HIRSCHMAN MEETS WILLIAMSON: RELATIONSHIP-SPECIFIC INVESTMENT AND LOYALTY

KÁROLY MIKE

ABSTRACT In Albert Hirschman’s theory, loyalty plays a key role in the equilibrium between exit and voice. This article extends economic (rational choice) analysis to the emergence of loyalty, which Hirschman considers an exogenous factor. This is accomplished by linking Williamson’s theory of specific investment to Hirschman’s model. Three cases are distinguished: (1) loyalty is due to specific investment; (2) loyalty is due to (intermediate) factors influenced by specific investment; and, (3) loyalty is independent of specific investment. A simple model formalizes the first case. A paradoxical dynamic of loyalty is identified: a lower degree of specificity may lead to a weakening of loyalty in the short run but a strengthening of loyalty in the long run. An application to the process of European integration is sketched.

KEYWORDS loyalty, exit, voice, embeddedness, relationship-specific investment

INTRODUCTION

Albert Hirschman’s Exit, Voice and Loyalty (1970) is beyond doubt a true classic of the social sciences. Its central thesis – that human relationships are supported by two fundamental mechanisms: exit and voice – is widely cited and...
often used to structure thinking in economics, political science and sociology. However, Hirschman went beyond identifying the exit-voice dichotomy and proposed that the two mechanisms should and could be studied together. In fact, he added the concept of loyalty so as to explain the interaction between exit and voice. I think it is not unfair to say that Hirschman’s proposal has not generated a particularly fruitful research program. There have been very few formal elaborations of Hirschman’s numerous intuitions or the simple geometrical models presented in his book (Moses 2005, Gehlbach 2006, Lorz and Nastassine 2007). Empirical applications have been more numerous but also relatively limited, considering the fame of the original theory (e.g. Orbell – Uno 1972, Sharp 1984, Squire et al. 1987, Feld 1997, Dowding et al. 2000).

Hirschman himself suggested a reason for this failure by emphasizing “a stress on the possibility of unstable equilibria between exit and voice as one drives out the other and on the lack of once-and-for-all solutions through optimal mixes of exit and voice. In other words, the new approach does not satisfy our craving for equilibrium, harmony and final repose.” (1976: 386). I believe that an additional explanation for the limited fruitfulness of Hirschman’s approach is a missing element in his theory. Hirschman stressed the role of loyalty in the interaction of exit and voice but left the emergence of loyalty unexplained. While he explained exit and voice as rational action, he discussed loyalty as an exogenous trait of unclear origin that may or may not characterize a group or an individual. In other words, loyalty remained outside the scope of economic (rational choice) analysis. By arguing that loyalty is the key to understanding the interaction between exit and voice but it is not amenable to economic analysis, Hirschman seemed to close down the avenue for future research that he had opened. I propose that Hirschman’s theory can be augmented and perhaps revitalized by extending rational choice analysis to the phenomenon of loyalty. My thesis is that the concept of relationship-specific (or idiosyncratic) investment is fundamental for understanding loyalty as rational action. Thus, Hirschman’s theory should be modified by borrowing the central concept from Oliver Williamson’s theory of transaction costs.

Like Hirschman, Williamson departed from the neoclassical view of the market which stresses the importance of competition and assumes that the key mechanism for dealing with underperforming economic partners is moving on to their rivals. Building on Coase’s (1937) insight, Williamson set out to explain the existence of hierarchical and other long-term economic relationships (1975, 1985). He introduced the notion of relationship-specific – or, as he prefers to call it, idiosyncratic – investment, which is investment that loses
a significant part of its value outside a given relationship. Such investments play a key role in transforming easy-to-exit, more or less impersonal contract-based transactions into long-run, difficult-to-exit relationships whose functioning depends on internal governance mechanisms. In particular, unified or hierarchical governance is to be expected for recurring transactions characterized by uncertainty and highly idiosyncratic investments.

In a brief exchange, Hirschman and Williamson themselves noted that their theories share some basic ideas. Of particular interest to us is Hirschman’s remark that “hierarchy can be considered as a special variety of institutionalized voice” (1976: 386-7). Williamson agreed by stating that “both EVL [Exit, Voice and Loyalty] and MH [Markets and Hierarchies] feature the use of market alternatives to autonomous contracting in which voice is prominent” (1976: 373). Having recognized the co-existence of voice and exit mechanisms in the economy and other social fields, both authors attempted to identify the circumstances under which exit or voice is likely to emerge. Hirschman argued that a fundamentally important circumstance is the presence of loyalty, whereas Williamson pointed to the importance of idiosyncratic investments. All the more surprising that the authors did not reflect upon the possible correspondence between the two concepts.3 I propose that Williamson was more successful in unearthing the microanalytics of why people may choose to stay in relationships and try to influence them from within, rather than leave them. In fact, he identified a rational choice explanation for the emergence of loyalty; the undertaking of relationship-specific investment. This explanation can be transferred to Hirschman’s more general exit-voice framework. By linking the two theories, loyalty can be brought (at least partially) within the ambit of rational choice explanation. At the same time, Williamson’s concept is proven to be relevant for all types of human cooperation, regardless of the specific mechanisms of governance.

After briefly summarizing Hirschman’s theory and Williamson’s concept of relationship-specific investment, I integrate the latter into Hirschman’s framework through his concept of loyalty. Three cases are distinguished: (1) loyalty is due to specific investment; (2) loyalty is due to (intermediate) factors influenced by specific investment; and, (3) loyalty is independent of specific investment. In the first case, the concept of loyalty can basically be replaced in Hirschman’s model by the concept of specific investment. An advantage of this replacement is that decisions about specific investment and specific assets are usually easier to identify and explain as responses to factors in the social environment than loyalty per se.

3 To the author’s knowledge, this has not yet changed. See, for example, Hirschman 2008.
Although loyalty seems to be especially prevalent in relationships where investments tend to be highly idiosyncratic, the multi-faceted phenomenon of loyalty cannot simply be reduced to the concept of specific investment. However, the presence of specific investments can make the emergence of loyalty more likely even in cases when loyalty is not a direct consequence of specific investment (case 2). To explain how, I must look beyond economics to social psychology and moral philosophy. Finally, loyalty may emerge even in the absence of specific investments. This is an acknowledgement of the limits of the approach proposed here. No doubt loyalty may sometimes be beyond economic analysis.

