minor-difficulty fees required by the client's incentive compatibility constraint becomes smaller. Consequently, as the relationship grows, the consultant's temptation to cheat falls, and a lower probability of termination in equilibrium is sufficient to deter cheating. Further, as the cost of effort decreases over time, the value of the relationship increases. The resulting path for the relationship is as above: initially, when the cost of effort is high, the parties choose the low assignment to minimize inefficient termination. As they continue working together and the cost of collaborating in the assignments goes down, the parties switch to the high assignment.

6 Concluding remarks

This paper presents a model of relationship building that captures the main features and problems of the consultant-client relationship. I show that the combination of private monitoring and moral hazard that characterizes this relationship results in a risk-versus-return tradeoff, and study how this tradeoff changes and influences the choice of assignments and probability of breakup over the course of the relationship. The parties typically start with low-risk, low-return assignments, which minimize the risk of inefficient termination in equilibrium. Over time, as relationship capital is accumulated and this risk becomes less relevant, the parties switch to high-risk, high-return assignments, which generate a higher per-period expected surplus. The probability of breakup is decreasing over the course of the relationship, but may jump when the parties move to a new type of assignment. Evidence from the management consulting industry is consistent with these results.

The model could be modified or extended in different directions. For example, assignments or tasks could be classified according not to the distribution of difficulty, but rather to whether difficulty is relatively easy or hard to assess objectively. In this case, the predictions of the model would be unchanged if high-risk assignments are defined as those for which the probability that difficulty cannot be objectively evaluated, or observed by the client, is relatively high. This definition of risk, indeed, may be more appropriate for some real-world examples. Another possible extension would be to incorporate more interesting paths for the stock of relationship capital. The rate at which relationships accumulate relationship capital may depend on different factors, including the parties' efforts to collaborate and the type of assignment in which they engage. The model can readily be used to study how these factors may affect the optimal path of the relationship.

The model could also be enriched to explore how individuals on the one hand and firms on the other contribute to the building of relationships. The model presented here does not distinguish between these parties; in fact, it proposes a theory that can be used to understand the relationship between individual consultants and clients as well as the evolution of consulting firms. In reality, however, individuals and firms may play different roles and interact in a way that influences how relationships optimally grow. Likewise, the framework can be adapted to study network evolution. Starting from how a single principal-agent relationship is developed, one can build upon the model to explore how a network of relationships optimally evolves. The analysis could offer predictions not only on how individual relationships within the network will grow, but also on how the size and shape of the network will change over time.

While the model is constructed to describe the relationship between consultants and clients, the results of this paper may also shed light on how other real-world relationships are developed. In particular, the model may be well-suited to explain relationships between sellers and buyers of services that require the collaboration of both parties and are subject to informational asymmetries. Many of the examples discussed in the literature on credence goods or expert markets, like lawyers, home improvement contractors, car mechanics, and doctors, may enter this category. Anecdotal evidence suggests that, as the model predicts, these experts typically start with small and simple tasks until they build up a reputation and can switch to riskier and more profitable tasks.

The model may also be useful to explain relationships between firms, in supply chains, alliances and business groups. For instance, it could be applied to study the evolution of trade between firms that are located in different countries, for which informational asymmetries play an important role. Furthermore, the model may help explain not only how the choice of tasks changes over the course of a relationship, but also how the choice of governance structure changes. For example, using multi-industry data on alliances made between 1970 and 1989, Gulati (1995) finds that firms with prior collaborative agreements are more likely to choose non-equity alliances over equity alliances. As the value of the relationship between partner firms increases over time, they substitute formal contractual provisions with riskier and more efficient informal practices.

To conclude, I briefly review the related literature. As mentioned in the introduction, the paper is related to Sobel (1985), Ghosh and Ray (1996), Kranton (1996), and Watson (1999, 2002), which propose a theory of relationship building under incomplete information about the parties' willingness to cooperate.[32] Like the present paper, this theory predicts a gradual path for the relationship, but the mechanism is different: parties start small and increase the stakes of the relationship gradually in order to sort out cooperative from non-cooperative types. Also, in contrast with the results shown above, this theory predicts that the probability of termination is increasing over time, and becomes zero once full separation of types has occurred.[33]

Other two related recent papers are Chassang (forthcoming) and McAdams (2009). Chassang presents a model where a party cannot observe her partner's cost of cooperation in a given pe-

[32]Halac (2009) considers an agency setting with a similar form of incomplete information and studies how the optimal contract depends on the allocation of bargaining power and the distribution of types.

