Switching costs and screening efficiency of incomplete contracts

Mehmet Bac  
Department of Economics, Bilkent University

Abstract. We study the role of switching costs in a dynamic buyer-seller relationship where quality is not contractible and the sellers retain private information about their quality-relevant abilities. In this environment buyer switching costs increase the seller’s bargaining power in negotiations for the second contract, but they also induce the seller to improve quality during the first contract, in signalling his type. The overall effect is to enhance efficiency and increase the buyer’s welfare. This beneficial effect stems from the link between quality, the buyer’s posterior beliefs, and ex post distribution of bargaining power as a function of the switching cost. JEL Classification: L14

1. Introduction

Switching costs are naturally or artificially implied in many dynamic relationships: employers incur search and selection costs to fill a vacant job, and the buyer of an idiosyncratic equipment spends time and resources to learn how to use it.

I wish to thank two referees for very helpful comments and suggestions. The paper also benefited from presentation at the EEA congress, Toulouse 1997. Remaining errors are mine. Email: bac@bilkent.edu.tr
Relationship-specific investments and contract cancellation or termination fees also generate switching costs. An extensive literature surveyed by Klemperer (1995) is devoted to the study of the consequences of switching costs. Klemperer (1987), Farrell and Shapiro (1988), and Beggs and Klemperer (1992), for example, have studied the impact of consumer switching costs on firms’ pricing behaviour and market structure.

Switching costs would be inconsequential in a world of complete and enforceable long-term contracts, but drafting such contracts requires a suprahuman effort (complete specification of concepts such as performance and quality, as well as a complete list of obligations for the parties in all states of nature). When contracts are incomplete with respect to important aspects of trade, relationship-specific investments and transmission of type information lead inevitably to a bilateral or unilateral monopoly. Switching costs can have important efficiency and surplus-sharing consequences in such contexts. For instance, ex post, the supplier can manipulate quality parameters if the buyer’s switching cost is prohibitive; as a result, the discouraged buyer may withdraw altogether from the market. If the seller has relationship-specific investments and switching is costless for the buyer, the latter may threaten to switch away. Anticipation of the fact that termination will make the supplier’s specific investments obsolete yields the well-known underinvestment result. In the first case the buyer’s switching cost generates inefficiency by putting her in the power of the seller who has all the incentives to behave opportunistically. In the second case, having the buyer offer a ‘hostage’ of an appropriate size will provide protection to the vulnerable party (the seller) who undertakes relationship-specific investments.1 Efficiency consequences of switching costs thus depend crucially on their impact on individual incentives.

In this paper an environment is identified where switching costs have potentially beneficial effects: an important aspect of trade, such as ‘quality’ or ‘performance,’ is not contractible, and suppliers differ unobservably in their innate abilities to provide quality or performance. In this environment, increasing the switching cost of the buyer can increase the total surplus by improving efficiency during the screening process. Intuition might suggest the opposite, because a higher switching cost seems to expose the buyer further to the seller’s power. The latter has all the incentives to withhold noncontractible quality. Why would the buyer benefit from the ‘glue’ brought into the relationship when no direct mechanism is available to check for the hazard of opportunism?

The link between the ex post welfare transfer generated by the switching cost and the buyer’s ex ante screening motive is the key to our result. Under noncontractible quality and asymmetric information, the buyer’s screening motive requires that the good or service be supplied below cost during a ‘probationary’ first contract. A high innate value seller would accept this contract and incur the cost

---

1 In his classic essay on bargaining, Schelling (1956) was first to note that exchange of hostages serves incentive purposes. Williamson (1976) has further developed the idea in a formal model of transaction costs.
differential to signal his type only if he anticipates he can later capture a compensating surplus. The buyer’s switching cost increases this surplus of the seller by enhancing his ex post bargaining power, which indirectly induces the seller to exert more effort (hence, improve quality) in signalling his type during the probationary contract. This is the mechanism that potentially improves the buyer’s welfare.

The mechanism relies crucially on the evolution of the parties’ bargaining power. The buyer has all the bargaining power initially, justified by the fact that she faces a large number of observationally identical potential sellers. She may lose some bargaining power to her trading partner as the relationship unfolds, however, because parity between outsiders and the incumbent seller may not obtain. The incumbent seller will have some bargaining power if he has convincingly demonstrated a high innate quality, because he enjoys a clear advantage over any other seller of unknown type. We should expect – and assume in the model – that such an incumbent seller’s enhanced bargaining power enables him to capture part of the continuation surplus.

Our basic model, then, has a buyer with non-contractible idiosyncratic needs, called ‘quality,’ and a large pool of potential sellers. The buyer’s information is incomplete as to the types of potential sellers, that is, their innate quality, which is a substitute for quality, as, for example, job-specific abilities are for performance in the labour context. Quality can also be improved through effort. The buyer’s learning about innate qualities is a primary concern in choosing contract terms, but the learning process is complicated by the possibility that a ‘bad’ seller accepts the contract and exerts a sufficiently high effort to imitate the performance of a ‘good’ seller. These are the basic features of the two-period model presented in section 2 and analysed in section 3. In section 4 a mechanism is identified through which switching costs improve the buyer’s welfare. The paper is concluded in section 5.

