A Game Theoretic Analysis of Second-language Acquisition and the World's Language Regime

Author. Kentaro HATSUMI

Address. Faculty of Policy Studies, Kansai University, Suita, Osaka, 564-8680, Japan

Email. hatsumi *atmark* kansai-u.ac.jp

This Version. June 2, 2014

Keywords. bilingualism, foreign language policy, sociolinguistics, large coalition formation game, interdisciplinary study

JEL Classification Codes. C72, D85, F51, F68

Acknowledgement. I thank Atsushi Kajii, Takeshi Ogawa and seminar participants at Tokyo Tech for helpful comments. Any errors are my responsibility.

Abstract. With a strong motivation to represent the current world's language regime of English domination, we construct a game theoretic model in which there are multiple countries with their own languages and each citizen can gain from additional communication in her secondly acquired language. We demonstrate that in any equilibrium, a hegemonic language, which is a language that all citizens in other countries want to study, emerges. We study in detail the conditions under which the language of a country without the largest population becomes hegemonic. Such an equilibrium is more likely to exist if the size of the population of a country that is not the largest increases, or if the ratio of the gain from the additional communication in the second language to the cost for acquisition increases.

1 Introduction

Language has a quite important role in various social aspects, including politics and economics. In political contexts, the classical and influential work by Anderson (1983) suggests that print media written in one language was critical in constituting a sense of unity of a modern nation state. In economic contexts, even within a country, differences in spoken language often work as entry barriers preventing language minorities from obtaining job opportunities. Even in the United States or the United Kingdom, which are countries that are believed to have respect for diversities of ethnicities and languages, empirical studies such as Tainer (1988), Dustmann and Fabbri (2003), and Bleakley and Chin (2004) suggest that workers who understand English can gain significantly higher wages. It is worthwhile for social scientists to constitute models that focus on language and related issues and analyze them in detail.

Recently, one of the most interesting and important phenomena in sociolinguistics is the focus on English as a foreign language that people all over the world want to study. Table 1 lists the top 12 languages by the number of native speakers. As can be seen, English is in third place. It is difficult to calculate the actual number of speakers of a language as a second language.[1] Ethnologue arguably reports the number of second-language speakers of English as more than 430 million, which is the largest among all languages and much greater than the second-placed language, Mandarin, which has 178 million second-language speakers.[2]

[Table 1 around here]

Concentration on English is observed more clearly in several specific

[1] In sociolinguistics, a distinction is often made between second language and foreign language. In Graddol (1997), second language is an additional language used in some daily contexts more frequently than just a foreign language. In spite of the importance of this difference, we do not differentiate between second language and foreign language. We refer to both as second language, and construct a simple big-picture model of second-language acquisition.

[2] See Lewis et al. (eds., 2013). Mandarin is a dialect or a group of dialects of Chinese that has 840 million native speakers.

fields. Graddol (1997) reports that, even in the early 1990s, English was ranked first in terms of the number of titles published in the form of books. English accounted for 28% of book titles in the world, much more than Chinese, which was ranked second with 13.3%. Miniwatts Marketing Group (2013) reports on the Web site "Internet World Stats" that English is ranked first by the number of Internet users by language, with a 26.8% share. Moreover, many scholars state that English is now the universal language in many scientific disciplines, and they are forced to use English for scientific activities even though, in the 20th century, a number of quite important researches were published in several other languages such as French, German, and Russian.[3]

The historical reason why English has become so hegemonic is often considered to be the combination of Britain's colonial expansion and the rise of the United States as a superpower in the 20th century.[4] Then the questions arise: *Can we represent this situation in a formal way? And can we analyze the formal model and find some intuition about the stability of the current world's language regime?*

Because there is very little research on the discipline of game theory to model second-language choices, especially focusing on the problem why English has become hegemonic, we construct a model with very basic assumptions, especially on the payoff structure. The model we construct is as follows. There are four countries.[5] This is for the sake of simplicity, and

[3]See, for example, Graddol (1997), Ammon (ed., 2001), and Drubin and Kellogg (2012).

[4]See Graddol (1997). Phillipson (1992) also emphasizes in his influential book that the United Kingdom and the United States maintain the continuous inequalities between English and other languages for political and industrial purposes. Graddol (2006) notes that "the English language teaching sector directly earns nearly £1.3 billion for the UK in invisible exports and our other education related exports earn up to £10 billion a year more."

[5]A "country" throughout this paper is more like a cultural community based on a language than a political nation.

we can easily increase the number of countries.[6] Each country has its own language and a continuum of agents (or citizens). Each agent can speak her native language and can learn, at most, one foreign language. Each agent can gain linearly from the number of foreign agents with whom she can communicate in her acquired second language. Each agent has her own effort cost to learn a foreign language that is hidden at first. Even though this is a two-stage game for introducing the uncertainty of cost, it is essentially simultaneous. In the first stage, all agents choose their second language, and in the second stage, after the cost is revealed, they decide whether actually to study. We only care for pure strategies. We employ a relatively strong equilibrium concept, a partially strong subgame perfect Nash equilibrium, which prohibits all coalitional deviations within a country to study other languages in the first stage. We obtain all equilibria of this game.

What we can obtain by analyzing this model in detail is the following. In any equilibrium, there exists a hegemonic language, which is a language that all agents in other countries want to study (Proposition 2). Without any condition, an equilibrium exists in which the language of the country with the largest population is hegemonic (Proposition 3). For languages of the countries with the second, third, and fourth largest populations to become hegemonic, we need conditions. These conditions are essentially the same. As the relative size of the population of the second, third, or fourth largest country to that of the largest country increases, any equilibrium in which the language of the second, third, or fourth largest country is hegemonic is more likely to exist (Propositions 5 and 7). If the ratio of gain by the additional communication to effort cost increases, then any equilibrium in which the language of the second, third, or fourth largest country is hegemonic is more

[6] In the three-country case, a different result is obtained than for the case of four or more countries. That result is discussed in Section 5.

likely to exist (Propositions 6 and 7).

Theoretically speaking, this study is in the class of large coalition formation games, with the special feature that an agent can belong to two coalitions through her native and second languages.[7] After obtaining Proposition 3, the model turns out to be a large coordination game in which agents choose a hegemonic language. This coordination game is a rather simple one. However, it may be worthwhile to examine the conditions for an equilibrium in which the language in a country whose population is not the largest is hegemonic.

We focus only on the difference between population sizes among countries as a causative role for the propositions obtained. Even though the model fails to address important issues, such as the degree of similarity among various languages in the world and the variety of motivations people have for wanting to learn second languages, we succeed in obtaining a rough large-scale picture of the world language regime.[8] This model suggests that as the population of the largest country increases relative to those of other countries, a language in the second, third, or fourth largest countries is hardly likely to be hegemonic, in equilibrium. This indicates that the number of native Chinese speakers may have an important role in the hegemony of English in the future. In some specific areas, such as the sciences and popular culture, if the number of native Chinese speakers with specific knowledge and skills becomes large enough, this may possibly affect the language regime in such areas. This model also suggests that reducing the learning cost of English as well as increasing the gain from it may help to maintain the hegemony of English.

[7]For a survey of recent studies on coalition formation games, see Ray (2008).

[8]Although the current model depicts the language hegemony problem among larger countries, some important linguistic problems within relatively smaller countries such as diminishing minority languages cannot be followed.

The structure of the rest of the paper is as follows. Section 2 reviews the previous game theoretic models studying behavior of second-language acquisitions. Section 3 constructs the formal model. Section 4 shows several consequences of analyzing this model in detail. Section 5 discusses several remaining problems. Section 6 concludes.

2 Related Literature

Selten and Pool (1991) is the closest to our model in the sense of the theoretical framework. In their model, there are multiple countries and each country has a continuum of agents. An agent has her native language and can learn foreign languages. She can gain from the additional communication in her foreign languages, but has an effort cost to learn them. Agents within a country can conduct some coalitional actions. Those features are the same as ours. They actually construct quite a general model. An agent can learn more than one foreign language, and her payoff and cost structures can be picked up from varieties of functions rather than just linear ones. However, their model is so general that they can state only the existence of an equilibrium, while our model is more tractable and derives some implications for understanding the linguistic situations in the world. Ginsburgh et al. (2007) modify Selten and Pool (1991) with a different motivation from ours. They construct a two-country model with empirical evidence on how the sizes of countries and the similarity of languages affect the behavior of foreign language acquisitions.[9]

Laitin (1993) studies the English hegemony problem directly with game theory. He focuses on the problem within a country, and suggests that the

[9] Church and King (1993) constructs a simple two-country model in which agents gain from the additional communication in their foreign language similarly to Selten and Pool (1991). They focus on the efficiencies of equilibria.

coordination game with a tipping point by Schelling (1978) represents the actual situation in Ghana well. Educating children in English is currently in equilibrium in Ghana, and indigenous languages are expected to disappear. To enable indigenous languages to survive, he claims that the number of children educated in indigenous languages must be larger than the tipping point. Caminal (2010) constructs a model of industrial organization with the English hegemony problem. He studies choices of languages by firms through which they produce cultural goods such as books and films. Wickström (2005) constructs a dynamic model of a society with two languages and studies the conditions for the survival of bilingualism. Mélitz (2007) studies translation activities of literature, both theoretically and empirically, and discusses how the domination of English in such activities affects welfare.

3 The Model

There are four **countries**, X_1, X_2, X_3, and X_4. Let $N := \{1, 2, 3, 4\}$, which is the set of the suffixes of the countries. The suffixes of arbitrary countries are often denoted by a, b, c, and d. In each country, there is a continuum of **agents** (or citizens). **Population** of each country is the length of the continuum of agents. Let x_a denote the population of X_a. We normalize that $x_1 := 1$ and assume that $x_1(= 1) > x_2 > x_3 > x_4$. Each country has a different **language**. Let L_a denote the language of X_a. An agent i in X_a can speak L_a and can learn at most one other language.