[33]Ghosh and Ray (1996) and Kranton (1996) assume that non-cooperative types are completely myopic, so separation of types occurs immediately in their models. Sobel (1985) studies the optimal path for a relationship in a simple lender-borrower model; see Section 6 of his paper. It is easy to verify there that the probability of termination is increasing over time until it becomes zero. Watson (1999, 2002) considers a more general setting and allows for two-sided incomplete information. Like in Sobel (1985), termination increases as information is being revealed.

riod, but can learn to predict this cost as the parties gain common experience, so the monitoring problem is reduced over time. The probability that the relationship is terminated is positive while learning occurs, and becomes zero once learning is over. McAdams abstracts from informational asymmetries; he studies an economy where each period parties choose costly efforts and whether to leave the relationship and be re-matched. Given an exogenous stochastic process, he shows that relationships in higher states persist longer into the future.

The paper is also related to the literature on repeated games under private monitoring and, more specifically, to recent work on relational incentive contracts under subjective performance evaluations. Most closely related are Levin (2003), which examines optimal contracting in a principal-agent model with moral hazard and unobservable outcomes, and MacLeod (2003), which extends Levin's model by introducing risk aversion and considering the case where outcomes are observable with positive (but less than one) probability. The present paper incorporates dynamics and different trading opportunities into this type of settings to study how relationships are optimally developed.

Finally, another related literature is that on credence goods (see Dulleck and Kerschbamer [2006] for a survey). This literature studies the informational problems associated with expert markets and the institutions that can deal with these problems. The model presented above departs from the credence goods literature by considering the dual informational problem that arises when private monitoring on one side is accompanied by moral hazard on the other side, and examining how this informational problem affects contracting and the path of relationships in a dynamic environment.

Appendix: Proofs

Propositions 1 and 2 are proven by the discussion in the text.

Proof of Proposition 3. First, note that condition (7) is necessary and sufficient for trade to occur in a perfect public equilibrium given relationship capital k. For necessity, note that if (7) does not hold, then both (3) and (5) cannot hold, so a contract that induces truthful reporting and positive effort every period does not exist. For sufficiency, suppose that (7) holds. Then let

$$\begin{aligned} f(i_k, m, k) &= \frac{\delta}{1-\delta}\big[(1-v)\phi(i_k, k)(U(i_{k+1}, k+1) - U(i_0, 0)) \\ &\quad -(1-(1-v)\phi(i_k, k))(\pi(i_{k+1}, k+1) - \pi(i_0, 0))\big], \\ f(i_k, s, k) &= \frac{\delta}{1-\delta}\big(U(i_{k+1}, k+1) - U(i_0, 0)\big), \end{aligned}$$

and w so that $U(i_k, k) = U(i_0, 0) + (1-\lambda)(S(i_k, k) - S(i_0, 0))$, $\pi(i_k, k) = \pi(i_0, 0) + \lambda(S(i_k, k) - S(i_0, 0))$. Then both (3) and (5) hold. Furthermore, payments are self-enforcing, since

$$\begin{aligned} -f(i_k, m, k) &= \frac{\delta}{1-\delta}[\pi(i_{k+1}, k+1) - \pi(i_0, 0) - (1-v)\phi(S(i_{k+1}, k+1) - S(i_0, 0))] \\ &\leqslant \frac{\delta}{1-\delta}\big(\pi(i_{k+1}, k+1) - \pi(i_0, 0)\big), \end{aligned}$$

and

$$f(i_k, s, k) \leqslant \frac{\delta}{1-\delta}\big(U(i_{k+1}, k+1) - U(i_0, 0)\big).$$

Finally, note that the client's self-enforcement constraint (given by the last inequality above) ensures that the client has no incentives to lie about his observation of d. That is, given the proposed fee $f(i_k, s, k)$, the client's payoff from claiming that he has not observed d when $d = s$ is smaller than his payoff from being truthful:

$$\phi[(1-\delta)(-f(i_k,s,k))+\delta U(i_{k+1},k+1)]+(1-\phi)\delta U(i_0,0) \leqslant (1-\delta)(-f(i_k,s,k))+\delta U(i_{k+1},k+1) \tag{8}$$