Though the primary focus in this paper is the impact of switching costs, its multiperiod framework relates to the literature on dynamics of incentive contracts. In a class of models where output or performance is contractible and information is asymmetric, dynamics are shown to have a negative effect on intertemporal incen-

---

2 That contracts may transform an initially competitive trading environment into bilateral monopolies is a salient feature of many real-life relationships. In the labour context, revelation of a good employer-employee match transforms an ex ante competitive environment into a bilateral monopoly. Another example is cable franchising, where it has been observed that bidding parity does not obtain at contract renewal stages (Schmalansee 1979; Williamson 1976; Zupan 1989). As initial contracts are executed, the relationship between the operator (seller) and the city or community (buyer) becomes a bilateral monopoly. Several aspects of dynamic procurement relationships fit into our model, including the dynamics of prices. To quote from Anton and Yao (1987), ‘contractors frequently submit bids below their expected cost, with the knowledge that the price will be renegotiated over time’ under bilateral monopoly conditions.

3 Several papers, some of which are mentioned below, make a similar assumption. Non-contractibility of an important aspect of trade is explicitly assumed in MacLeod and Maleckson (1988), Hart and Moore (1988, 1990), and Bac (1993). Bac (2000) applies the basic model of this paper to the labor context by including relationship-specific investments. The incomplete contracting models in which the hold-up problem is studied also assume that investment decisions are not contractible (see, e.g., Tirole 1986; Gonzalez 1998). We do not consider investment decisions in our model, but the equilibrium signalling effort of the selected seller can be interpreted as a non-contractible buyer-specific investment.
tives. This result is known as the ratchet effect (see Freixas, Guesnerie, and Tirole 1985; Laffont and Tirole 1988). The ratchet effect can be reversed if an important aspect of trade is not contractible, as in Holmstrom’s (1982) analysis of intertemporal managerial incentives. A (noisy) signal about the manager’s performance becomes public information in Holmstrom’s model, and the manager exerts effort, thanks to career concerns. Bac (1993) considers an adverse selection, bilateral monopoly framework in continuous time, and shows that durations of probationary contracts are quite instrumental in screening. A common message of these dynamic incentive models is that the contractual arrangement must provide the agent (here, the incumbent seller) a sufficient continuation surplus at later contract renewals to generate successful signaling/screening. Various modes of screening in multi-period relationships have been emphasized in the literature. In the context of dynamic insurance under hidden information, Cooper and Hayes (1987) have shown that contract lengths can serve as a screening device, provided that the insurer’s long-term commitment is credible. In Anton and Yao’s (1987) procurement model, the procurer is able to circumvent the ratchet effect by committing to an auction scheme, because the (correctly) expected cost advantage generated by learning-by-doing will favour the incumbent producer at the auction stage. Thanks to the surplus the producer expects to capture from continuation, the buyer may initially obtain full type information through a menu of contracts inducing self-selection. The main different feature of this paper is that non-contractibility of quality rules out quality-contingent menus of contracts and generates a wide scope for opportunism. There are both hidden information and moral hazard; hence, the buyer has to use (as we show) probationary contracts in order to successfully screen the seller. In this paper the focus is on the impact of switching costs on the efficiency of the screening process.

2. The model

We consider a buyer and a large number of potential sellers. There are two periods. In each period the buyer demands one unit of a specific good or service. Her per-period utility $u = q - p$ depends on the price $p$ and some idiosyncratic attributes of the good or service, called ‘quality’ and denoted $q$. Quality is affected by effort $x$ and by the seller’s innate quality $\theta$ through the following simple technology: $q(\theta, x) = \theta + x$. The seller’s effort disutility function $d(x)$ is assumed to be strictly convex and increasing in $x$. If he is on contract with the buyer, the per period utility of a seller is $v = p - c - d(x)$, where $c$ denotes the production cost net of quality effort.

4 Laffont and Tirole (1988) show that no initial contracts exist to separate types when the buyer cannot commit to ignore the information she obtains. Baron and Besanko (1987) also discuss this result.

5 Hosios and Peters (1989) draw attention to the importance of commitment in Cooper and Hayes’s results and extend their analysis to the case without commitment. Dionne and Doherty (1994) is a further extension to the case of renegotiation-proof long-term contracts.
parties’ utilities are simply sum of their per period utilities, \( U = u_1 + u_2 \) for the buyer and \( V = v_1 + v_2 \) for the sellers, and the outside options of all the parties are normalized to zero.