Payoff. There are agents who actually study a foreign language and understand it and agents who do not study it. **Active** agents are the former agents who actually study a foreign language. It is endogenously determined whether an agent becomes active in the game studied in this paper. The

net payoff of an active agent i in X_a learning L_b is $r(x_b + y_b) - t_i$, where r is a constant for all agents in all countries, y_b is the total length of active agents studying L_b who are not in X_a, and t_i is the type of agent i that represents effort cost to study and understand a foreign language. We simply refer to t_i as **cost** for agent i. t_i is independently, identically, and uniformly distributed over $[0, 1]$. We assume that $0 < r < 1/(1 + x_2 + x_3 + x_4)$ to ensure that $0 < r(x_b + y_b) < 1$. The first term of the net payoff $r(x_b + y_b)$ is referred to as the **gross payoff**, and is denoted by u_{ab}. Note that the gross payoff is the same for all active agents in X_a studying L_b. The interpretation of the gross payoff is that the gain from studying a foreign language depends on the length of agents with whom she can communicate in her second language. r represents the ratio of this gain to the cost. If r is large, the gain from communication is large relative to the cost. When i does not study a foreign language, she need not pay her cost t_i, and so her net payoff is normalized to 0.

We assume that the net payoff increases linearly with the length of agents with whom she can communicate in her secondly acquired language. If some agents in other countries study her native language, she may also gain from the additional length of agents with whom she can communicate in her native language. Thus, her actual utility varies depending on the length of agents studying her native language even though her net payoff is normalized to 0 when she does not choose to study a foreign language, and to $r(x_b + y_b) - t_i$ when she chooses to study one. We need to take this normalization into account, especially when welfare analysis is executed.

Note that, in the model, an agent can increase her payoff by communication with other agents in a foreign language even if they can speak her native language. On the other hand, she can gain nothing directly from the positive length of agents learning the same language from the same country. Remember that agents within a country can easily and deeply understand each other in their native language, and so communication among them in their second language brings relatively small benefits. However, conversation with a foreigner is usually imperfect, and therefore communication in a second language is quite helpful. The setting of the payoff in this paper approximates these situations well. Later, in Section 5, we discuss the payoff structure in detail.

Timing of action and information revelation. This is a two-stage game. In the first stage, an agent chooses a foreign language. At this stage, an agent does not know costs for herself and other agents. After the first stage, her cost to study a foreign language is revealed not only to her, but also to all agents. In the second stage, an agent decides whether to study the language chosen in the first stage, *i.e.*, she decides whether to be active or not. All of the game structures except the costs are common knowledge among all agents in all countries.

Equilibrium Concept. A **coalition** is a continuum of agents in a country with a length smaller than or equal to that population. We employ a stronger version of the subgame perfect equilibrium as the equilibrium concept, which requires partial stability to some coalitional deviations in the first stage.

In the second stage, all agents' choices should constitute a Nash equilibrium, *i.e.*, any active agent $i \in X_a$ studying L_b should satisfy the following condition:

$$u_{ab} - t_i > 0.$$

We refer to this condition as **individual rationality** (**IR**). Agents who are not active in the second stage should not satisfy **IR**.

In the first stage, based on the prediction for the outcome in the second stage, agents choose languages to maximize their expected net payoffs. Their choices should constitute a Nash equilibrium that is even stable against deviation by any coalition.[10] Remember that when agents form a coalition in the first stage, they do not know each agent's cost, including their own, in the coalition. They only know that agents' costs are uniformly distributed over $[0, 1]$ in the coalition.

There are several possibilities for coalitional deviations in one country; for example, a change in the foreign language study policy by the central government of a country and a boom in some foreign culture and the related foreign language in a country. Thus, it is natural to request stability against coalitional deviations in equilibria.

Hereafter, we refer to this equilibrium concept in this two-stage game just as **equilibrium**

Now, we start preliminary investigations. First, we present how to calculate actual proportions of active agents and their final net payoffs when decisions by all agents in the first stage are given. Let $a, b, c, d \in N$. Let $z_{ab} \in [0, x_a]$ be a coalition in X_a in which members choose to study L_b in the first stage. Let $p_{ab} \in [0, 1]$ be the proportion of active agents in z_{ab}.

IR requires that the net payoff for an active agent i with t_i should be larger than 0. Let $\bar{t}_{ab} = u_{ab}$, which is the threshold of cost for agents in X_a to decide whether to learn L_b in the second stage. If t_i is smaller than $\bar{t}_{ab}$, then an agent i actually studies L_b. Otherwise, she does not study L_b in the second stage. As we assume that t_i is uniformly distributed over $[0, 1]$, the proportion of agents whose costs are smaller than $\bar{t}_{ab}$ is $\bar{t}_{ab}$, *i.e.*, $p_{ab} = \bar{t}_{ab}$.

[10]The equilibrium concept in the first stage is a partially "strong Nash equilibrium" presented by Aumann (1959). This partiality means that we only account for a coalition of agents in one country.

Thus, we have $p_{ab} = u_{ab}$.

Note that the gross payoff of agents in X_a studying L_b is $u_{ab} = r(x_b + p_{cb}z_{cb} + p_{db}z_{db})$. Thus, we also have $p_{ab} = r(x_b + p_{cb}z_{cb} + p_{db}z_{db})$. This equation states that p_{ab} depends on x_b and the lengths of active agents in other countries studying L_b, *i.e.*, $p_{cb}z_{cb}$ and $p_{db}z_{db}$. Similarly, we have $p_{cb} = r(x_b + p_{ab}z_{ab} + p_{db}z_{db})$ and $p_{db} = r(x_b + p_{ab}z_{ab} + p_{cb}z_{cb})$. By solving these three equations simultaneously, we have the actual proportion of active agents in X_a studying L_b as

$$p_{ab} = u_{ab} = \frac{rx_b(rz_{cb}+1)(rz_{db}+1)}{1 - r^2(z_{ab}z_{cb} + z_{cb}z_{db} + z_{db}z_{ab}) - 2r^3 z_{ab}z_{cb}z_{db}}.$$

Hence, both p_{ab} and u_{ab} are functions of z_{ab}, z_{cb} and z_{db}. When functional forms are appropriate, we write $p_{ab}[z_{ab}, z_{cb}, z_{db}]$ and $u_{ab}[z_{ab}, z_{cb}, z_{db}]$ to represent the terms in the above formula.[11]

The next lemma states some properties of p_{ab} and u_{ab} that are useful in the following investigation.

Lemma 1. Let $a, b, c, d \in N$. Let an agent i in X_a choose L_b in the first stage. (i) If z_{cb} increases, then p_{ab} and u_{ab} increase. (ii) If z_{ab} increases and $z_{cb} > 0$ or $z_{db} > 0$, then p_{ab} and u_{ab} increase.[12]

Proof of Lemma 1. (i) Suppose that z_{cb} increases. Then, the numerator of p_{ab} increases since $r > 0$, $x_b > 0$ and $z_{db} \geq 0$. If $z_{ab} > 0$ or $z_{db} > 0$, the denominator of p_{ab} decreases, otherwise it is unchanged. Thus, p_{ab} increases.

(ii) Suppose that z_{ab} increases. The numerator of p_{ab} is unchanged. If $z_{cb} > 0$ or $z_{db} > 0$, the denominator of p_{ab} decreases, otherwise it is unchanged. Thus, p_{ab} increases as long as $z_{cb} > 0$ or $z_{db} > 0$. □

[11] The first variable of p_{ab} and u_{ab} should be the proportion of agents in X_a choosing L_b in the first stage. Because in the formula, z_{cb} and z_{db} are symmetric, we need not consider the order of the second and third variables of the functions.

[12] We can define p_{ab} and u_{ab} even if $z_{ab} = 0$.

It is easy to interpret Lemma 1. As we assume that the gross payoff of studying a foreign language depends on the length of agents with whom she can communicate in her second language, (i) of Lemma 1 is straightforward. The increase of agents in the same country choosing the same foreign language in the first stage may not directly affect her gross payoff. At first, it increases the proportions of active agents studying the same foreign language in other countries. Then, it increases the length of agents with whom she can communicate, and increases her gross payoff, as a consequence, which is (ii) of Lemma 1.

Now, we consider the equivalent condition for an equilibrium in the first stage. The expected net payoff of agent i in X_a deciding to study L_b in the first stage is $\int_0^{\bar{t}_{ab}}\{u_{ab}-t_i\}dt_i$. This is because if $t_i < \bar{t}_{ab}$, i will choose to be active in the second stage, otherwise she does not choose to be active and her payoff is 0. This is also because t_i is uniformly distributed over $[0,1]$. Then, the expected net payoff can be rewritten as:

$$\int_0^{\bar{t}_{ab}}\{u_{ab}-t_i\}dt_i = \int_0^{u_{ab}}\{u_{ab}-t_i\}dt_i = [u_{ab}t_i - \frac{t_i^2}{2}]_0^{u_{ab}} = \frac{r^2u_{ab}^2}{2}.$$

To check an equilibrium, we need to compare the expected net payoff in the equilibrium with those in deviations. Since $u_{ab} > u'_{ab} \iff r^2u_{ab}^2/2 > r^2u'^2_{ab}/2$, we only need to compare the gross payoffs to confirm an equilibrium. Hence, in the first stage in an equilibrium, for any coalition $0 < z'_{ac} \leq x_a$,

$$u_{ab}[z_{ab}, z_{cb}, z_{db}] \geq u_{ac}[z'_{ac}, z_{bc}, z_{dc}].$$

We refer to this condition as **incentive compatibility** (**IC**). What we are actually interested in are the choices of agents in the first stage. Hence, we rather focus on the first stage and **IC** condition in the following investigation.