If elements in the social environment that affect specific investment decisions can be identified, the emergence and persistence of loyalty can be explained. In the second part of this article I focus on the case where loyalty is due to specific investment. I construct a simple model to explore how changes in external circumstances may affect specific investments (and therefore loyalty), and ultimately exit and voice. I envisage a situation where the transferability of investments changes due to an external factor. A paradoxical dynamic of loyalty is identified: a lower degree of specificity may lead to a weakening of loyalty in the short run but a strengthening of loyalty in the long run. I conclude by sketching an empirical application of the model.

2. A SUMMARY OF HIRSCHMAN’S THEORY

Hirschman was led by his analysis of firm-customer and intra-organizational relationships to identify two fundamental preference revealing and incentive mechanisms: exit and voice. Exit corresponds to market behaviour as perceived by neoclassical economics: market actors always choose their most preferred alternative, and if they become dissatisfied with a good, a service or a contractual relationship, they will opt for another. Hirschman’s innovation was to draw the attention of economists to the fact that exit is not the only mechanism of preference revelation and incentive giving. A consumer or a member of an organization may be able to improve their welfare by changing an existing relationship from within rather than leaving it. A customer may call a producer with their complaint, subordinates may ask their superiors to improve their work, etc. Such methods of expressing dissatisfaction are labeled ‘voice’ because, unlike exit, they usually involve communication between the affected parties.

As Hirschman amply illustrates (1978, 2008), the categories of exit and voice have relevance beyond consumer markets and economic organizations.
People who wish to cooperate can more or less always choose between these two mechanisms. We often try to improve our personal relationships, civil associations or political communities ‘from within’. Except for extreme situations, we can also quit these relationships. Exit is the most common strategy in some relationships, while voice prevails in others. Hirschman argues that a fundamental difference between politics and the economy is that the former is mainly characterized by voice, while exit dominates the latter. However, he emphasizes that both mechanisms are present in both spheres and their joint analysis will provide a better understanding of both.

The parallel consideration of exit and voice led Hirschman to the concept of loyalty. Exit and voice may be substitutes or complements. They are substitutes inasmuch as exit opportunities may make voice less attractive. They are complements inasmuch as the exit option strengthens the bargaining power of an individual who raises his voice. If exit is absolutely impossible, voice may not be seriously considered since it is not backed up by the threat of leaving the group. By contrast, if there are plenty of alternative relationships to choose from, no-one may bother to engage in costly voice activities. This may explain, for example, why a totalitarian party as well as small parties in a multiparty system may be less sensitive to internal criticism than large parties in a two-party democracy.

Hirschman identifies loyalty as a type or trait: ‘loyal’ people suffer some sort of psychic cost – a ‘cost of disloyalty’ (1970: 136) – when they quit a group or relationship. According to this definition, loyalty raises the cost of exit. As a consequence, it increases the probability of voice but diminishes the bargaining power of those who engage in voice. Very strong loyalty is undesirable because it prevents group members from even considering the option of exit and robs them of internal bargaining power. An intermediate level of loyalty may, however, facilitate a healthy equilibrium between exit and voice by creating a situation in which ‘there should be the possibility of exit, but exit should not be too easy or too attractive as soon as deterioration of one’s own organization sets in’ (83). In a nation state, for example, a restrained form of patriotism may be desirable.

Hirschman argues that loyalty is socially useful if exit threatens to undermine voice that is efficient from the group’s perspective (i.e. it would increase the group’s welfare). This threat is gravest if those who are likely to leave first can raise their voices most effectively.

Hirschman also identifies another aspect of loyal behaviour. Some loyal people tend to participate in action aimed at changing the activities or goals of their group, while others are simply unwilling to exit and suffer in silence. That is, loyalty has two ‘components’: the passive component is always
there, and it may (or may not) be accompanied by the active component.\textsuperscript{4} Loyalty is usually more than a mere reluctance to leave a relationship. It refers to a positive commitment to improving the welfare of other parties in a relationship (Barry 1974). The active component of loyalty may reinforce the favourable effect of (moderate) passive loyalty on the equilibrium between exit and voice. While passive loyalty functions as an ‘exit tax’, active loyalty can be interpreted as a ‘voice subsidy’ (Gehlbach 2006).

3. RELATIONSHIP-SPECIFIC INVESTMENT AS A CAUSE OF LOYALTY

Hirschman stops short of explaining where and why loyalty emerges and what influences its strength. So the criticism that loyalty serves as an ‘ad hoc equation filler’ in his theory has merit (Barry 1974). To include loyalty into economic analysis, loyalty must be grasped as rational behaviour. Individual decisions that accompany the emergence and survival of loyalty must be identified. For this purpose, I define loyalty as perseverance in a relationship in spite of difficulties, which may be accompanied by a willingness to act in order to improve the relationship from within. This is in line with Hirschman’s idea that loyalty functions as an ‘exit tax’ and occasionally also as a ‘voice subsidy’. However, I move beyond assuming that loyalty is merely a behavioral type and bring economic analysis into the examination of decisions that underlie loyalty.

3.1. Relationship-specific investments

An idiosyncratic or relationship-specific investment creates a (specific) asset which has higher value in a given relationship than in the best alternative relationship. That is, it loses some of its value outside the relationship (Milgrom – Roberts 2005).\textsuperscript{5} For our purposes, specific investments in human capital have special importance. Learning and cooperative experience in a relationship often lead to the accumulation of specific human assets. Indeed, economic theories of human capital tend to stress the role of specific

\textsuperscript{4} The distinction is suggested by Balázs (1997), although with a somewhat different meaning.