An optimal self-enforcing contract then maximizes the expected surplus subject to condition (7). In the optimum, this condition must bind; otherwise, $(1-\phi(i_k,k))$ could be reduced and thus the expected surplus increased. Then, substituting $(1-\phi(i_k,k))$ from condition (7) with equality in expression (4), the expected surplus is

$$\begin{aligned} S(i_k,k) &= (1-\delta)(y^{i_k}-\bar{q}_1(i_k,k)-c)+\delta S(i_{k+1},k+1) \\ &\quad -\max\left\{0,(1-p_1^{i_k})\left[\frac{c(1-\delta)}{\Delta p}-v\delta\big(S(i_{k+1},k+1)-S(i_0,0)\big)\right]\right\}. \end{aligned}$$

Assumptions A2b and A2c imply that $S(\ell,k) > S(h,k)$ when $S(i_{k+1},k+1)-S(i_0,0)$ is sufficiently low, and $S(\ell,k) < S(h,k)$ otherwise. The claim then follows from the fact that $S(i_{k+1},k+1)-S(i_0,0)$ is low initially, but increases over the course of the relationship as k increases. □

Proof of Proposition 4. The probability $\phi(i_k,k)$ that satisfies (7) with equality is strictly less than one if $c/\Delta p$ is sufficiently high and v sufficiently low. Also, this probability increases as $S(i_{k+1},k+1)-S(i_0,0)$ increases. Hence, given i fixed, the probability of termination in equilibrium, $(1-p^i)(1-v)(1-\phi(i_k,k))$, falls over time as k increases and thus $S(i_{k+1},k+1)-S(i_0,0)$ increases. When the parties switch from assignment ℓ to h, if the relationship capital is k, the probability of termination in equilibrium increases if and only if $(1-p^\ell)(1-v)(1-\phi(\ell,k)) < (1-p^h)(1-v)(1-\phi(h,k+1))$, which is equivalent to the condition given in the proposition. □

References

Abreu, Dilip, "On the Theory of Infinitely Repeated Games with Discounting," *Econometrica*, 1988, *56* (6), 383–396.

Bajari, Patrick and Steven Tadelis, "Incentives versus Transaction Costs: A Theory of Procurement Contracts," *RAND Journal of Economics*, Autumn 2001, *32* (3), 387–407.

Bennett, Robert and Colin Smith, "The Selection and Control of Management Consultants by Small Business Clients," *International Small Business Journal*, 2004, *22* (5), 435–462.

Bhide, Amar, "McKinsey & Company (A): 1956," 1992. Harvard Business School Case #393-066.

____ , *The Origin and Evolution of New Businesses*, Oxford University Press, USA, 2000.

Biswas, Sugata and Daryl Twitchell, *Management Consulting: A Complete Guide to the Industry*, John Wiley and Sons, Inc., New York, 2002.

Chassang, Sylvain, "Building Routines: Learning, Cooperation and the Dynamics of Incomplete Relational Contracts," *American Economic Review*, forthcoming.

Clark, Timothy, "The Market Provision of Management Services, Information Asymmetries and Service Quality – Some Market Solutions: An Empirical Example," *British Journal of Management*, 1993, *3* (2), 235–251.

____ , *Managing Consultants: Consultancy as the Management of Impressions*, Open University Press, Buckingham, 1995.

Dawes, Philip, Grahame Dowling, and Paul Patterson, "Criteria Used to Select Management Consultants," *Industrial Marketing Management*, 1992, *21* (3), 187–193.

Dulleck, Uwe and Rudolf Kerschbamer, "On Doctors, Mechanics, and Computer Specialists: The Economics of Credence Goods," *Journal of Economic Literature*, 2006, *44*, 5–42.

Fuchs, Williams, "Contracting with Repeated Moral Hazard and Private Evaluations," *American Economic Review*, 2007, *97* (4), 1432–1448.

Ghosh, Parikshit and Debraj Ray, "Cooperation in Community Interaction without Information Flows," *Review of Economic Studies*, 1996, *63*, 491–519.