4 Results

4.1 properties of equilibria

In this section, we derive all equilibria and investigate their properties. First, we define two notions related to equilibria.

A **countrywide coalition** (**CC-**) **equilibrium** is an equilibrium in which all agents in each country choose the same language in the first stage. A **CC-deviation** is a deviation from an equilibrium by all agents in a country. The next remark is relatively obvious; however, it is quite useful in the following investigation.

Remark 1. To qualify as an equilibrium, we only need to check the **IC** condition against CC-deviations, *i.e.*, $u_{ab}[z_{ab}, z_{cb}, z_{db}] \geq u_{ac}[x_a, z_{bc}, z_{dc}]$, where z_{ab}, z_{cb}, z_{db}, z_{bc}, and z_{dc} are the coalitions in an equilibrium.

Proof of Remark 1. By Lemma 1 (ii), a CC-deviation brings a higher or equal payoff to a deviation with a smaller coalition for an agent in it. □

The following is the first proposition in this model.

Proposition 1. *An equilibrium is always a CC-equilibrium.*

Proof of Proposition 1. Suppose that there is an equilibrium that is not a CC-equilibrium. Then a positive proportion z_{ab} of agents in X_a chooses L_b and another positive proportion z_{ac} chooses L_c. **IC** implies that $u_{ab} = u_{ac}$.

Note that there is a positive proportion of active agents studying L_b or L_c in countries other than X_a. Suppose not. Then $u_{ab} = rx_b$ and $u_{ac} = rx_c$. Since $x_b \neq x_c$ by definition, this contradicts $u_{ab} = u_{ac}$.

Let L_b be a language that positive proportion of active agents in countries other than X_a studies. Consider a CC-deviation x_a of X_a to choose L_b.

Then, by Lemma 1 (ii), an agent in this coalition gains a higher payoff than in the equilibrium. This is a contradiction. □

Remember that Lemma 1 (ii) states that agents in a country obtain higher payoffs by forming a larger coalition. This induces Remark 1 and Proposition 1, suggesting that we only need to take into consideration the biggest coalition, which is **CC**. Proposition 1 also suggests that because any coalitional deviation is possible, a situation where agents in a country study different languages is quite unstable in this model.

Now, we can summarize situations in terms of directed graph theory.[13] Suppose that countries are **vertices** and the choices of agents of those countries in the first stage are represented by **arcs**. A country where agents stay is represented by the **tail** of the arc, and the country whose language is chosen by the agents is represented by the **head** of the arc.

By Proposition 1, there are four arcs and each country must be the tail of only one arc. The number of possible heads of each arc is three. Thus, there is the possibility of $3^4 = 81$ situations (or graphs).

We divide the situations into three types. **Situation 1** is that any country is head of at most one arc; **Situation 2** is that any country is head of at most two arcs; and **Situation 3** is that there is a country that is head of three arcs. See Figure 1 for the illustrations. Then, we have the following proposition.

[Figure 1 around here]

Proposition 2. *Situations 1 and 2 do not occur in an equilibrium.*

[13] We borrow some elementary terms of directed graph theory to represent situations and do not demand knowledge of directed graph theory. If recent developments in directed graph theory are of interest, see Bang-Jensen and Gutin (2000) for a survey. The terms of graph theory used in this paper follow them. For applications of graph theory to social sciences, see Jackson (2010).

As the proof of Proposition 2 is relatively long, we have placed it in the Appendix. The essence of the proof is as follows. In both situations, without loss of generality, while agents in X_a choose L_b, agents in X_c do not choose L_b. However, because agents in both X_a and X_c essentially have the same preferences in the sense that they gain from the larger number of agents with whom they can communicate in their second languages, the existence of agents in X_a and X_c choosing different languages itself results in a contradiction to **IC**.

By Proposition 1 and 2, if an equilibrium exists, then it belongs to Situation 3. Actually, there are four equilibria in this model and all belong to Situation 3 as long as certain conditions discussed in the following subsection are satisfied.

A language is called **hegemonic** if all agents in all other countries choose it in the first stage. Because any case of Situation 3 has a hegemonic language, there always exists an equilibrium. We can show that an equilibrium in which each of the four languages is hegemonic is unique. We refer to the equilibrium in which L_a is hegemonic as **hegemony** a **(**H_a**-) equilibrium**.

The next proposition states that, in this model, there is an equilibrium without any condition.

Proposition 3. *There is H_1-equilibrium in which all agents in X_2, X_3, and X_4 choose L_1, and all agents in X_1 choose L_2 in the first stage.*

Proof of Proposition 3. We show that H_1-equilibrium surely satisfies **IC**. An active agent in X_1 has gross payoff $u_{12}[x_1, 0, 0] = rx_2$ in the equilibrium. Consider $a \in N\backslash\{1, 2\}$. Suppose that agents in X_1 CC-deviate to X_a. Then, an active agent X_1 has gross payoff $u_{1a}[x_1, 0, 0] = rx_a$. Because $x_2 > x_a$, **IC** for agents in X_1 is satisfied.

Let $b, c, d \in N\backslash\{1\}$. In equilibrium, an active agent in X_b has net payoff

$u_{b1}[x_b, x_c, x_d] = \{r(rx_c+1)(rx_d+1)\}/\{1-r^2(x_bx_c+x_cx_d+x_dx_b)-r^3x_bx_cx_d\}$. Suppose that agents in X_b CC-deviate to X_c. Then, an active agent has net payoff $u_{bc}[x_b, 0, 0] = rx_c$. Obviously $u_{b1} > 1$ and $1 > rx_c$, **IC** for agents in X_b is satisfied. □

The existence of H_1-equilibrium without any condition in this model is not surprising. Because agents gain from having more agents to communicate with, it is quite natural that the language in the largest country can be hegemonic and all agents in other countries study this language. However, the hegemonic language is sometimes not the language in the real world. There may be several historical reasons why a language in a relatively small country becomes hegemonic. In the next subsection, we show that even in this basic model, there is an equilibrium in which a language in a small country becomes hegemonic under some conditions. As these conditions have important roles when we predict the stability of a hegemonic language in the future, the investigation of the details of the conditions is worthwhile.

4.2 Conditions when the hegemonic country is not the largest country

In this subsection, we first focus on the H_2-equilibrium as it is the most possible and tractable equilibrium except the H_1-equilibrium. Later, we show that the H_3- and the H_4-equilibria actually have essentially the same existence conditions.

We first check the conditions under which the H_2-equilibrium exists, and then investigate how changes in the parameters affect the existence of this equilibrium.

Proposition 4. *The H_2-equilibrium exists when all agents in X_1, X_3, and*

X_4 choose L_2 and all agents in X_2 choose L_1 in the first stage if the gross payoff of agents in X_3 in the equilibrium is larger than or equal to their gross payoff when CC-deviating to study L_1, i.e., if $u_{32}[x_3, x_1(=1), x_4] - u_{31}[x_3, x_2, 0] \geq 0$.

Proof of Proposition 4. It is obvious that in this equilibrium, agents in X_1 and those in X_2 have no incentive for a CC-deviation. It is also obvious that agents in X_3 have no incentive to CC-deviate to study L_4 and those in X_4 have no incentive to deviate to study L_3. The net payoff of an active agent i in X_3 in the equilibrium is

$$u_{32}[x_3, 1, x_4] = \frac{rx_2(r+1)(rx_4+1)}{1 - r^2(x_3 + x_4 + x_3x_4) - 2r^3x_3x_4}. \quad (1)$$

Her net payoff in a CC-deviation to study L_1 is

$$u_{31}[x_3, x_2, 0] = \frac{r(rx_2+1)}{1 - r^2x_2x_3}. \quad (2)$$

The net payoff of an active agent j in X_4 in the equilibrium is

$$u_{42}[x_4, 1, x_3] = \frac{rx_2(r+1)(rx_3+1)}{1 - r^2(x_3 + x_4 + x_3x_4) - 2r^3x_3x_4}. \quad (3)$$

Her net payoff in a CC-deviation to study L_1 is

$$u_{41}[x_4, x_2, 0] = \frac{r(rx_2+1)}{1 - r^2x_2x_4}. \quad (4)$$

The condition for agents in X_3 staying in the equilibrium is (1) − (2) ≥ 0 and that for agents in X_4 is (3) − (4) ≥ 0. Note that the numerator of (1) is smaller than that of (3) and the denominators of (1) and (3) are equal. Thus, (1) is smaller than (3). Also note that the numerators of (2) and

(4) are equal and the denominator of (2) is smaller than that of (4). Thus, (2) is larger than (4). Hence, if (1) − (2) ≥ 0 holds, (3) − (4) ≥ 0 always holds. Hence, we only need (1) − (2) ≥ 0 for the necessary condition for the H_2-equilibrium. □

We investigate how each parameter of this model affects the existence of the H_2-equilibrium. Of course, in this investigation, the necessary condition $u_{32}[x_3, 1, x_4] - u_{31}[x_3, x_2, 0] \geq 0$ plays an important role. Table 2 contains numerical examples of how changes in parameters such as r, x_2, x_3 and x_4 affect the existence of equilibria including the H_2-equilibrium. As Table 2 suggests, we have the following proposition.

[Table 2 around here]

Proposition 5. *The H_2-equilibrium is more likely to exist if x_2, x_3, or x_4 increases, in the sense that $u_{32}[x_3, 1, x_4] - u_{31}[x_3, x_2, 0]$ increases in x_2, x_3, and x_4.*

Even though showing the increase of $u_{32}[x_3, 1, x_4] - u_{31}[x_3, x_2, 0]$ in x_2, x_3, and x_4 is not difficult, they are slightly technical and therefore we have included them in the Appendix. We can show that the derivative of $u_{32}[x_3, 1, x_4] - u_{31}[x_3, x_2, 0]$ with respect to x_2 is positive. We can obtain the proofs more straightforwardly in the cases of x_3 and x_4.