\textsuperscript{5} An investment may be specific to a transaction rather than the parties to a transaction, who may change. For example, human capital might be specific to a machine, no matter who owns that machine. I ignore this complication here.
knowledge in economic relationships. Human capital is usually defined as income-generating capacity (Rosen 1987). However, the concept can be defined more broadly as capacity to engage in productive cooperation. The deepest and most general cause of the specific nature of human capital is man’s limited cognitive capacity, which makes his knowledge highly context-specific. The concept of bounded rationality may be helpful here. As Herbert Simon puts it:

“the bounded rationality of humans does not allow us to grasp the complex situations that provide the environments for our actions in their entirety. The first step in rational action is to focus attention on specific (strategic) aspects of the total situation… One dimension of simplification is to focus on particular goals, and one form of focus is to attend to the goals of an organization or organizational unit” (1991: 37).

Bounded rationality ties our knowledge to our own situation and limits it to our closer environment. This is especially so when dispersed, practical, time and space specific forms of knowledge are needed (Hayek 1995), as opposed to abstract, scientific knowledge.

The economic role of this type of knowledge was explained by Alchian (1984) in discussing team production. Technical advantages are not the only explanation why teamwork is more productive than individual work. In fact, the value of a successful group partly derives from the value of creating a successful group. The source of cooperative surplus is the group members’ specific investment in getting to know each other and establishing methods of cooperation. This may sound tautological, since cooperation is impossible without acquaintance or some modus vivendi. In fact, what I want to argue is that learning about and cooperating with a person or a group always creates human capital that is to some degree relationship-specific. The more the result of learning or the method of cooperation is tied to a person or a group, the greater the degree of specificity.

In the economy, highly idiosyncratic investments tend to characterize long-term relationships, especially organizations (Williamson 2007, Hart 2006). Beyond the economy, specific investments appear in close personal relationships, such as with families and friendships. The more personal a relationship, the greater the investments that are specific to it. Ben-Porath (1980) notes the following: “The most important characteristic of the family contract is that it is embedded in the identity of the partners, without which it loses its meaning. It is thus specific and nonnegotiable or non-transferable.” (3-4).
Specific investments are also very important for political communities. As Furubotn and Richter (1998: 415) explain (while discussing North’s (1981) ‘neoclassical theory of the state’):

“Citizens of a country usually make much more significant specific investments than, for example, employees of a relationship. They are born and grow up in a country. Their sunk costs include their years of youth spent in their country, while they learn its language, formal and informal rules of behaviour, religion, culture, etc. Moreover, specific investments include resources spent on family, friends and business relationships.”

3.2. The effect of specific investments on exit and voice

Williamson (1975) introduced the concept of idiosyncratic investment into economic theory in order to grasp the logic of durable long-term relationships. A specific investment makes it profitable for parties to hold on to their relationships even in the midst of unexpected, unfavorable changes. Such an investment ‘locks in’ the parties in the relationship. It is therefore straightforward to identify specific investment as an important cause of loyalty; i.e. perseverance in a relationship. The literature following Hirschman introduced the intuition that “the degree of loyalty that one has depends upon... the amount one has invested in that object” (Dowding et al. 2000: 477). The examples of investment discussed include resources invested by employees that are directly related to their jobs (e.g. length of employment, work efforts, non-transferable training experience) or which are workplace-specific in other ways (e.g. having a home near one’s workplace, adopting travelling solutions, workplace friends, making payments into corporate pension funds – as in Farrell – Rusult 1992; ‘investing’ into local schools by sending one’s children there – as in Dowding – John 2007). However, these authors do not generalize these examples or follow up their intuitions using theory. I think the primary reason for this failure is that they do not use the concept of relationship-specific investment. The rather narrow field of management science which deals with customer satisfaction and loyalty has gone one step further by identifying ‘consumers’ irreplaceable asset generated from past transactions’ as having ‘strong effects on their loyalty responses’ (Chiou – Pan 2009: 327 – emphasis added). However, no attempt was made to theorise the mechanism.

In fact, relationship-specific investments have exactly the same effects on exit and voice decisions as loyalty, according to Hirschman. First, such
an investment increases the value of the existing relationship relative to alternatives. Thus, the relative price of exit increases compared to staying, as if there were an ‘exit tax’. The investment also induces more voice: it increases the potential value of the relationship so the stake is greater. Second, some specific investments raise the value of the relationship by increasing the effectiveness of voice mechanisms. Like active loyalty, they provide a ‘subsidy’ to voice.

It is worth clarifying how specific investment and voice relate to one another. Voice efforts are themselves a special category of specific investments since they bring benefits only within the relationship. Nevertheless, it seems sensible to distinguish between activities that are directly aimed at improving a relationship and other investments that increase its value. Decisions about the former and the latter can usually be distinguished and analysed separately. Like Hirschman, I shall continue calling the former ‘voice’ and restrict the use of the term ‘specific investment’ to the latter.

Some specific investments in this restricted sense may work through voice. Others increase the value of the relationship independently of voice; i.e. even if the investor remains passive. Therefore, following the distinction between active and passive loyalty, one can distinguish between voice-related and voice-independent specific investments. An example of voice-related investment is the establishment of a bilateral governance structure that facilitates cooperation between two firms (Williamson 1979). An example of voice-independent specific investment may be a supplier’s relocation closer to one manufacturer and farther away from the latter’s rivals. This example suggests, however, that specific investments are probably never completely voice-independent. Geographical proximity may, for example, diminish communication costs within the relationship. Nevertheless, from the exit-voice perspective it is useful to distinguish between the two ideal types.

### 3.3. The complex relationship between specific investments and loyalty

The discussion so far raises the question whether the concept of loyalty could not simply be ‘replaced’ by relationship-specific investment in Hirschman’s theoretical framework. Such a replacement may seem further justified by the observation that specific investment is a regular feature of those relationships where loyalty is typically expected. In the economy, loyalty is usually associated with organizations. One of Hirschman’s key examples is ‘organizational loyalty’. Simon (1991) compares organizations and markets
and clearly associates loyalty with the former. As noted, Williamson invented the concept of idiosyncratic investment with the very purpose of explaining hierarchical governance, which he thought was the defining feature of modern economic organizations. Outside the economy, loyalty is typically associated with intense personal relationships and political communities; i.e. those social relations that are characterized by highly specific investments.

However, it would be wrong to equate the presence of specific investment with loyalty, i.e. perseverance in a group or relationship. The two concepts relate to each other in more complex and more interesting ways. Loyalty has other *rational aspects* beyond the undertaking of specific investments, which must be examined. Several rational decisions can be identified that play a role in the establishment and preservation of loyalty.