The innate quality \( \theta \) (type) of a seller is his private knowledge. We assume two possible realizations of types: \( \theta = \theta_H \) with probability \( \pi \) and \( \theta = \theta_L \) with probability \( 1 - \pi \). We have \( \theta_H > \theta_L \), so a \( \theta_H \)-seller has an absolute advantage in providing quality. Moreover, \( \theta_H > c > \theta_L \), which implies that matching with a \( \theta_L \)-seller who exerts zero effort generates a negative per period total surplus (i.e., the sum of the parties’ per period utilities \( [\theta_L - p_1] + [p_1 - c] \) is negative). We shall also assume that the expected surplus from matching with a seller of unknown type who exerts zero effort is positive, that is, \( \pi \theta_H + (1 - \pi) \theta_L - c \geq 0 \).

Thus, the only information that the buyer ignores is the innate quality of each potential seller. The analysis will be focused on efficiency consequences of switching costs and the interaction between switching costs and the process of screening will be highlighted. In this model, efficiency requires that the buyer be matched with a \( \theta_H \)-seller who exerts the effort \( x = \arg\max [x - d(x)] \).

Obviously, the buyer could induce the efficient outcome if quality were contractible, through a quality-contingent payment scheme. We assume that this direct mechanism is not feasible: the seller’s effort choice \( x \) and the quality of the good or service demanded by the buyer are not contractible. The buyer observes delivered quality but not the seller’s effort. Court ordering with respect to other aspects of trade are assumed to be efficacious. These are verifiable and contractible variables, such as the unit price \( p \), switching costs \( k \) or cancellation fees \( f \), contract durations, and delivery of the good or service.6 Bonding or any quality-contingent payment scheme is impossible because quality itself is not contractible.

In terms of duration, the buyer’s relationship with a seller can be governed by two alternative contractual arrangements: a short-term contract \( C_1 = \{p_1; f; k\} \) that governs only the first period; and a long-term contract \( C = \{p_1, p_2; f; k\} \) that governs both periods. The parties’ especially the buyer’s possibility of changing the trading partner will obviously depend on the contents of the initial contract. Switching to another seller is impossible or too costly under a two-period long-term contract that stipulates a sufficiently high cancellation fee, whereas at the other extreme a one-period contract that makes no reference at all to the second-period price allows for costless switching.7

6 Thus, trade will be governed by incomplete contracts, say, of the form ‘in period t, one unit of the good manufactured by the seller S should be delivered to buyer B for the payment \( p \). The party i = S,B who terminates the relationship without the consent of the other party will pay the fee \( f \) (or incur a switching cost of \( k \)).’ A cancellation fee is a transfer from one party to the other, while a switching cost is a real resource cost or a breach penalty to be paid to a third party.

7 The long-term/short-term distinction is, of course, a relative one. We say that a contract \( C_1 \) has a longer term than \( C_2 \) if \( C_1 \) covers (specifies corresponding prices for) a larger number of periods than \( C_2 \). We allow short-term contracts to specify a fee or a switching cost to be paid by one party or both if the relationship is terminated, even though the terms of a second contract – if the parties choose the continuation option – is left open to negotiation.
The two-period relationship described above is modelled as a dynamic game that involves the following sequence of events. The buyer determines the form and content of an initial contract and offers it to many observationally identical sellers.\(^8\) The sellers make acceptance choices, the buyer chooses one among those who accept, and the relationship begins. In each period covered by the contract, the incumbent seller determines \(x\), his quality-improving effort. Observing the (non-contractible) quality, the buyer updates her beliefs about the incumbent seller’s type. The parties (especially the buyer) may have the option to terminate the relationship at the interim date if the initial contract does not stipulate a high switching cost or cancellation fee. Under a short-term contract that does not cover the second period, the parties negotiate the second contract. The parties’ relative bargaining power will depend on the type information transmitted by the incumbent seller, as explained further in the next section. Finally, under a long-term contract that specifies the second-period price, the relationship will extend to the second period and trade will occur at the specified price.

An implication of the model that is worth noting at this stage is that non-contractibility of quality and hidden type information together generate the potential for a \(\theta_L\)-seller to outperform a \(\theta_H\)-seller if the latter does not exert a sufficient effort. That is, effort can be used to conceal differential innate qualities because the two are substitutes.\(^9\) The buyer should therefore take into consideration the type-concealing effect of effort at both the outset and the contract renewal stage to avoid being locked-in with a ‘bad’ seller.

### 3. Equilibrium analysis

Let us consider, first, the benchmark case in which the sellers’ innate qualities are common knowledge, to highlight the role of asymmetric information in our results. When the buyer knows the innate qualities of the sellers, she will make her initial contract offer exclusively to \(\theta_H\)-sellers. Since quality and effort are not contractible and effort is costly, the selected seller will set \(x_t = 0\) in each period \(t\) and provide the quality \(q_t = \theta_H\). The seller’s utility is then \(p_1 - c + [p_2 - c]\). As a result, the buyer’s initial offer is a long-term contract with prices \(p_1 + p_2 = 2c\) (or a sequence of two short-term contracts with prices \(p_1 = p_2 = c\), which yields the buyer the utility \(2[\theta_H - c]\) and the seller his outside option utility, zero. In contrast to the case of asymmetric information, under symmetric information switching costs can have no

---

\(^8\) The analysis goes through if, alternatively, the buyer invites the sellers to submit contract proposals.