It is very intuitive that if the population of X_2 increases, the H_2-equilibrium is more likely to exist. Actually, the increase of x_2 positively affects both $u_{32}[x_3, x_1, x_4]$ and $u_{31}[x_3, x_2, 0]$, whose precise expressions are in (1) and (2). Proposition 5 suggests that surely the effect of the increase of x_2 on $u_{32}[x_3, x_1, x_4]$ is larger than that on $u_{31}[x_3, x_2, 0]$.

The fact that the increase of x_3 or x_4 makes the existence of the H_2-equilibrium more likely is rather interesting. x_3 obviously positively affects both $u_{32}[x_3, x_1, x_4]$ and $u_{31}[x_3, x_2, 0]$ since the CC-deviation by the agents in X_c itself is the key for the condition for the existence of the equilibrium. It is shown that the effect of x_3 on $u_{32}[x_3, x_1, x_4]$ is larger. x_4 positively affects only $u_{32}[x_3, x_1, x_4]$. Overall, increases in the populations of relatively small countries whose agents choose the hegemonic language make the existence of the H_2-equilibrium more likely.

The following proposition states that essentially, if r increases, then the H_2-equilibrium is more likely to exist.

Proposition 6. *The H_2-equilibrium is more likely to exist if r increases, in the sense that, given r, r', such that $r < r'$, then if the H_2-equilibrium exists in r, it exists in r'.*

Figure 2 depicts typical graphs of $u_{32}[x_3, 1, x_4] - u_{31}[x_3, x_2, 0]$ with respect to r. The outcome of this function should be less than 0 when r is close to 0 since the first derivative of (19) with respect to r is negative when r is close to 0. The proposition comes from the fact proved in the Appendix that the second derivative of (19) with respect to r is positive.

[Figure 2 around here]

r is interpreted as the ratio of gain by the additional communication by the learned hegemonic language relative to effort cost. Thus, if r becomes larger, the number of actual agents in each country increases. This makes the existence of the H_2-equilibrium more likely as a consequence. This proposition is interesting with respect to language policy. If the country with the hegemonic language wishes to maintain their position, a policy of reducing the language learning cost is quite effective. Also, this is consistent with intuition.

Now, we focus on the H_3- and the H_4-equilibria. The next proposition is relatively surprising. The essentially symmetric arguments to the existence of the H_2-equilibrium hold on both the H_3- and the H_4-equilibria.

Proposition 7. *Let $a, b \in \{3, 4\}$. There is the H_a-equilibrium in which all agents in X_1, X_2, and X_b choose L_a and all agents in X_a choose L_1 in the first stage if the gross payoff of agents in X_2 in the equilibrium is larger than or equal to their gross payoff when deviating to study L_1, i.e., if $u_{2a}[x_2, x_1(=1), x_b] - u_{21}[x_2, x_a, 0] \geq 0$.*

The necessary condition $u_{2a}[x_2, 1, x_b] - u_{21}[x_2, x_a, 0] \geq 0$ and the symmetric argument to Proposition 5 suggest that the increase of x_2 brings the emergence of the H_a-equilibrium. However, if x_2 increases, $u_{1a}[1, x_2, x_b] - u_{12}[1, 0, 0]$ decreases and the CC-deviation by agents in X_1 to study L_2 may possibly occur. Proposition 7 states that we need not consider the latter condition. As we explain in the proof in the Appendix, if the necessary condition of this proposition $u_{2a}[x_2, x_1, x_b] - u_{21}[x_2, x_a, 0] \geq 0$ holds, then $u_{1a}[1, x_2, x_b] - u_{12}[1, 0, 0] \geq 0$ holds.

Note that in the proofs for Propositions 6 and 7, we never use the fact that $x_2 > x_3$ and $x_2 > x_4$. Hence, the symmetric statements to Propositions 5 and 6 also hold on both the H_3- and the H_4-equilibria. Increases of x_2, x_3, x_4, and r contribute to the existence of the H_3- and the H_4-equilibria.

5 Discussion

5.1 The effect of the number of countries

In this article, we have investigated the model of second-language choice with four countries. In this subsection, we discuss what happens when the number of countries changes.

First, we note the three-country case. In this case, there is only one equilibrium, which is the H_1-equilibrium.

Proposition 8. *Suppose that there are three countries X_1, X_2, and X_3. Then, there is a unique equilibrium. It is the H_1-equilibrium in which all agents in X_2 and X_3 choose L_1, and all agents in X_1 choose L_2 in the first stage.*

The proof is in the Appendix. One of the main purposes of this paper is determining the condition for the existence of equilibrium in which the hegemonic country is not X_1, and checking for stability. Therefore, we investigated the four-country case in the main part.

On the other hand, if the number of countries is greater than or equal to five, the investigation and all propositions for the case of four countries hold.

5.2 Payoff function reconsidered

We define the net payoff of an agent i in X_a studying L_b as $r(x_b+y_b)-t_i$. As already mentioned when defining this payoff structure, an agent can increase her payoff by communication with other agents in a foreign language even when they can speak her native language. If we define the net payoff as $r(x_b + y_b - p_{ba}z_{ba}) - t_i$, then an agent i can gain only by the additional communication with agents who cannot speak her native language. This setting may be clear and appropriate from the viewpoint of an abstract coalition formation game.

Even though we agree to the importance of modelling using the latter payoff condition, there are two reasons why we adopt the payoff structure of the current model in this paper. One reason is mathematical tractability. Since the payoff structure of our model is simple, we can obtain p_{ab} and

u_{ab} by solving only three simultaneous equations. On the other hand, if we impose the latter alternative net payoff, naively the number of simultaneous equations rises to $4 \cdot 3 = 12$ and the model becomes more complicated and loses its tractability. The second reason is that our simple model well approximates some particular language acquisition situations. Additional communications with foreigners in the secondary language are often quite helpful for deeper understanding, even though they can understand her native language. The payoff structure in the current paper is closer to these situations.

5.3 Welfare analysis

As has already been mentioned when defining net payoff, the actual utility of an agent i in X_a depends not only on her net payoff but also on y_a, the number of foreign agents with whom she can communicate in her own language. If we take this into account, it is obvious that all four equilibria in this model are Pareto efficient because an agent with quite a high cost never learns a foreign language, and she can only increase her actual utility if her native language becomes hegemonic.

5.4 The ratio of gain to cost reconsidered

We have assumed throughout the paper that r represents the ratio of the gain from understanding a foreign language and communicating with other agents in this additional language to its effort cost. In the real world, r varies among linguistic relations.[14] For example, native English speakers can learn French or Spanish more easily than Chinese, Korean, or Japanese.

[14] Isphording and Otten (2013) provide empirical studies suggesting that the similarity of two languages positively affects not only the learning ability of each of the languages but also their bilateral trade volumes.

Moreover, because possessing the hegemonic language is usually considered to bring welfare improvement for agents in the country, it may often promote campaigns to increase the gain/cost ratio to learn their language for foreigners via media and schools. The targets of these campaigns are sometimes not all countries, but one country in particular.

Suppose that for an agent in X_1, the learning cost of L_4 is lower than those of other foreign languages, and vice versa. For an agent in X_2, the learning cost of L_3 is lower than those of other foreign languages, and vice versa. Then, our model with this extension easily suggests the emergence of two blocks of languages and the disappearance of a hegemonic language. This is an interesting direction of analysis, especially when considering language policies of a country.

However, in the real world, many believe that the hegemonic language, English, exists. At least at the moment, the model of this paper with a constant r over all countries can be considered as a good approximation of the actual situation in the world.

One implication of this model is that if the population of the largest country x_1 is much larger than that of the second largest country x_2, then the H_2-, H_3-, and H_4-equilibria are unlikely to exist. Currently, in many areas such as the sciences, English is in the hegemonic position. If Chinese, or any other language, has a large enough number of native speakers who can be considered as qualified scientists, our model suggests the hegemony of English will disappear in the equilibrium. It is very hard to forecast the situation after the English hegemony era. A naive application of the current model implies the emergence of another hegemonic language. Alternatively, there is a possibility that several blocks of languages will appear. More sensitive analyses beyond the current model are anticipated.[15]

6 Concluding remarks

In this paper, we have developed a large-scale game model of second-language acquisition. In an equilibrium for this model, a hegemonic language always exists and we have studied the conditions for the existence of a hegemonic language of which the number of native speakers is not the largest.

As we have employed several simplified assumptions, such as the fact that the gain of an agent is determined only by the number of agents with whom she can communicate in the additionally learned language, there remain many important features of second-language acquisition that have not been studied. In the present model, agents in a country learn the same language, even though in the real world, a significant number of people in various countries study several minor languages with several purposes. This model is quite static, and lacks the dynamics of convergence to an equilibrium with one hegemonic language and its collapse. However, we believe that this model is a simple approximation of the real world with a hegemonic language. We believe that this can contribute to the understanding of the linguistic situation in the world today, and stimulate further research on sociolinguistics via formal modelling approaches.

Appendix

Proof of Proposition 2. **Situation 1.** Let $a, b, c, d \in N$. Situation 1 constitutes two subtypes. Situation 1-i is such that arcs construct a cyclic

[15] In his very influential work, Graddol (1997) predicts that "no single language will occupy the monopolistic position in the 21st century which English has – almost – achieved by the end of the 20th century. It is more likely that a small number of world languages will form an 'oligopoly', each with particular spheres of influence and regional bases."

structure and in this structure, X_a is the head of the arc to X_b, X_b is that of X_c, X_c is that of X_d and X_d is that of X_a. Situation 1-ii is such that if X_a is the head of the arc to X_b, then X_a is the tail of the arc from X_b. See Figure 3 for an illustration. Any other graph does not belong to Situation 1. We suppose that those situations are equilibria and derive contradictions.