Hirschman made a few scattered remarks about the origins of loyalty. I shall first discuss these and then go on to analyse other rational decisions that may underlie loyalty. By drawing on economics, social psychology and moral philosophy, I shall show that all of these decisions are, to some extent, linked to specific investments.

Hirschman states that loyalty can be explained by “less rational, though far from wholly irrational” factors (38). At several points, he suggests that the irrational component dominates. He emphasizes that loyalty is especially important when voice activities require time, effort and innovativeness. If this is so, then loyalty is needed precisely when it is irrational from the individual’s perspective. Moreover, he calls loyalty a special form of attachment; i.e. an unreflected emotion.

**Emotions.** Without doubt, emotional attachment often accompanies loyalty. It is questionable, however, that non-rational attachment plays such an important role in securing loyalty as presumed by Hirschman. As Balázs (1997) points out, in the absence of rational thinking, emotions are more easily extinguished. The test of loyalty is persistence in a crisis, which is more securely supported by a will that can partly rely on rational arguments than by ‘blind loyalty’. Perseverance often requires the suppression of instinctive emotions (Royce 1908). In fact, there is some contradiction between Hirschman’s claim that *moderate* loyalty is socially desirable and his statement that loyalty is individually *irrational* when it is really needed. It is unclear how an irrational emotion can be moderated. There are two ways out of this contradiction. First, rational decisions are just as important components of loyalty as emotions. Second, emotions associated with loyalty may not be entirely devoid of rationality.

As far as the rationality of emotions is concerned, perhaps it is not inappropriate to speak of an ‘economy of emotions’. One can nurture friendly
emotions only towards a small number of people (as was noted by Adam Smith). One must therefore ‘specialise’ one’s friendly emotions within a few relationships. Similarly, one must choose among friendly, erotic, patriotic and other emotions. For example, the emotional aspect of friendship can be perceived as the investment of one’s ‘emotional capital’ which brings an ‘emotional benefit’ mainly in one relationship.

**Trust.** As far as the non-emotional, rational aspects of loyalty are concerned, I shall first consider what Hirschman considered the (only?) rational component of loyalty. He emphasized that a loyal person has trust in a relationship and is optimistic about its future. Loyal members of a group often believe that they are able to influence the functioning of the group. Should they not have significant influence, they still expect that “someone will act or something will happen to improve matters” (78). Otherwise, loyalty is unlikely to survive. These two types of trust – the first based on a belief in one’s own capabilities, the other on a belief in others’ actions – correspond clearly to active and passive loyalty, respectively.

Trust within limits may be rational if it is based on past experience. As explained above, *acquaintance* with another person and *cooperative experience* are specific assets. The decision to trust may be rational even in the absence of such assets if a relationship can be maintained in the long run. Simulations of repeated two-person prisoners’ dilemmas (e.g. Axelrod 1984, Vanberg – Congleton 1992) prove that ‘prudent moral’ strategies that combine initial trust and subsequent reciprocity provide higher payoffs than opportunism in a variety of circumstances. Hence, it may pay to offer trust at the beginning of a relationship and cooperate even in the absence of positive past experience. However, this assumes that it is a reasonable expectation that the relationship will continue. The likelihood of continuation is greater if there are specific investments which induce the parties to remain in the relationship.

Besides the decision to trust, there are other important rational decisions that should be included in a comprehensive explanation of loyalty (Ewin 1992). If I hold on to a relationship in a crisis because I trust (based on my past experience or future expectations) that better times will come, I act like a prudent investor. Loyalty beyond such calculating behaviour (i.e. staying even if I expect further deterioration or personal suffering) is not necessarily irrational, however.

**Morality.** What can make such behaviour rational? The answer is partly suggested by the etymology of the word loyalty. Its Latin root is *lex*, i.e. law.
The reason for my loyalty to you may be that I owe you something or have an obligation to you. My obligation may result from my own promise, an action of yours in the past which calls for consideration, a moral law (e.g. between a parent and a child) or simply gratitude (Ewin 1992). It is rational to fulfil a moral obligation if its violation (i) activates social sanctions (e.g. loss of reputation or ostracism), or (ii) feelings of guilt which serve as internalised sanctions (Ellickson 1991). Thus, social or internal sanctions may explain loyalty even in the absence of trust. Both passive and active forms of loyalty can be explained in this way: I may feel obliged either simply to persist in a relationship or to contribute actively to its improvement.

What is the relationship between morality and specific investments? The moral obligations I take upon myself may (and perhaps should) be influenced by my ability to help others. This ability, in turn, depends on my knowledge of the other person and our shared experience, i.e. specific investment.

Identity. Identity is another concept which many social scientists and philosophers associate with loyalty. People’s self-perception or identity is partly shaped by the social groups to which they belong. Internalization of group membership changes motivations and often leads to behaviour which contradicts narrow self-interest. The welfare of those who identify with a group is linked to the welfare of the group (Van Vugt – Hart 2004: 586). If group members consider the likely effects of their behaviour on their group mates, they may choose to stay (and perhaps raise their voice) even when their narrower self-interest would dictate otherwise. Moreover, identity may influence the range of people towards whom we feel morally obliged (Oldenquist 1982). So identification may also have an indirect effect on loyalty by linking one’s sense of justice or gratitude to a group or relationship.

Identification with a social group always presumes to some extent acquaintance with the group, which is a specific investment. In fact, the noted link between identity and moral obligations is perhaps best explained and justified by the fact the specific investments may increase the effectiveness of our moral endeavors.