\(^9\) For example, this phenomenon manifests itself in the labour context where a less talented but hard-working employee occasionally performs better than a highly talented employee, in vertical relationships where ‘bad’ sellers invest to provide a temporary quality improvement and satisfy the buyer’s needs, thereby to obtain an increase in price/volume of future orders. We shall show that in equilibrium the buyer avoids adverse selection, mainly because she is able to observe quality without noise. Including a noise in observed quality, though not our purpose in this paper, can generate selection errors of the type mentioned above.
efficiency consequences in this relationship. Since the buyer will match with a $\theta_H$-seller, all that a switching cost can do is to transfer some bargaining power from the buyer to the incumbent seller at the interim contract renewal date (if the initial contract covers only the first period). In particular, the buyer obviously cannot expect to gain from locking herself in with the incumbent seller. Under asymmetric information, however, the $\theta_H$-seller may have to exert effort in proving his type. This quality-improving effort can be affected, though indirectly, by the level of buyer switching costs, through the second-period surplus that the incumbent seller expects from continuation. In the rest of the paper we focus on the case of asymmetric information.

If the buyer chooses a long-term contract $C = \{p_1, p_2, f\}$ with a sufficiently high cancellation fee $f$, the seller faces no risk of termination. The parties are locked in for two periods; therefore, the seller will set $x_t = 0$ in each period, which yields him the utility $p_1 + p_2 - 2c$. This utility should be non-negative to satisfy the individual rationality constraint of the sellers; thus, the buyer should choose $p_1$ and $p_2$ so that $p_1 + p_2 = 2c$, which yields her the utility $2[(\pi \theta_H + (1 - \pi))\theta_L - c]$. Note that the selected seller’s utility $p_1 + p_2 - 2c = 0$ does not depend on his type; thus, the contract $C = \{p_1, p_2, f\}$ has no screening power. This unfortunate outcome is due to the prohibitively high cost of breach and absence of any induced link between the seller’s second-period utility $v_2$ and the buyer’s beliefs about the seller’s type. Thus, to screen out $\theta_L$-sellers, the contractual arrangement should allow for termination when the buyer learns that she is matched with a $\theta_L$-seller. By inducing a continuation equilibrium in which the buyer’s beliefs affect the second-period price, the contractual arrangement should also provide the selected seller with an incentive to exert effort. Long-term contracts that specify $p_2$ in advance cannot achieve these objectives. We therefore consider all-or-nothing contract offers of the form $C_1 = \{p_1; k; f\}$ eliminating all references to and commitments for the second-period price. We start the analysis below under the assumption $f = k = 0$ and show in section 4 the potential role of a positive switching cost or cancellation fee.

Under a short-term, incomplete contract $C_1 = \{p_1\}$ that allows for costless switching, the buyer observes $q_1$ and updates her beliefs $\mu(\theta_H|q_1) = \text{prob}(\theta = \theta_H|q_1)$ about her trading partner’s innate quality. As the contract expires, she makes a binary continuation decision denoted $b \in \{0, 1\}$. If she opts for keeping the seller ($b = 1$), then the parties negotiate the terms of the second contract $C_2$, namely, the price $p_2$. We study the relationship backward in time, starting with period two. Note that the seller will exert no effort in period two because quality and effort are not contractible; hence, $q_2 = \theta$ and the corresponding total surplus is $\theta - c$.

---

10 See Bac (1993) for a discussion of this observation. The result that when quality is not contractible, commitments for contractible variables beyond the actual period decrease the buyer’s welfare is termed ‘the principle of negative protection’ by Farrell and Shapiro (1989). It holds, except trivially, when the relevant variables are chosen at levels that would have been chosen under non-commitment.
3.1. The contract for period two
The initial contract may (in equilibrium, it will) transform the parties’ options and bargaining power at the interim negotiation stage. Absent a switching cost, the crucial parameter determining the parties’ relative bargaining position is the buyer’s learning about her trading partner; in fact \( \mu (\theta_H | q_1) \) provides the only link between the two periods. For instance, the incumbent seller who proves an innate quality of \( \theta_H \) will enjoy a great advantage over outsiders because at the interim date the buyer will strictly prefer keeping the seller, who she knows to be of type \( \theta_H \), rather than contracting with another seller of unknown type. This is clear from the fact that the buyer’s second-period expected utility \( \mu \theta_H + (1 - \mu) \theta_L - p_2 \) is increasing in \( \mu \). Expected gains from continuation are strictly positive; hence, the incumbent seller has some bargaining power whenever \( \mu > \pi \). The buyer has all the bargaining power otherwise, if \( \mu \leq \pi \).