[Figure 3 around here.]

1-i. Suppose that Situation 1-i is an equilibrium. The gross payoff of agents in X_a in the equilibrium is

$$u_{ab}[x_a, 0, 0] = rx_b. \tag{5}$$

The payoff of agents in X_a CC-deviating to study L_d is

$$u_{ad}[x_a, x_c, 0] = \frac{r(1 + rx_c)x_d}{1 - r^2 x_a x_c}. \tag{6}$$

IC implies that (5) − (6) ≥ 0, which is

$$rx_b - \frac{r(1 + rx_c)x_d}{1 - r^2 x_a x_c} \geq 0 \iff \frac{x_b}{x_d} \geq \frac{1 + rx_c}{1 - r^2 x_a x_c}. \tag{7}$$

Similarly, the gross payoff of agents in X_c in the equilibrium is

$$u_{cd}[x_c, 0, 0] = rx_d. \tag{8}$$

The payoff of agents in X_c CC-deviating to study L_b is

$$u_{bc}[x_c, x_a, 0] = \frac{r(1 + rx_a)x_b}{1 - r^2 x_a x_c}. \tag{9}$$

IC implies that (8) − (9) ≥ 0, which is

$$rx_d - \frac{r(1+rx_a)x_b}{1-r^2x_ax_c} \geq 0 \iff \frac{x_d}{x_b} \geq \frac{1+rx_a}{1-r^2x_ax_c} \iff \frac{x_b}{x_d} \leq \frac{1-r^2x_ax_c}{1+rx_a}. \tag{10}$$

Since $\{1+rx_c\}/\{1-r^2x_ax_c\} > 1$, (7) implies that $x_b/x_d > 1$. Since $\{1-r^2x_ax_c\}/\{1+rx_a\} < 1$, (10) implies that $x_b/x_d < 1$. These imply a contradiction.

1-ii. Suppose that Situation 1-ii is an equilibrium. As the X_b is the tail of the arc from X_a and X_d is that of X_c similarly to 1-i, we can derive a contradiction in the same way.

Situation 2. The proof for this case is essentially the same as that for Situation 1. In Situation 2, there is a country that is the tails of the arcs from two countries. Let X_b be the heads of the arcs from X_a and X_c. Without loss of generality, let X_a be the head of the arc from X_d. Note that Situation 2 consists of Subsituations 2-i in which X_b is the tail of the arc to X_a, 2-ii in which X_b is the tail of the arc to X_c, and 2-iii in which X_b is the tail of the arc to X_d. (See Figure 4 for an illustration.) Suppose that Situation 2, which can be any of Subsituations 2-i, 2-ii and 2-iii at first, is an equilibrium, and we derive a contradiction.

[Figure 4 around here.]

The gross payoff of agents in X_d in the equilibrium is

$$u_{da}[x_d, 0, 0] = rx_a. \tag{11}$$

The gross payoff of agents in X_d CC-deviating to study L_b is

$$u_{db}[x_d, x_a, x_c] = \frac{r(1+rx_a)(1+rx_c)x_b}{1-r^2(x_ax_c+x_ax_d+x_cx_d)-2r^3x_ax_cx_d}. \tag{12}$$

IC implies that (11) − (12) ≥ 0, which is

$$
\begin{aligned}
rx_a &\geq \frac{r(1+rx_a)(1+rx_c)x_b}{1-r^2(x_ax_c+x_ax_d+x_cx_d)-2r^3x_ax_cx_d} \\
\Longleftrightarrow \quad \frac{x_a}{x_b} &\geq \frac{(1+rx_a)(1+rx_c)}{1-r^2(x_ax_c+x_ax_d+x_cx_d)-2r^3x_ax_cx_d}.
\end{aligned} \tag{13}
$$

The gross payoff of agents in X_c in the equilibrium is

$$
u_{cb}[x_c, x_a, 0] = \frac{r(1+rx_a)x_b}{1-r^2x_ax_c}. \tag{14}
$$

In Subsituations 2-ii and 2-iii, the gross payoff of agents in X_c CC-deviating to study L_a is

$$
u_{ca}[x_c, x_d, 0] = \frac{r(1+rx_d)x_a}{1-r^2x_cx_d}. \tag{15}
$$

IC implies that (14) − (15) ≥ 0, which is

$$
\begin{aligned}
&\frac{r(1+rx_a)x_b}{1-r^2x_ax_c} - \frac{r(1+rx_d)x_a}{1-r^2x_cx_d} \geq 0 \\
\Longleftrightarrow \quad &\frac{(1+rx_a)(1-r^2x_cx_d)}{(1+rx_d)(1-r^2x_ax_c)} \geq \frac{x_a}{x_b} \\
\Longleftrightarrow \quad &\frac{(1+rx_a)(1-r^2x_cx_d)}{1+rx_d-r^2x_ax_c-r^3x_ax_cx_d} \geq \frac{x_a}{x_b}.
\end{aligned} \tag{16}
$$

Note that the left-hand side (LHS) of (13) equals the right-hand side (RHS) of (16). The denominator of the RHS of (13) is smaller than that of the LHS of (16), and the numerator of the RHS of (13) is larger than that of the LHS of (16). Thus, the RHS of (13) is larger than the LHS of (16). This fact, together with (13) and (16), implies a contradiction.

In Subsituation 2-i, the payoff of agents in X_c CC-deviating to study L_a is

$$
u_{ca}[x_c, x_b, x_d] = \frac{r(1+rx_b)(1+rx_d)x_a}{1-r^2(x_bx_c+x_bx_d+x_cx_d)-r^3x_bx_cx_d}. \tag{17}
$$

IC implies that (14) − (17) ≥ 0, which is

$$
\begin{aligned}
&\frac{r(1+rx_a)x_b}{1-r^2x_ax_c} - \frac{r(1+rx_b)(1+rx_d)x_a}{1-r^2x_bx_c-r^2x_bx_d-r^2x_cx_d-r^3x_bx_cx_d} \geq 0 \\
\Longleftrightarrow\ &\frac{(1+rx_a)x_b}{1-r^2x_ax_c} \geq \frac{(1+rx_b)(1+rx_d)x_a}{1-r^2x_bx_c-r^2x_bx_d-r^2x_cx_d-r^3x_bx_cx_d} \\
\Longleftrightarrow\ &\frac{(1+rx_a)(1-r^2x_bx_c-r^2x_bx_d-r^2x_cx_d-r^3x_bx_cx_d)}{(1-r^2x_ax_c)(1+rx_b)(1+rx_d)} \geq \frac{x_a}{x_b}. \qquad (18)
\end{aligned}
$$

The denominator of the RHS of (13) is smaller than that of the LHS of (18) and the numerator of the RHS of (13) is larger than that of the LHS of (18). Thus, the RHS of (13) is larger than the LHS of (18). This fact, together with (13) and (18), implies a contradiction. □

Proof of Proposition 5. $u_{32}[x_3, x_1(=1), x_4] - u_{31}[x_3, x_2, 0]$ is explicitly written as

$$
\begin{aligned}
&u_{32}[x_3,1,x_4] - u_{31}[x_3,x_2,0] \\
=&\{r(x_2-1)+r^2x_2x_4+r^3(-x_2^2x_3+x_2x_4+x_3+x_3x_4+x_4) \\
&+r^4(x_2x_3-x_2^2x_3+x_2x_3x_4-x_2^2x_3x_4+x_2x_4+2x_3x_4) \\
&+r^5(2x_2x_3x_4-x_2^2x_3x_4)\} \\
&/\{(1-r^2x_2x_3)(1-r^2x_3-r^2x_4-r^2x_3x_4-2r^3x_3x_4)\}. \qquad (19)
\end{aligned}
$$

1. By differentiating (19) with respect to x_2, we have that

$$
\begin{aligned}
&\{r+r^2x_4+r^3(-2x_2x_3-x_3+x_4) \\
&+r^4(-2x_2x_3-2x_2x_4x_3+x_4x_3+x_3+x_4) \\
&+r^5(x_2^2x_3^2+x_4x_3^2+x_3^2-2x_2x_3x_4+3x_3x_4) \\
&+r^6(x_2^2x_3^2+x_2^2x_4x_3^2+2x_4x_3^2)+r^7x_2^2x_3^2x_4\} \\
&/\{(1-r^2x_2x_3)^2(1-r^2x_3-r^2x_3x_4-r^2x_4-2r^3x_3x_4)\}. \qquad (20)
\end{aligned}
$$

The denominator of (20) is obviously positive. By the definition of r, $r < 1/(1 + x_2 + x_3 + x_4) \iff 1/r > (1 + x_2 + x_3 + x_4)$. Thus, $r > r^3(1 + x_2 + x_3 + x_4)^2 > r^3(1 + x_2 + x_3 + x_4) + r^4(1 + x_2 + x_3 + x_4)^2(x_2 + x_3 + x_4)$. As $x_2, x_3, x_4 < 1$, the RHS of this inequality is larger than $-\{r^3(-2x_2x_3 - x_3) + r^4(-2x_2x_3 - 2x_2x_4)\}$. Also note that $r^5(-2x_2x_3x_4 + 3x_3x_4) > 0$ since $x_2 < 0$. Thus, the numerator of (20) is positive. Hence, (20) is positive and (19) is increasing in x_2.