Stability of choices. Comte-Sponville (1998) argues in his treatise on virtues that all forms of loyalty, or as he puts it, fidelity, are ultimately based on one’s fidelity to oneself: one’s previous thoughts and intentions. If I do not have a record of a certain past and do not accept it as mine, I cease to be a consistent actor and moral subject. “Fidelity is the virtue of the Same, by which the Same exists or survives” (32). In the more mundane language of economics, fidelity is the constancy of preferences.
Economists usually ignore the possibility that preferences might change due to methodological reasons. If they acknowledge it, they seek to explain concrete and well-defined changes in preferences (Becker – Stigler 1989). The problem of changeability itself is hardly ever discussed. Economists implicitly assumes that the task of preserving one’s self-identity is accomplished successfully. By contrast, if the starting point for analysis is the assumption of changeable preferences, ‘loyalty’ can be identified as a special personal type: one with a stable preference ordering. This concept of loyalty is in line with Hirschman’s theory. If an individual previously chose relationship A rather than relationship B, they are, ceteris paribus, more likely to persist in relationship A if their preference ordering is stable rather than changeable, and the preference relation between A and B reverses with a positive probability.

Loyalty to oneself may be a matter of personal disposition but is also, to some extent, a matter of rational choice. Rational decisions which increase one’s self-loyalty include (i) a conscious suppression of weakness of will (see Elster 1979), and (ii) gathering and digesting information to provide better and more secure foundations for our evaluations. The latter is again a form of specific investment. For example, if an alternative (not yet examined) relationship becomes available, the leaving of an existing relationship for this new one requires at least some investment in gathering additional information about the new alternative. In other words, the previous investment in information about the existing relationship loses some of its value.

To summarize our findings, loyalty may be based on emotional attachment and trust as well as moral motives, identification with a group and the constancy of preferences. I have argued that specific investments can be linked to each of these factors of loyalty. Although these factors do not presuppose specific investments, they would be rather weak in their absence. Acquaintance with the other parties in a relationship and the accumulation of cooperative experience are especially important.

It should also be noted that large specific investments do not automatically lead to loyalty through these intermediate factors. Investments that are specific to a marriage, a business relationship or a country may be accompanied by different degrees of trust, morality, identity or preference stability. It is perhaps most appropriate to say that specific investments create a favorable environment for the emergence of loyalty, or increase the probability of loyal

---

6 A couple of exceptions who at least raise the issue are Hahn (1996) and Skog (2000). In political science, the volatility of electoral preferences is a much discussed topic. See Huckfeldt (1983).
behaviour. If a relationship is characterized by highly specific investments, parties: (i) have a more secure base on which to fund their trust; are more likely to obtain specific knowledge and experience which; (ii) helps them identify with others in the relationship; (iii) makes their moral commitment reasonable; (iv) explains their emotional attachment; and, (v) mitigates the volatility of their choices. By contrast, if the presence of specific investment is negligible (e.g. because the role of non-transferable information and experience is negligible) then the ease of exit weakens the incentives to trust, reduces the likelihood of identification, casts doubt on the reasonability of focusing moral obligations on the relationship and weakens steadfastness in one’s choices.

In sum, three analytical cases can be distinguished: (1) loyalty is the result of specific investment; (2) loyalty is the result of one or several (intermediate) factors which are strengthened by specific investments; and (3) loyalty emerges independently of specific investments. In reality, loyalty of course tends to be a mixture of these analytically clear cases. To put it differently, loyalty may be explained partly by specific investments, partly by other factors strengthened by specific investments and it may also have a component which cannot be explained in this way.

Finally, this discussion of specific investments and loyalty would not be complete if I did not mention the possibility of an inverse relationship between the two concepts. A specific investment requires a ‘contractual guarantee’ which protects the investor against the other party’s efforts to appropriate the return (or quasi-rent) from the investment. Loyalty (not founded on the specific investment in question) may serve as such a guarantee. A contracting party’s ability to appropriate (or his bargaining power) depends on the size of loss they would suffer upon leaving the relationship. Loyalty as an ‘exit tax’ increases this loss and weakens the party’s bargaining power. Of course, there may be other guarantees than loyalty. So loyalty is not a prerequisite for specific investment (at least theoretically).

**4. HOW DOES A CHANGE IN INVESTMENT SPECIFICITY AFFECT EXIT AND VOICE? A FORMAL MODEL**

Among the three cases distinguished above, ‘loyalty as a result of specific investment’ represents the analytically simplest case. If the sole explanation for loyalty, i.e. perseverance in a relationship, is the presence of specific investment, loyalty can be practically replaced by specific investment as an equilibrator of exit and voice in Hirschman model. The main advantage
of this replacement is that it makes room for an analysis of the emergence and strength of loyalty. I present a simple model to illustrate how changes in the social environment may affect decisions about specific investments and ultimately exit and voice. More concretely, I assume that relationships become more similar and, as a result, investments become less specific. One could think of several real-world applications: (i) productive technologies used by different firms become more standardized so human capital becomes less firm-specific; (ii) smaller communities in a society become culturally more homogeneous; or, (iii) it becomes easier for potential business partners to establish contact with each other (e.g. information technologies or travel conditions improve), etc. I will elaborate on only one application: European integration makes national political communities more homogeneous and therefore investments by citizens become less country-specific.

I assume that an individual who participates in a relationship can make two types of relationship-specific investment. One type is voice-related – it increases the value of the relationship by facilitating voice. It raises the benefit from any given voice effort. The other type is voice-independent – its positive effect on the value of the relationship is independent of voice. The assumption that both types of investment are present makes room for the analysis of both as well as their interaction. I distinguish between the short run and the long run. In the short run, investment levels are fixed. They are the consequences of past investment decisions which were taken when the change in specificity was not yet foreseeable. In the long run, investment levels can be re-optimized in response to changes in specificity. First, I explore how changes in specificity affect exit and voice decisions in the short run when investment levels are fixed.

4.1. Short-run adjustments

(i) Assume that the levels of voice-related investment \( (I) \) and voice-independent investment \( (Y) \) are given.  
(ii) The investor learns about the usefulness of investments in the existing relationship \( (R(I) = I^a + b, \ R(I) = I^2 + b) \), and the best alternative relationship \( (R(I) = I^a + b, \ R(I) = I^a + b) \).  
(iii) The investor then decides whether to exit or stay in the existing relationship.  
(iv) Finally, they decide about their voice effort in the chosen relationship.

When the individual makes the exit/stay decision, they weigh the return of their investments in both relationships as well as the level of welfare they can enjoy in each relationship irrespective of their investments. Assume that these welfare levels are stochastic variables, and the individual learns about
their values just before they make the exit/stay decision. By this assumption, investment levels influence the *probability* of exit. (This assumption is useful because, as we shall see, it provides a common framework for analyzing short-term and long-term effects.)