Using the parties’ disagreement utilities, we can determine the support of the ‘bargaining set,’ including the negotiated price \( p_2^* \), as follows. The seller’s disagreement option determines the lower bound for \( p_2^* \), which is clearly \( c \). The best the buyer can do by switching is to sign a contract \( \{ p = c \} \) for the last period and obtain the expected utility \( U^d = \pi \theta_H + (1 - \pi) \theta_L - c > 0 \). Given the buyer’s updated beliefs, the upper bound for \( p_2^* \), denoted \( \tilde{p}_2(\mu) \), can be determined through the condition stating the buyer’s indifference between switching and keeping the incumbent seller: \( \mu \theta_H + (1 - \mu) \theta_L - \tilde{p}_2(\mu) = U^d \). As mentioned above, if the incumbent seller conveys type information leading to \( \mu > \pi \), then \( \tilde{p}_2(\mu) > c \) and continuation generates a strictly positive expected surplus. Given the (commonly known) negotiation procedure and rational expectations about the outcome of negotiations, the price \( p_2^*(\mu) \in [c, \tilde{p}_2(\mu)] \) will be anticipated in advance. The only assumption we make about the bargaining solution is that \( p_2^*(\mu) \) is an increasing function of \( \mu \). For example, the ‘equal splitting of the expected pie’ solution \( p_2^N = (\tilde{p}_2(\mu) - c)/2 \) satisfies this assumption. First, the size of the expected second-period surplus is increasing in \( \mu \); hence, it is natural that the seller captures some of this incremental surplus through a higher price. Second, the more the buyer is convinced that \( \theta = \theta_H \), the higher is the seller’s bargaining power and thus the higher the price he obtains for the second contract. We can therefore write the seller’s second-period welfare as \( v_2^*(\mu) = p_2^*(\mu) - c \), provided that \( \mu > \pi \) and the buyer opts for continuation.

3.2. First-period contract and efficient equilibria
We now proceed to period one and study the equilibrium first contract and corresponding post-contract strategies, given \( p_2^*(\mu) \). The combined strategies and

---

11 The expression for \( U^d \) assumes that the number of potential sellers is large enough that the probability \( \pi \) remains roughly constant when the buyer decides to switch to an outside seller.

12 It should be emphasized that \( p_2^*(\mu) \) is not a reward function to which the buyer commits herself. It is the anticipated reduced-form solution to the negotiation problem.
the belief system must form a perfect Bayesian equilibrium (PBE), denoted \( \{ p_1^*, x_H^*, x_L^*, b^*, \mu^* \} \).

The game has many screening PBEs, depending on the buyer’s postulated belief system. Here, we focus on the (possibly constrained) efficient screening PBE in which \( \theta_L \)-sellers are screened out and the selected \( \theta_H \)-seller exerts the efficient quality-improving effort. The efficient effort is \( x^O = \arg\max_x [x - d(x)] \) and the efficient quality is \( q^O = \theta_H + x^O \).

The following conditions must hold in any screening (not necessarily efficient) PBE. First, provided that the effort \( x_H \) leads the buyer to update her beliefs to \( m = 1 \), the incumbent \( \theta_H \)-seller’s individual rationality constraint must hold:

\[
p_1 - c - d(x_H) + [p_2^*(1) - c] \geq 0. \tag{1}
\]

Thus, accepting the initial price \( p_1 \) and exerting the effort \( x_H \) should yield a non-negative utility to \( \theta_H \)-sellers. The second condition is that the buyer’s contract offer \( \{ p_1^* \} \) must screen out \( \theta_L \)-sellers, that is, offer no ‘free lunch’ to \( \theta_L \)-sellers who would accept the contract, exert zero effort, and switch back to their outside options at the end of period one. In other words, the first-period price should not exceed the cost \( c \):

\[
p_1 \leq c. \tag{2}
\]

Finally, the strategy of accepting \( \{ p_1^* \} \) to exert an effort \( x_L \) and produce a quality \( \theta_L + x_L \) that leads the buyer to revise her beliefs to \( \mu = 1 \) should not be beneficial for a \( \theta_L \)-seller:

\[
p_1 - c - d(x_L) + [p_2^*(1) - c] \leq 0. \tag{3}
\]

In a screening (not necessarily efficient) PBE, the buyer can update her beliefs to \( \mu = 1 \) on observing the quality \( \theta_L + x_L \) only if it is at least as high as the quality \( \theta_H + x_H \) provided by the \( \theta_H \)-seller. Thus, given the effort strategy \( x_H \) and the quality \( \theta_H + x_H \) of a \( \theta_H \)-seller, the minimum effort that a \( \theta_L \)-seller should exert to imitate this quality is

\[
x_L = x_H + (\theta_H + \theta_L). \tag{4}
\]