2. Suppose that x_3 increases. The two factors of the denominator of (19) $(1 - r^2x_2x_3)$ and $(1 - r^2x_3 - r^2x_4 - r^2x_3x_4 - 2r^3x_3x_4)$ both decrease. Thus, the denominator of (19) decreases. Since $x_2, x_4 < 1$, $\{-x_2^2x_3 + x_3\}$, $\{x_2x_3 + -x_2^2x_3 + x_2x_3x_4 - x_2^2x_3x_4\}$, and $\{x_2x_3x_4 - x_2^2x_3x_4\}$ increase. Thus, the numerator of (19) increases. Hence, (19) increases.

3. Suppose that x_4 increases. Then, $\{1 - 2r^3x_3x_4 - r^2x_3 - r^2x_4 - r^2x_3x_4\}$ decreases. Thus, the denominator of (19) decreases. Now $x_2, x_3 < 1$, $\{x_2x_3x_4 - x_2^2x_3x_4\}$ and $\{x_2x_3x_4 - x_2^2x_3x_4\}$ increase. Thus, the numerator of (19) increases. Hence, (19) increases. □

Proof of Proposition 6. We show that the second derivative of $\{u_{32}[x_3, 1, x_4] - u_{31}[x_3, x_2, 0]\}$ with respect to r is positive. Let F_N denote the numerator of (1), F_D denote the denominator of (1), S_N denote the numerator of (2), and S_D denote the denominator of (2). Then, $\{u_{32}[x_3, 1, x_4] - u_{31}[x_3, x_2, 0]\} = \{F_N/F_D - S_N/S_D\}$, and its second derivative is

$$\begin{aligned}
&\{\frac{2F_N F_D'^2}{F_D^3} - \frac{2F_N' F_D'}{F_D^2} - \frac{F_N F_D''}{F_D^2} + \frac{F_N''}{F_D}\} \\
&- \{\frac{2S_N S_D'^2}{S_D^3} - \frac{2S_N' S_D'}{S_D^2} - \frac{S_N S_D''}{S_D^2} + \frac{S_N''}{S_D}\} \\
=&\{\frac{2F_N' F_D'^2}{F_D^3} - \frac{2S_N S_D'^2}{S_D^3}\} + \{-\frac{2F_N' F_D'}{F_D^2} + \frac{2S_N' S_D'}{S_D^2}\}
\end{aligned}$$

$$+\{-\frac{F_N F''_D}{F_D^2}+\frac{S_N S''_D}{S_D^2}\}+\{\frac{F''_N}{F_D}-\frac{S''_N}{S_D}\}. \tag{21}$$

We show that all four terms in the parentheses in (21) are positive. First, note that $F_N > 0$, $F_D > 0$, $S_N > 0$, $S_D > 0$, and, particularly, that $F_D < S_D$. Also note that $F'_N = x_2 + 2rx_2 + 2rx_2x_4 + 3r^2x_2x_4 > 0$, $F''_N = 2x_2 + 2x_2x_4 + 6rx_2x_4 > 0$, $F'_D = -2rx_3 - 2rx_4 - 2rx_3x_4 - 6r^2x_3x_4 < 0$, $F''_D = -2x_3 - 2x_4 - 2x_3x_4 - 12rx_3x_4 < 0$, $S'_N = 1 + 2rx_2 > 0$, $S''_N = 2x_2 > 0$, $S'_D = -2rx_2x_3 < 0$, and $S''_D = -2x_2x_3 < 0$. Thus, it is sufficient to show that (i) $F'_N F'^2_D > S_N S'^2_D$, (ii) $F'_N F'_D < S'_N S'_D$, (iii) $F_N F''_D < S_N S''_D$, and (iv) $F''_N > S''_N$. Note that in (ii) and (iii), both sides of the inequalities are negative.

(i) $F_N F'^2_D - S_N S'^2_D = \{rx_2(r+1)(rx_4+1)\}\{-2rx_3 - 2rx_4 - 2rx_3x_4 - 6r^2x_3x_4\}^2 - \{r(rx_2+1)\}\{-2rx_2x_3\}^2 > rx_2(r+1)(-2rx_3)^2 - \{r(rx_2+1)\}\{-2rx_2x_3\}^2 = 4r^3(r+1)x_2x_3^2 - 4r^3(rx_2+1)x_2^2x_3^2 > 0$.

(ii) $F'_N F'_D - S'_N S'_D = (x_2+2rx_2+2rx_2x_4+3r^2x_2x_4)(-2rx_3-2rx_4-2rx_3x_4-6r^2x_3x_4)-(1+2rx_2)(-2rx_2x_3) < (x_2+2rx_2)(-2rx_3)-(1+2rx_2)(2rx_2x_3) = -4rx_2(1+r+rx_2)x_3 < 0$.

(iii) $F_N F''_D - S_N S''_D = \{rx_2(r+1)(rx_4+1)\}\{-2x_3-2x_4-2x_3x_4-12rx_3x_4\} - \{r(rx_2+1)\}\{-2x_2x_3\} < rx_2(r+1)(-2x_3) - r(rx_2+1)(-2x_2x_3) = 2r^2(-1+x_2)x_2x_3 < 0$.

(iv) $F''_N - S''_N = (2x_2 + 2x_2x_4 + 6rx_2x_4) - (2x_2) > 0$. □

Proof of Proposition 7. First, we consider the case of $a = 3$ and $b = 4$. We only show that if $u_{23}[x_2, 1, x_4] - u_{21}[x_2, x_3, 0] \geq 0$ holds, then $u_{13}[1, x_2, x_4] - u_{12}[1, 0, 0] \geq 0$ holds. By the similar argument to the proof of Proposition 4, we can easily find that the necessary condition of this proposition implies any other **IC** conditions.

$$
\begin{aligned}
&u_{23}[x_2, 1, x_4] - u_{21}[x_2, x_3, 0] \\
=&\frac{rx_3(r+1)(rx_4+1)}{1 - r^2(x_2 + x_4 + x_2x_4) - 2r^3x_2x_4} - \frac{r(rx_3+1)}{1 - r^2x_2x_3} \\
=&r\{-1 + x_3 + rx_3x_4 + r^2(x_2 + x_4 + x_2x_4 + x_3x_4 - x_2x_3^2) \\
&+ r^3(2x_2x_4 + x_3x_4 + x_2x_3 + x_2x_3x_4 - x_2x_3^2 - x_2x_3^2x_4) + r^4(2x_2x_3x_4 - x_2x_3^2x_4)\} \\
&/(1 - r^2(x_4 + x_2 + x_2x_4) - 2r^3x_2x_4)(1 - r^2x_2x_3)
\end{aligned}
$$

Since we assume that this is larger than or equal to 0, we have that

$$
\begin{aligned}
&x_3 + rx_3x_4 + r^2(x_2 + x_4 + x_2x_4 + x_3x_4 - x_2x_3^2) \\
&+ r^3(2x_2x_4 + x_3x_4 + x_2x_3 + x_2x_3x_4 - x_2x_3^2 - x_2x_3^2x_4) \\
&+ r^4(2x_2x_3x_4 - x_2x_3^2x_4) \geq 1. \qquad (22)
\end{aligned}
$$

$u_{13}[1, x_2, x_4] - u_{12}[1, 0, 0]$ is rewritten as

$$
\begin{aligned}
&r\{-x_2 + x_3 + r(x_3x_4 + x_2x_3) + r^2(x_2x_4 + x_2^2 + x_2^2x_4 + x_2x_3x_4) + 2r^3x_2^2x_4\} \\
&/(1 - (r^2x_4 + r^2x_2 + r^2x_2x_4) - 2r^3x_2x_4).
\end{aligned}
$$

Thus, to show that this is greater than or equal to 0, we need to show that

$$
\begin{aligned}
&-x_2 + x_3 + r(x_3x_4 + x_2x_3) + r^2(x_2x_4 + x_2^2 + x_2^2x_4 + x_2x_3x_4) + 2r^3x_2^2x_4 \geq 0 \\
\iff& x_3 + r(x_3x_4 + x_2x_3) + r^2(x_2x_4 + x_2^2 + x_2^2x_4 + x_2x_3x_4) + 2r^3x_2^2x_4 \geq x_2 \\
\iff& x_3/x_2 + r(x_3x_4/x_2 + x_3) + r^2(x_4 + x_2 + x_2x_4 + x_3x_4) + 2r^3x_2x_4 \geq 1. \qquad (23)
\end{aligned}
$$

To show (23), it is enough to obtain that the LHS of (23) is greater than the LHS of (22). Note that $\{x_3/x_2 + rx_3x_4/x_2\} > \{x_3 + rx_3x_4\}$. By the definition of r, we also have that $rx_3 > r^3x_3(1+x_2+x_3+x_4)^2 > \{r^3(x_3x_4 + x_2x_3) + 2r^4x_2x_3x_4\}$. Hence, we obtain the anticipated inequality.

As we do not use the fact that $x_3 > x_4$ to prove for the case of $a = 3$ and $b = 4$, the same argument can be applied to the case of $a = 4$ and $b = 3$. $\square$

Proof of Proposition 8. By similar arguments to those used in the case with four countries, we can obtain Lemmas 1 and Proposition 1. Thus, we only take into account CC-equilibria. For $a \in \{2, 3\}$, $b \in \{1, 3\}$, and $c \in \{1, 2\}$, let $S(a, b, c)$ denote the situation in which all agents in X_1 choose to study L_a, all agents in X_2 choose to study L_b, and all agents in X_3 choose to study L_c. As S can represent all prospective CC-equilibria, the number of possibilities is $2^3 = 8$. $S(211)$, which is the H_1-equilibrium in the three-countries case, which is obviously an equilibrium. We show that the other seven situations are not equilibria.