Gibbons, Robert, "What the Folk Theorem Doesn't Tell Us," *Industrial and Corporate Change*, February 2006, pp. 1–6.

Gulati, Ranjay, "Does Familiarity Breed Trust? The Implications of Repeated Ties for Contractual Choice in Alliances," *Academy of Management Journal*, February 1995, *38* (1), 85–112.

Halac, Marina, "Relational Contracts and the Value of Relationships," December 2009. Mimeo, Columbia University.

Kipping, Matthias, "American Management Consulting Companies in Western Europe, 1920 to 1990: Products, Reputation, and Relationships," *Business History Review*, 1999, *73* (2), 190–220.

Kranton, Rachel, "The Formation of Cooperative Relationships," *Journal of Law, Economics, and Organization*, 1996, *12* (1), 214–233.

Kubr, Milan, *Management Consulting: A Guide to the Profession*, International Labour Office, Geneva, 2002.

Levin, Jonathan, "Relational Incentive Contracts," *American Economic Review*, 2003, *93* (3), 835–857.

MacLeod, Bentley, "Optimal Contracting with Subjective Evaluation," *American Economic Review*, 2003, *93* (1), 216–240.

McAdams, David, "Performance and Turnover in a Stochastic Partnership," June 2009. Mimeo, Duke University.

Nayyar, Praveen, "Information Asymmetries: A Source of Competitive Advantage in Diversified Firms," *Strategic Management Journal*, 1990, *11* (4), 513–519.

O'Shea, James and Charles Madigan, *Dangerous Company: Management Consultants and the Businesses They Save and Ruin*, Times Business, New York, 1997.

Shapiro, Eileen C., Robert G. Eccles, and Trina L. Soske, "Consulting: Has the Solution Become Part of the Problem?," *Sloan Management Review*, 1993, *34* (4), 89–95.

Sobel, Joel, "A Theory of Credibility," *Review of Economic Studies*, 1985, *52* (4), 557–573.

Stroh, Linda and Homer Johnson, *The Basic Principles of Effective Consulting*, Lawrence Erlbaum Associates, Inc., 2006.

Sweeney, Jack, "Raising Bain," *Consulting Magazine*, February 2001.

Watson, Joel, "Starting Small and Renegotiation," *Journal of Economic Theory*, 1999, *85*, 52–90.

———, "Starting Small and Commitment," *Games and Economic Behavior*, 2002, *38*, 176–199.

Williamson, Oliver, "The Vertical Integration of Production: Market Failure Considerations," *American Economic Review Papers and Proceedings*, 1971, *61* (2).

———, *The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting*, Free Press, New York, 1985.

Wright, Christopher, "Promoting Demand, Gaining Legitimacy and Broadening Expertise: The Evolution of Consultancy-Client Relationships in Australia," in Matthias Kipping and Lars Engwall, eds., *Management Consulting: Emergence and Dynamics of a Knowledge Industry*, Oxford University Press, New York, 2002.

Zylbersztajn, Decio and Sérgio Lazzarini, "On the Survival of Contracts: Assessing the Stability of Technology Licensing Agreements in the Brazilian Seed Industry," *Journal of Economic Behavior and Organization*, 2005, *56*, 103–120.

Figure I

Timing

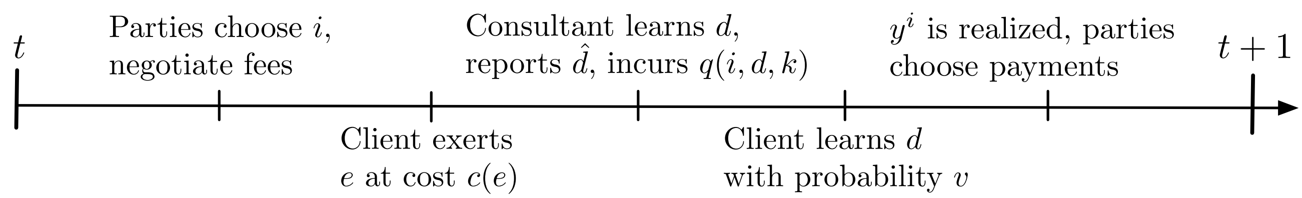

Figure II

Numerical Example