The buyer’s utility from matching and continuing with a \( \theta_L \)-seller who exerts the effort \( x_H \) in period one is

\[
U = \theta_H + x_H - p_1 + \theta_H - p_2^*(1). \tag{5}
\]

13 Note that the effort \( x_L \) is not an effort that a \( \theta_L \)-seller actually exerts in a screening PBE by accepting the initial contract; it is, rather, the effort that is necessary to imitate the quality provided by the \( \theta_H \)-seller.
The efficient screening PBE strategies can now be obtained by solving the problem of choosing \( p_1 \) and \( x_H \) to maximize (5) subject to (1), (2), (3), and (4). The following proposition, proved in the appendix, characterizes the equilibrium.

**Proposition 1.** Given the expected price \( p_2^*(\mu) \) to be negotiated for the second contract, the efficient screening PBE has the following properties: only \( \theta_H \)-sellers accept the initial contract \( \{ p_1^* \} \), and

i) if \( p_2^*(1) - c \geq d(x^O) \), the selected \( \theta_H \)-seller exerts the effort \( x_H^* = x^O \) and the buyer sets \( p_1^* = c + d(x^O) - \{ p_2^*(1) - c \} \);

ii) if \( p_2^*(1) - c < d(x^O) \), then \( x_H^* = d^{-1}(p_2^*(1) - c) < x^O \) and the buyer sets \( p_1^* = c \).

Two types of equilibria may arise, depending on whether the screening PBE is efficient or constrained by the ‘no free lunch’ constraint (2). The buyer obtains the efficient quality \( q^O = \theta_H + x^O \) in the type-(i) equilibrium and sets a price \( p_1^* \leq c \) that binds the individual rationality constraint (1) of the \( \theta_H \)-seller. A price below cost, which is necessary to screen out \( \theta_L \)-sellers, is feasible because the second-period surplus \( p_2^*(1) - c \) that the \( \theta_H \)-seller expects more than compensates him for the cost of exerting the efficient effort \( x^O \). Accepting the contract and exerting the effort \( x^O \) is individually rational because condition (1) holds, and it is incentive compatible to exert the effort \( x^O \) because the buyer interprets a lower quality as coming from a \( \theta_L \)-seller and accordingly terminates the relationship. In a type-(ii) equilibrium the buyer is not able to induce the first-best level of quality; she therefore sets \( p_1^* = c \) and contents herself with the highest possible but suboptimal quality \( \theta_H + x_H^* = \theta_H + d^{-1}(p_2^*(1) - c) \). In both types of equilibria the \( \theta_H \)-seller’s individual rationality constraint (1) is binding, while (3) is not binding – quite naturally so because the \( \theta_H \)-seller has an innate quality advantage.

Two remarks are in order. First, compared with the benchmark case of symmetric information, the buyer’s utility is unambiguously higher in the equilibrium characterized in proposition 1. This is intuitive, because the \( \theta_H \)-seller exerts effort in demonstrating his type and improves non-contractible quality. Second, we have focused on the efficient screening PBE. We could have selected an alternative screening PBE, for instance, the one satisfying a refinement such as the Intuitive Criterion. Applying this criterion to the continuation game that extends from the buyer’s contract offer, however, has no qualitative impact on our results concerning the potential efficiency effects of buyer switching costs.\(^{14}\)

\(^{14}\) Roughly, in this model the Intuitive Criterion will eliminate all equilibria supported by a belief system that interprets as coming from a \( \theta_L \)-seller an out-of-equilibrium quality signal that a \( \theta_L \)-seller can never beneficially produce. Accordingly, in an equilibrium satisfying the Intuitive Criterion the \( \theta_H \)-seller would exert the minimum (not necessarily efficient) effort that is just sufficient to differentiate himself from \( \theta_L \)-sellers; that is, (3) should be binding. Even if we impose the Intuitive Criterion, the equilibrium may be constrained by the ‘no free lunch’ condition (2) and, as we show in section 4, switching costs would then improve efficiency. The analysis of the screening PBE satisfying the Intuitive Criterion is available from the author upon request.
4. The role of switching costs

A buyer-switching cost \( k > 0 \) will decrease the buyer’s disagreement utility at the interim negotiation stage and hence indirectly increase the negotiated price that the seller expects from the second contract. We denote this price by \( p_2^*(\mu, k) \) and accordingly assume that \( p_2^*(\mu, k) \) is increasing in \( k \) (as long as a higher \( k \) decreases the buyer’s disagreement utility). A buyer cancellation fee (a resource transfer from the buyer to the seller if the buyer decides to terminate the relationship) also decreases the buyer’s disagreement utility, but unlike a switching cost, it increases the seller’s disagreement utility because the fee goes to the seller. The qualitative impact of a cancellation fee on the second-period price should therefore be similar. In this relationship, however, these two types of frictions can have quite different efficiency consequences. Since the buyer’s switching cost is a resource cost that she incurs or pays to a third party only if she decides to switch to another seller, it has no impact on her ‘no trade’ option in period two. This is in contrast to the case of a cancellation fee, which the buyer has to pay to the incumbent seller if she terminates the relationship whether or not she switches to another seller in period two. Therefore, as we explain below, a high cancellation fee can upset the screening property of the PBE.