In the following three situations, agents in X_3 have incentive to CC-deviate to study L_1. In $S(312)$, we have $u_{32}[x_3, 0] - u_{31}[x_3, x_2] = rx_2 - \{r(rx_2+1)\}/\{1-r^2x_2x_3\} < 0$.[16] In $S(332)$, we have $u_{32}[x_3, 0] - u_{31}[x_3, 0] = rx_2 - r < 0$. In $S(212)$, we have $u_{32}[x_3, 1] - u_{31}[x_3, x_2] = \{r(rx_2 + 1)\}/\{1 - r^2x_2x_3\} - \{rx_2(r + 1)\}/\{1 - r^2x_3\}$

$$= \frac{r(x_2 - 1)(1 - r^2x_2x_3 - r^2x_3 - r^3x_2x_3)}{(1 - r^2x_2x_3)(1 - r^2x_3)} < 0.$$

In the following three situations, agents in X_3 have incentive to CC-deviate to study L_1. In $S(231)$, we have $u_{23}[x_2, 0] - u_{21}[x_2, x_3] = rx_3 - \{r(rx_2 + 1)\}/\{1 - r^2x_2x_3\} < 0$. In $S(232)$, we have $u_{23}[x_2, 0] - u_{21}[x_2, 0] = rx_3 - r < 0$. In $S(331)$, we have $u_{23}[x_2, 1] - u_{21}[x_2, x_3] = \{rx_3(r + 1)\}/\{1 -$

[16] Analogical notations to the four-country case are applied.

$r^2x_2\} - \{r(rx_3+1)\}/\{1-r^2x_2x_3\}$

$$= \frac{r(x_3-1)(1-r^2x_2x_3-r^2x_3-r^3x_2x_3)}{(1-r^2x_2x_3)(1-r^2x_2)} < 0.$$

In $S(311)$, agents in X_1 have incentive to CC-deviate to study L_2 since $u_{13}[1,0] - u_{12}[1,0] = rx_3 - rx_2 < 0$. □

References

Ammon, Ulrich (ed.) (2001), "The Dominance of English as a Language of Science: Effects on Other Languages and language Communities", de Gruyter, Berlin.

Anderson, Benedict (1983), "Imagined Communities: Reflections on the Origin and Spread of Nationalism", Verso, London.

Aumann, Robert (1959), Acceptable Points in General Cooperative N-Person Games, in Albert W. Tucker, R. Duncan Luce (eds.) "Contributions to the Theory of Games IV", Princeton University Press, Princeton.

Bang-Jensen, Jorgen; Gutin, Gregory (2000), "Digraphs: Theory, Algorithms and Applications", Springer-Verlag, Berlin.

Bleakley, Hoyt; Chin, Aimee (2004), Language Skills and Earnings: Evidence from Childhood Immigrants, Review of Economics and Statistics 86 (2), 481-496. DOI: 10.1162/003465304323031067

Caminal, Ramon (2010), Markets and Linguistic Diversity, Journal of Economic Behavior and Organization 76 (3), 774-790. DOI: 10.1016/j.jebo.2010.09.009

Church, Jeffrey; King, Ian (1993), Bilingualism and Network Externalities, Canadian Journal of Economics 26 (2), 337-345. URL: http://www.jstor.org/stable/135911

Drubin, David G.; Kellogg, Douglas R. (2012) English as the Universal Language of Science: Opportunities and Challenges, Molecular Biology of the Cell 23 (8), 1399. DOI: 10.1091/mbc.E12-02-0108

Dustmann, Christian; Fabbri, Francesca (2003), Language Proficiency and Labour Market Performance of Immigrants in the UK, Economic Journal 113 (489), 695-717. DOI: 10.1111/1468-0297.t01-1-0015

Ginsburgh, Victor; Ortuño-Ortín, Ignacio; Weber, Shlomo (2007), Learning Foreign Languages: Theoretical and Empirical Implications of the Selten and Pool model, Journal of Economic Behavior and Organization 64 (3-4), 337-347. DOI: 10.1016/j.jebo.2006.10.005

Graddol, David (1997), "The Future of English?: A Guide to Forecasting the Popularity of the English Language in the 21st Century", British Council, London.

Graddol, David (2006), "English Next: Why Global English May Mean the End of English as a Foreign Language", British Council, London.

Isphording, Ingo Eduard; Otten, Sebastian (2013), The Costs of Babylon: Linguistic Distance in Applied Economics, Review of International Economics 21 (2), 354-369. DOI:10.1111/roie.12041

Jackson, Matthew O. (2010), “Social and Economic Networks”, Princeton University Press, Princeton.

Laitin, David D. (1993), The Game Theory of Language Regimes, International Political Science Review 14 (3), 227-239. DOI: 10.1177/019251219301400302

Lewis, M. Paul; Simons, Gary F.; Fennig, Charles D. (eds.) (2013) “Ethnologue: Languages of the World, Seventeenth Edition”, SIL International, Dallas. URL: http://www.ethnologue.com.

Mélitz, Jacques (2007), The Impact of English Dominance on Literature and Welfare, Journal of Economic Behavior and Organization 64 (2), 193-215. DOI: 10.1016/j.jebo.2006.10.003

Miniwatts Marketing Group (2013) “Internet World Stats”, Bogota. URL: http://www.internetworldstats.com

Phillipson, Robert (1992), “Linguistic Imperialism”, Oxford University Press, Oxford.

Ray, Debraj (2008), “A Game-Theoretic Perspective on Coalition Formation (The Lipsey Lectures)”, Oxford University Press, Oxford.

Schelling, Thomas C. (1978), “Micromotives and Macrobehavior”, Norton, New York.

Selten, Reinhard; Pool, Jonathan (1991), The Distribution of Foreign Language Skills as a Game Equilibrium, in Reinhard Selten (ed.), “Game Equilibrium Models IV: Social and Political Interaction”, Springer-Verlag, Berlin.

Tainer, Evelina (1988), English Language Proficiency and the Determination of Earnings among Foreign-Born Men, Journal of Human Resources 23 (1), 108-122. URL: http://www.jstor.org/stable/145847

Wickström, Bengt-Arne (2005), Can Bilingualism be Dynamically Stable?: A Simple Model of Language Choice, Rationality and Society 17 (1), 81-115. DOI: 10.1177/1043463105051776

Rank	Language	Primary Country	Speakers
1	Chinese	China	1,197
2	Spanish	Spain	406
3	English	United Kingdom	335
4	Hindi	India	260
5	Arabic	Saudi Arabia	223
6	Portuguese	Portugal	202
7	Bengali	Bangladesh	193
8	Russian	Russian Federation	162
9	Japanese	Japan	122
10	Javanese	Indonesia	84.3
11	German, Standard	Germany	83.8
12	Lahnda	Pakistan	82.7

Table 1: The top 12 languages by the number of native speakers

Source: Lewis et al. (eds, 2013)

r	x_2	x_3	x_4	$U_{32}[x_3,1,x_4]$	$U_{31}[x_3,x_2,0]$	U_{32}-U_{31}	H_2	H_3	H_4
0.2	0.8	0.7	0.6	0.2326	0.2373	-0.0047			
0.2	0.9	0.5	0.3	0.2386	0.2403	-0.0017			
0.2	0.9	0.5	0.4	0.2448	0.2403	0.0045	E		
0.2	0.9	0.8	0.3	0.2429	0.2430	-0.0001			
0.2	0.9	0.8	0.4	0.2497	0.2430	0.0067	E		
0.2	0.9	0.8	0.7	0.2710	0.2430	0.0280	E	E	
0.3	0.8	0.5	0.3	0.3752	0.3859	-0.0107			
0.3	0.8	0.5	0.4	0.3925	0.3859	0.0066	E		
0.3	0.9	0.7	0.1	0.3939	0.4039	-0.0100			
0.3	0.9	0.7	0.2	0.4139	0.4039	0.0100	E		
0.3	0.9	0.7	0.3	0.4349	0.4039	0.0310	E		
0.3	0.9	0.7	0.4	0.4568	0.4039	0.0529	E		
0.3	0.9	0.7	0.5	0.4796	0.4039	0.0757	E	E	
0.3	0.9	0.7	0.6	0.5036	0.4039	0.0996	E	E	
0.3	0.9	0.8	0.5	0.4890	0.4074	0.0816	E	E	
0.3	0.9	0.8	0.6	0.5146	0.4074	0.1072	E	E	E
0.4	0.7	0.6	0.3	0.5460	0.5489	-0.0029			
0.4	0.7	0.6	0.4	0.5899	0.5489	0.0410	E		
0.4	0.7	0.6	0.5	0.6377	0.5489	0.0889	E	E	E

E: Existence

Table 2: Numerical examples of how changes in parameters such as r, x_2, x_3 and x_4 affect the existence of equilibria

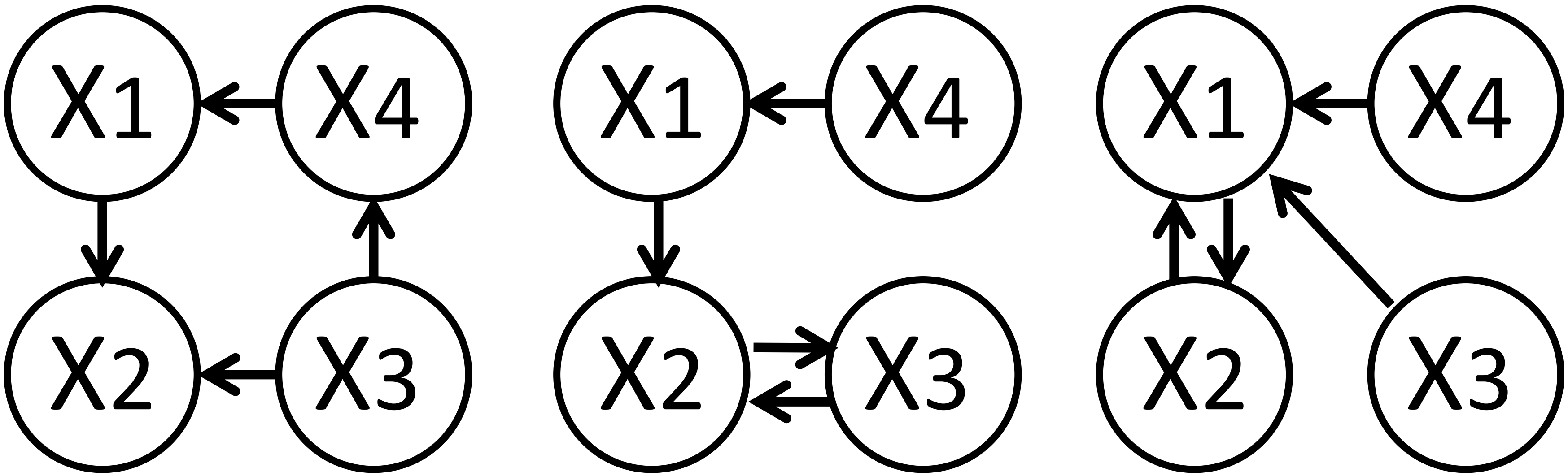

Figure 1. The left side is an example of Situation 2, the center is that of Situation 2, and the right side is that of Situation 3.