The individual’s short-run optimization problem can be solved by backward induction. The optimal voice effort \((v)\) in any relationship is:

\[
\max_v \text{NB} = B(v, s^j I) - C(v), \quad j = e, a,
\]

where \(\text{NB}\) is the net benefit from voice, \(B\) is the gross benefit from voice, and \(C\) is the cost of voice. Assume that the cost of voice is an upward-sloping and convex function of the voice effort: \(C_v > 0, C_{vv} > 0\). The benefit from voice increases with the level of voice-related investment and the voice effort at a diminishing rate: \(B_i > 0, B_{ii} < 0\); and \(B_v > 0, B_{vv} < 0\). Finally, voice-related investment increases the marginal benefit from voice effort: \(B_{sv} > 0\).

This last assumption implies that the less useful an investment is in a relationship (the smaller \(s^j\)), the lower the optimal level of voice effort \((v)\) and, as a consequence, the smaller the net benefit from voice \((\text{NB})\) (see *Graph 1*).

**Graph 1. Optimal Voice Effort when the Utility of Investment is Low \((s)\) or High \((s)\)**

The maximization problem above determines the optimal value of \(v\) and the maximum net benefit from voice, for given values of \(I\) and \(s^j\). This net benefit can be interpreted as the maximum return on voice-related investment for a given \(s^j\):
\( R(s^I_i) = NB(v^*, s^I_i) \).

The individual was assumed to consider two things when they chose between the two relationships: (1) the return on their investments in each relationship; and, (2) their welfare in each relationship, which is independent of their investments. Let the stochastic variable \( \zeta \) denote this welfare in the existing relationship, and the variable \( \zeta \) in the alternative relationship. If the return on voice-independent investment is \( W(s^I_i Y) \), the individual faces the following alternatives when they decide whether to exit or stay:

1. Overall welfare in the existing relationship:
   \[ W(s^I_i Y) + R(s^I_i) + \eta \]
2. Overall welfare in the alternative relationship:
   \[ W(s^a_i Y) + R(s^a_i) + \zeta \]

Before the values of \( \zeta \) and \( \zeta \) become certain, the probability of exit is:

\[
\Pr[W(s^I_i Y) + R(s^I_i) + \eta < W(s^I_i Y) + R(s^a_i) + \zeta] = \Pr[W(s^I_i Y) - W(s^I_i Y) + R(s^a_i) - R(s^I_i) < \zeta - \eta]
\]

**Voice-related investment becomes less specific**

First, assume that voice-related investment becomes less specific. More concretely, its usefulness in the alternative relationship increases. Since \( s^I_i \) is now greater, the probability of exit is higher. The individual is more likely to exit if their voice-related investment becomes less specific.

By assumption, \( I \) is fixed. Since \( s^I_i \) does not change, the individual will not alter their voice effort if they choose to stay.

**Voice-independent investment becomes less specific**

Second, consider the case when voice-independent investment becomes less specific. Equation (5) above shows that if \( s^I_i \) increases, the probability of exit increases too. Since the usefulness of the voice-independent investment in the existing relationship does not change, voice effort \( (v) \) will remain the same if the individual chooses to stay.
Summary of short-run effects

In the short run, if the usefulness of either type of investment in the alternative relationship increases, the consequences are the same. As existing investments become less specific, the individual (i) will have a stronger incentive to exit but (ii) will not change their voice effort if they choose to stay.

4.2. Long-run adjustments

Voice-related investment becomes less specific

Now assume that the individual foresees that the specificity of their investments will decrease and adjusts their investment levels accordingly. First, consider the case when their voice-related investment becomes more useful in the alternative relationship (while the usefulness of voice-independent investment does not change).

The change in specificity now affects not only the probability of exit and the voice effort but also the level of investments. The individual chooses the values of $I$ and $Y$ when they already foresee (correctly) their usefulness in the two relationships. To keep the analysis simple, I assume that the degree of usefulness of each investment type in the existing relationship is one: $R(I) = I^a + b$. This means that the individual can adjust the composition of their investment to changes within the existing relationship so that its return remains the same. For example, the assumption is valid if homogenization does not affect the cost of gathering information that helps the successful management of a relationship, although the composition of information needed may change. Since $R(I) = I^a + b$, the notations for the alternative relationship can be simplified: $s_y = s_y^a$, and $s_y = s_y^a$.

Primary effects

First, consider how the optimal level of voice-related investment, $I$, is affected by an increase in its external usefulness, $s_y$. The mechanism is somewhat complicated even in this simple model. To separate the various effects, I first assume that the level of voice-independent investment is fixed. Later on, I will lift this restriction and allow $Y$ to respond. The optimal level of $I$ is given by:

\[
\max_I (1 - \Pr_{\text{exit}})[\bar{W}(Y) + R(I) + E(\eta)] + \Pr_{\text{exit}}[\bar{W}(s_y Y) + R(s_y I) + E(\zeta)] - I
\]
where \( \Pr_{exit} = \Pr[R(I) - R(s, I) + \overline{W(Y)} - W(s, Y) < \zeta - \eta] \) is the probability of exit, while \( \Delta s > 0 \) and \( \Delta s > 0 \) are the conditional expected values of welfare levels.

The first-order condition is:

\[
(7) \quad FOC' = (1 - \Pr_{exit}) R'(I) + \Pr_{exit} s_r R'(s, I) + \frac{\partial \Pr_{exit}}{\partial s_r} [W(s, Y) - \overline{W(Y)} + R(s, I) - R(I) + E(\zeta) - E(\eta)] = 1
\]

\[
\text{I.} \quad \downarrow \quad 0 \quad \uparrow \quad \uparrow \quad ? \quad \text{III.} \quad \uparrow
\]

Component I represents the return on the investment if the individual stays. Component II represents the return if the individual exits. Components III reflects that the level of investment influences the probability of exit. The reason for this is that a change in \( \Delta s \) leads to a change in the difference between the two returns, \( R(I) - R(s, I) \), which in turn alters the attractiveness of the exit option.