In proposition 2 we consider the efficiency consequence of a buyer-switching cost. Let us define a critical level \( \tilde{k} \) for the switching cost,

\[
\tilde{k} = \pi \theta_H + (1 - \pi) \theta_L - c,
\]

such that the buyer is indifferent between two options: switching to another seller and not trading for the second period. Note that for \( k > \tilde{k} \), the buyer will prefer not to trade in period two, rather than switch to a seller of unknown quality. Thus, if \( k > \tilde{k} \), her disagreement utility in bargaining for the second-period price should remain constant and equal to zero. For \( k \in [0, \tilde{k}] \), the buyer’s disagreement utility is decreasing in \( k \); hence the price \( p_2^*(1, k) \) is increasing in \( k \).

**Proposition 2.**

i) If \( p_2^*(1, 0) - c \geq d(x^O) \), for any switching cost \( k > 0 \) there exists a first-period price \( p_1^k \) such that the contract \( \{p_1^k, k\} \) yields the buyer the same welfare as the efficient screening PBE under the contract \( \{p_1^*, 0\} \).

ii) If \( p_2^*(1, 0) - c < d(x^O) \), introducing a switching cost \( k > 0 \) will improve efficiency of the screening PBE and the buyer’s welfare. The efficient quality \( q^O \) can be induced by setting \( k^* \) that satisfies \( p_2^*(1, k^*) - c = d(x^O) \), provided \( k^* \leq \tilde{k} \). Otherwise, the maximum first-period quality that can be achieved by imposing a buyer switching cost is less than \( q^O \) but higher than the quality under \( k = 0 \).

**Proof**

i) The assumption \( p_2^*(1, 0) - c \geq d(x^O) \) implies that we are in case (i) of proposition 1. Since the switching cost \( k \) reduces the buyer’s disagreement utility, \( p_2^*(1, k) > p_2^*(1, 0) \) whenever \( k > 0 \) (the reader can verify this relation for the
Nash bargaining solution). Regardless of the anticipated price $p_2^*(1, k)$, the buyer can always adjust the first-period price $p_1$ so that $p_1 = c + d(x^O) - [p_2^*(1, k) - c]$, (binding the individual rationality constraint of the $\theta_H$-seller) and restore price optimality. This generates the same efficient level of equilibrium effort and quality. Now, since the same equilibrium quality obtains and the incumbent seller’s utilities under the contracts $\{p^k, k\}$ and $\{p_1^*, 0\}$ are identical, the buyer, too, must be indifferent between the two contracts.

ii) We are in case (ii) of proposition 1 and the induced effort and quality are suboptimal. A positive switching cost increases the second-period negotiated price from $p_2^*(1, 0)$ to $p_2^*(1, k)$, which yields the seller a strictly positive equilibrium utility, assuming the strategies to be fixed. The positive equilibrium utility of the seller, however, generates a ‘more efficient’ screening PBE in which the seller exerts a higher effort. As shown in section 3 above, this equilibrium effort strategy $x_H^k$ is given by the binding individual rationality constraint (1) of the $\theta_H$-seller. It satisfies

$$d(x_H^k) = p_2^*(1, k) - c, \tag{7}$$

where we used the fact that $p_1^* = c$. Since $p_2^*(1, k)$ is increasing in $k$ as long as $k \leq \hat{k}$, the first-best quality can be achieved if there is a $k \in [0, \hat{k}]$ such that (7) holds for $x_H^k = x^O$. Otherwise, the first-period quality will be suboptimal, although higher than the case $k = 0$. QED

One would expect higher buyer switching costs to impede effort incentives during period one. Why should the seller exert more effort when the buyer is almost locked in? Recall that the seller’s bargaining power (hence, second-contract surplus) depends on his first-contract performance, through the buyer’s beliefs. So the seller will not shirk when the buyer’s switching cost is higher, because in equilibrium he risks being interpreted as a $\theta_L$-seller, which would induce contract termination because the buyer’s (negative) utility from continuation with a $\theta_L$-seller is lower than both the expected utility from contracting with a seller of unknown type and the utility from not trading for the second period.

The intuition for why the buyer may benefit from the ‘glue’ brought into the relationship lies in the form of the intertemporal welfare transfer induced by the switching cost. The beneficial effect of a switching cost comes into being when, in the absence of switching costs, the first-period price $p_1^*$ is already at the maximal level $c$ consistent with the buyer’s motive of screening out $\theta_L$-sellers and the second-period price $p_2^*(1, 0)$ is not high enough to compensate the $\theta_H$-seller for the disutility of the efficient quality-improving effort. Introducing a switching cost increases the second-period price and induces a higher effort along the PBE path in period one. This intertemporal welfare transfer makes the incumbent seller indifferent but improves the buyer’s welfare and hence generates a Pareto improvement.