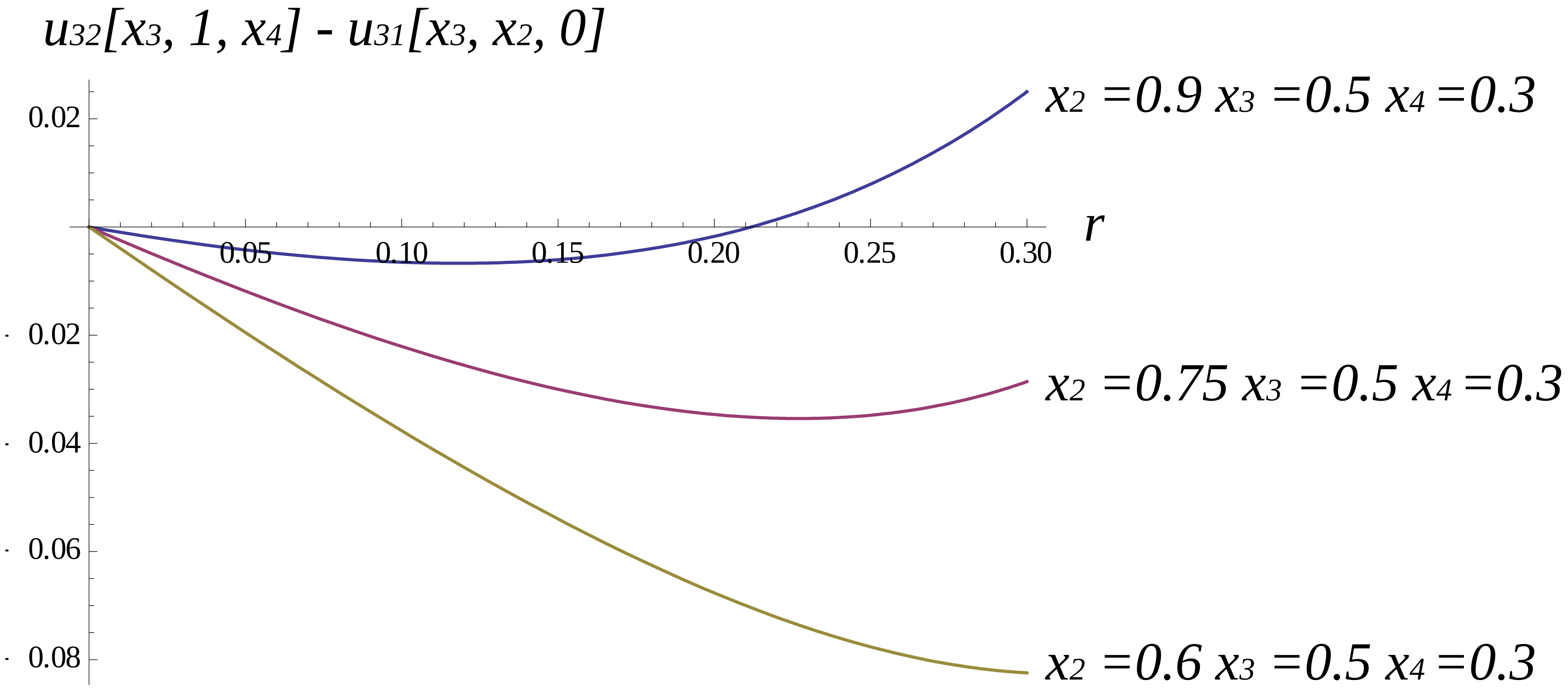

Figure 2. Typical graphs of $u_{32}[x_3, 1, x_4] - u_{31}[x_3, x_2, 0]$ with respect to r

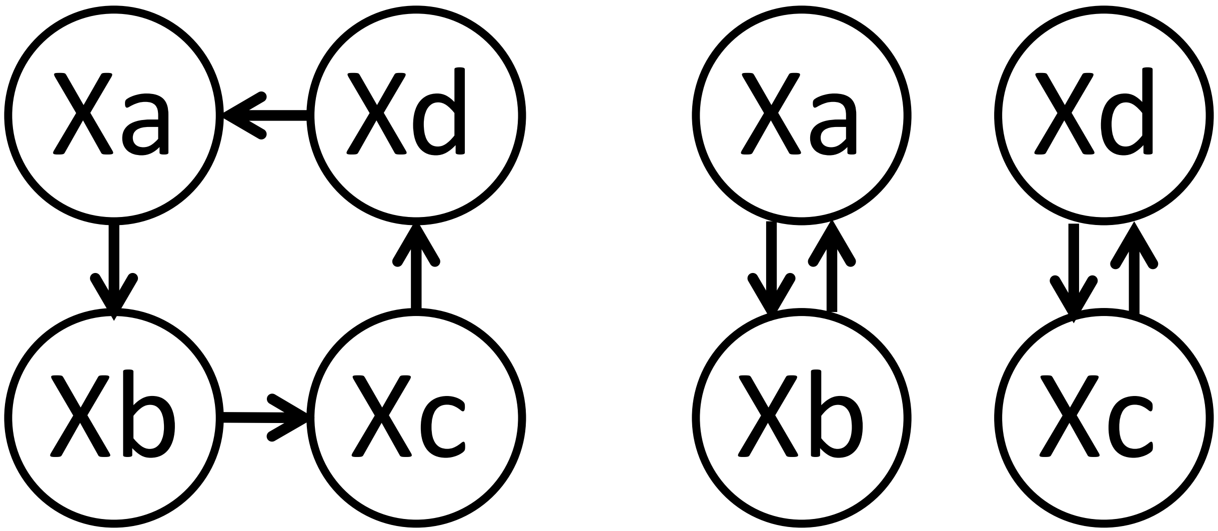

Figure 3. The left side illustrates Situation 1-i and the right side illustrates Situation 1-ii.

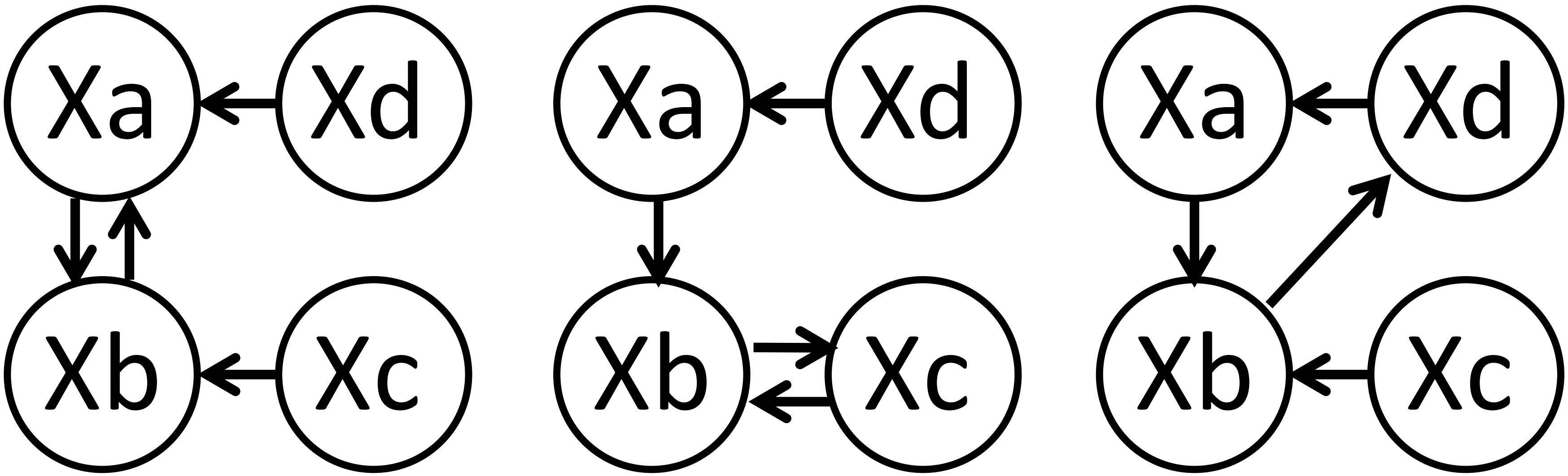

Figure 4. The left side illustrates Situation 2-i, the center illustrates 2-ii, and the right side illustrates 2-iii.