The first-order condition helps to track the effects of a higher \( s_I \) on optimal \( I \). As we have seen, in the short run a higher \( s_I \) (i.e. improved utility in the alternative relationship) increases the probability of exit. This effect is also present in the long run, provided that the return on the investment does not fall too rapidly as \( I \) increases: \( W'(Y) > s_r W'(s, Y) \) (this condition is satisfied if, for example, \( R(I) = I^a + b \), where \( 0 \leq a \leq 1 \).) In such a case, \( (1 - \Pr_{exit}) \) decreases in the first component, while \( \Pr_{exit} \) increases in the second component (the arrows below the first order condition signal the effects of \( \Delta s > 0 \) on the respective components for any fixed level of \( I \).) This implies that the investor chooses with greater probability the relationship in which part of the return will be lost. As a consequence, they are better off reducing their investment level in the long run.

At the same time, a higher \( s_r \) implies that the individual’s investment becomes more relevant for a broader range of relationships. The overall expected return on investment increases. Therefore, they are better off increasing the level of investment. The second factor in component II of the first-order condition reflects this: unless the return falls rapidly as investment increases, \(^7\) the expression \( s_r R'(s, I) \) rises with \( s_r \).

\[
\frac{\partial s_r R'(s, I)}{\partial s_r} = R'(s, I) + s_r R''(s, I) > 0
\]

\(^7\) The exact condition is:
Finally, the third component shows that a comprehensive analysis ought to consider one more effect: the value of $s$, which influences how the level of investment affects the probability of exit. However, the sign of the influence is uncertain:

$$\frac{\partial^2 \Pr_{exit}}{\partial I \partial s}$$ may be either positive or negative.

There is no general answer as to how the optimal level of voice-related investment is affected by an increase in its external usefulness. The sign of the effect depends on the shape of the return function and the probability distribution of the difference between the welfare levels in the existing and alternative relationships. However, it is intuitively more likely that the optimal level of investment will increase. Let’s focus on the first two mechanisms (the first two components in the first-order condition) and ignore the third one, whose effect is very uncertain and probably relatively weak, anyway. The first and second effects have opposite signs so their joint effect may be positive or negative. In ordinary circumstances, the first effect is likely to dominate, however. An improvement in external usefulness is unlikely to increase the probability of exit to such an extent that it would be worth reducing the level of investment.

The second question of interest is whether exit becomes more or less probable. The probability of exit when the level of investment is at an optimum $I = I^*$ is as follows:

$$\text{Pr}_{exit} = \Pr[R(I^*) - R(s, I^*) + W(Y) - W(s, Y) < \zeta - \eta].$$

This probability increases if a higher $s$ reduces the value of the left-hand side of the formula in brackets ($R(I^*) - R(s, I^*) + W(Y) - W(s, Y)$). By differentiating this expression according to $s$, we get the following:

$$[R'(I^*) - s, R'(s, I^*)] \frac{\partial I^*}{\partial s} - R'(s, I^*) I^*.$$

On the one hand, the improved external usefulness of the investment makes the exit option more valuable, as in the short run ($- R'(s, I^*) I^* < 0$). On the other hand, it may encourage increased investment (if $0 < \frac{\partial I^*}{\partial s}$). If investment is higher, a greater absolute value of return is lost as the individual exits.  

---

8 Provided the constraint in the previous footnote holds.
Therefore, the first effect which induces exit is augmented by a second effect which induces the individual to stay. Ultimately, the second effect will dominate and the probability of exit will fall if
\[ 0 < \frac{\partial l^*}{\partial s} \] is sufficiently high. Otherwise, the sign of the short-run effect (but not its strength) will remain the same in the long run: the probability of exit will increase.

**Secondary effects**

How does this result change if \( Y \) is allowed to respond to a higher external usefulness of \( I \) in the long run? The optimization problem is now the following:

\[
\max_{I,Y} (1 - \Pr_{\text{exit}})[W(Y) + R(I) + \eta] + \Pr_{\text{exit}}[W(s, Y) + R(s, I) + \zeta] - I
\]

The first-order conditions are:

\[
(11) \quad FOC^I = (1 - \Pr_{\text{exit}})R(I) + \Pr_{\text{exit}} s, R'(s, I) + \frac{\partial \Pr_{\text{exit}}}{\partial I}[W(s, Y) - W(Y) + R(s, I) - R(I) + E(\zeta) - E(\eta)] = 0
\]

\[
(12) \quad FOC^Y = (1 - \Pr_{\text{exit}})W'(Y) + \Pr_{\text{exit}} s, W'(s, Y) + \frac{\partial \Pr_{\text{exit}}}{\partial Y}[W(s, Y) - W(Y) + R(s, I) - R(I) + E(\zeta) - E(\eta)] = 0
\]

The second condition clarifies how a decrease in the specificity of voice-related investment indirectly affects the level of voice-independent investment. The effect is mediated by the probability of exit, which changes with \( s \) and \( I \). Let’s assume again that the first two components of the first-order condition dominate and the third effect can be ignored.\(^9\) If the probability of exit increases as a primary consequence of \( \Delta s \), the expected return on \( Y \) falls. Consequently, its optimal level will fall, too (since \( W'(Y) > s, W'(s, Y) \) can reasonably be assumed to hold). By contrast, if the probability of exit

\(^9\) Again, the effect of \( s \) and \( I \) through the third component is unclear because \( \Delta[R(I^*) - R(s, I^*)] \) is difficult to sign, as we have seen.
decreases, the expected return on $Y$ as well as its optimal level will increase. How does a change in the optimal level of $Y$ as a secondary effect modify the probability of exit and the voice effort? On the one hand, it is easy to see that it reinforces the primary effect on the probability of exit. If exit becomes more likely, optimal $Y$ falls, which makes exit even more likely. By contrast, if exit becomes less likely, optimal $Y$ increases, which strengthens the incentive to stay. The secondary effect on voice effort is more complicated. If optimal $Y$ increases, the option to stay becomes more attractive, which increases the expected return on $I$, hence its optimal level. If optimal $Y$ falls, the probability of staying decreases, which reduces the expected return on $I$, hence its optimal level. The secondary effect of a change in the optimal level of $Y$ may be relatively strong if $Y$ is more sensitive to changes in the probability of exit and/or makes up a larger part of the individual’s investment portfolio than $I$. In such a case, the ultimate consequence of a change in the specificity of voice-related investment may turn out to be the opposite of the primary effect. Voice effort may ultimately fall (increase) even if the direct effect of changing specificity was positive (negative).