As mentioned above, a buyer cancellation fee has a similar effect on the second-period price. A buyer cancellation fee decreases the buyer’s disagreement utility in
negotiations, however, whether she switches to another seller or decides not to trade at all. For this reason, switching costs and cancellation fees have potentially differential equilibrium effects. In contrast to the case of switching costs, if the buyer sets a cancellation fee too high, her option of continuation with even a $\theta_L$-seller will dominate other options at the contract renewal stage. The $\theta_L$-sellers who foresee this will accept the initial contract price $p_1^*$ if they expect a net positive benefit from locking in the buyer for two periods. Then, the initial contract and the induced PBE will no longer be of the ‘screening’ type. The buyer has to take into consideration the undesirable potential penetration of $\theta_L$-sellers and determine the (optimal) level of the cancellation fee accordingly. Such a penetration strategy from $\theta_L$-sellers is impossible under buyer switching costs because the buyer’s ‘no trade’ option will not be affected. In this relationship, switching costs have a relative advantage over cancellation fees.

5. Conclusion

To recapitulate, in this paper we have shown that frictions (in the form of switching costs or cancellation fees) may improve screening efficiency when an important aspect of trade is not contractible but is affected by sellers’ effort and privately known innate abilities. In such environments the optimal screening contractual arrangement consists of a sequence of two contracts that induces an intertemporal welfare transfer. The good-quality seller on probation must incur a loss to demonstrate his type. The seller would accept incurring this loss if he anticipates that through signalling he will improve his bargaining power and capture a compensating surplus later from the second contract. The process indirectly generates quality-improving effort through the link between the buyer’s posterior beliefs, second-period price, and first-period quality. The buyer’s motive to induce an efficient quality-improving effort may be in conflict with her motive to screen out ‘bad’ seller types at the outset. We have shown that an appropriate buyer-switching cost can improve the efficiency of first-period quality-improving effort while screening out ‘bad’ seller types. This beneficial effect arises when the first-period price is already at the maximal level consistent with a screening equilibrium and the continuation surplus that the seller expects is too low to induce the efficient effort during the first period.

Appendix

We prove proposition 1 in this appendix. We verify below that the following strategies and belief system form the PBE described in the proposition.

The buyer: Set $b^* = 1$ if $\mu = 1$, $b^* = 0$ if $\mu = 0$. Offer the (short-term) contract $p_1^* = c$ if $d(x^O) > p_2^*(1) - c$.

$\theta_H$-sellers: Accept the above contracts offered by the buyer. Reject the first-period price below $p_1^*$ given above. If selected, exert the effort $x_H^*$ such that $p_1 - c - d(x_H^*) + [p_2^*(1) - c] = 0$. 
\( \theta_L \)-sellers: Reject the above contracts offered by the buyer. Accept the price \( p_1 \) only if \( p_1 > c \).

The buyer’s belief system: If \( d(x^O) \leq p^*_2(1) - c \), then \( \mu(\theta_H|q_1) = 1 \) if \( q_1 \geq q^O = \theta_H + x^O \), and \( \mu(\theta_H|q_1) = 0 \) if \( q_1 < q^O \). If \( d(x^O) > p^*_2(1) - c \), then \( \mu(\theta_H|q_1) = 1 \) if \( q_1 \geq \tilde{q} = \theta_H + d^{-1}(p^*_2(1) - c) \), and \( \mu(\theta_H|q_1) = 0 \) otherwise.

The buyer’s continuation decision \( b^* \) is clearly optimal given the belief system. If the \( \theta_H \)-seller exerts the effort \( x^*_H \), the price \( p^*_H = c + d(x^*_H) - [p^*_2(1) - c] \) is the lowest price that the \( \theta_H \)-seller would accept and hence is an optimal strategy for the buyer. This price satisfies the equilibrium conditions (1), (2), and (3). Since (1) holds, it is optimal for \( \theta_H \)-sellers to accept the contract offer. The effort strategy \( x^*_H \) is optimal because an effort \( x^*_H < x^*_H \) leads to \( \mu = 0 \) and the buyer’s switch at the interim date. Finally, the \( \theta_L - \) sellers’ strategy to reject the contract is optimal because (3) is not binding; that is, accepting the contract will generate a negative utility. Note that the belief system is consistent with the strategies.

QED

References

Gonzales, Patrick (1998) ‘Specific investment, commitment and observability,’ mimeo, Université Laval


MacLeod, Bentley W., and James M. Malcomson (1988) ‘Reputation and hierarchy in dynamic models of employment,’ *Journal of Political Economy* 96, 832–54