**Voice-independent investment becomes less specific**

To what extent are the effects on the probability of exit and the voice effort different if voice-independent (rather than voice-related) investment generates a higher return outside the existing relationship and the individual optimizes their investment levels?

**Primary effects**

If voice-independent investment becomes more useful in the alternative relationship, there is no direct effect on the level or specificity of voice-related investment. Hence, the optimal level of voice effort does not change if one chooses to stay. The effect on the probability of exit is basically the same as in the previously examined case. Simply, $Y$ must be substituted for $I$, and $s_y$ for $s_r$. Higher external usefulness makes exit more attractive, whereas an increase in the optimal level of investment has the opposite effect. The latter effect is stronger if $0 < \frac{\partial Y^*}{\partial s_y}$ is greater. The probability of exit may increase or decrease, depending on the value $\frac{\partial Y^*}{\partial s_y}$. 
Secondary effects

A change in the probability of exit indirectly modifies the expected return on voice-related investment and, as a consequence, its optimal level. Therefore, a fall in the specificity of voice-independent investment influences the optimal level of the other investment type. If the primary effect is an increase in the probability of exit, optimal $I$ will fall. In the opposite case, it will rise. That is, as a secondary effect of $\Delta s_y$, the optimal level of voice-related investment and, as a consequence, the voice effort if one stays may rise or fall (but is unlikely to remain unchanged). Like in the previously examined case, the sign of the primary and secondary effects on the probability of exit will be the same.

5. INTERPRETATION OF RESULTS

Loyalty can be defined as perseverance in a relationship in spite of difficulties. According to Hirschman, loyalty increases the cost of exit and may decrease the cost of voice. Hence, it makes exit less probable and voice efforts more worthwhile. I have argued that relation-specific investment provides one explanation for the emergence of loyal behaviour. The larger the investment, the greater the quasi-rent that would be lost in the case of exit. Moreover, some relationship-specific investments are voice-related. They raise the value of the relationship through voice activities: they increase the benefit from a given voice effort and thereby make voice more worthwhile.

When loyalty is (mainly) the result of relationship-specific investment, the emergence of loyalty can be traced back to changes in the social environment. The model presented explores such a mechanism. The starting point is an exogenous change in the homogeneity of relationships. Investments made in existing relationships become more useful in alternative relationships. Intuition suggests that weaker asset specificity leads to more frequent exit and less intensive voice activities. The model partly proves this intuition correct but exposes a more complex dynamic.

In the short run, the model shows that if investment becomes less specific due to an increase in external usefulness, the individual will indeed have a greater incentive to exit. However, his voice effort will not change because if he chooses to stay the return on voice-related investment will not change. This may seem to be an artefact of the model which assumes that voice effort is chosen after the exit/stay choice has been made. One may object that a change in the probability of exit would influence voice effort before the exit/
stay decision. However, such an effect is considered in the long run version of the model in which voice-related investment is allowed to adjust to the probability of exit. One can think of voice-related investment to be voice effort intended to bring benefit in the long run.

The key difference between the short run and the long run is that the optimal level of specific investment may change. Under plausible assumptions, lower specificity will lead to an increase in the optimal level. First, a higher external return will increase the probability of exit and thereby the probability that part of the return will be lost. Hence, the effect on the optimal level of investment is negative. Second, a higher external return means that if the investor leaves for the alternative relationship, they will lose a smaller part of the return. Hence, this second effect on the optimal level of investment is positive. If the rate of return does not fall very rapidly as the level of investment rises, the net effect will be positive. The probability of exit will not increase to such an extent as to make it worthwhile to reduce the level of investment. A higher level of investment provides an incentive to stay because a higher investment implies a greater absolute loss of return in case of exit. In sum, the short run incentive to exit is counterbalanced (and may even be reversed) by a long term incentive to stay, which results from a higher optimal level of investment.

An important difference between voice-related and voice-independent investment is observed in the long run. Their direct effect on the voice effort when the individual has chosen to stay is different. The optimal level of voice effort in the existing relationship is only raised by an increase in voice-related investment, which makes effort more productive. However, a change in the specificity of voice-independent investment will have an indirect effect on the voice effort. It changes the probability of exit and thereby the expected return on voice-related investment, which determines the optimal voice effort in the existing relationship.

Our results show (somewhat paradoxically) that more homogeneity is likely to result in weaker loyalty in the short run but not necessarily in the long run. Higher external usefulness may increase the overall productivity of investment. The resulting higher optimal level of investment makes exit less attractive and, if voice-related, makes voice efforts more fruitful.

6. APPLICATION TO THE EUROPEAN UNION

A possible application of the model to some major changes propelled by European integration can be briefly sketched (for elaboration, see Mike
This process makes citizens’ investments less specific to their country. Thus, it affects their political activity (voice) as well as their migration (exit) choices. The model presented can help us understand these mechanisms. The distinctions between the short run and the long run and between voice-related and other specific investments appear to be particularly important. The temporal distinction must be considered because it takes time for the level and composition of investments to adjust to changes in the environment. Therefore, European integration is likely to have different effects in the short and the long run. One of the main conclusions to be drawn from the model is that a weakening of loyalty and an increase in exit is to be expected primarily in the short run. In the long run, a growth in the overall level of investment by citizens may counterbalance the lower country-specificity of these investments that already appear in the short run.

The model shows that the consequences of the integration process on national loyalty are far from certain. However, the conditions on which the signs of its effects on exit, voice and loyalty depend can be identified. One of the critical factors is the speed of adjustment in the level and composition of investments. The other factor is the degree of optimal adjustment. These factors could be the subject of empirical research in the future.

REFERENCES

Mike, Károly (2010), Contributions to the Economics of Federalism – with Applications to European Integration [in Hungarian]. PhD thesis, Corvinus University of Budapest.


