Essays on Social Interactions

by

Nicola Campigotto

Submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
in Economics

Doctoral advisor:
Nicola Dimitri

Head of programme:
Ugo Pagano

Dissertation committee:
Samuel Bowles
Andrés Perea
Paolo Pin
Robert Sugden

Cycle XXXI · Academic Year 2018–2019
Essays on Social Interactions

Nicola Campigotto
## Contents

3.2 Data ................................................................. 66  
  3.2.1 Descriptive Statistics .................................... 69  
3.3 Determinants of link formation .............................. 70  
  3.3.1 Gender, ethnicity, and religion .......................... 70  
  3.3.2 Cultural traits and school achievements ............... 73  
  3.3.3 A consistency check ...................................... 75  
3.4 Concluding remarks ........................................... 78  

A Appendices ......................................................... 81  
  A.1 Appendix to Chapter 2 ....................................... 81  
    A.1.1 MATLAB code ........................................... 81  
  A.2 Appendix to Chapter 3 ....................................... 85  
    A.2.1 Questionnaire items .................................... 85  
    A.2.2 Variables description .................................. 87  

Bibliography ......................................................... 91
On February 11, 1854, Commodore Matthew Perry entered Edo Bay with eight American warships and gave the order to anchor off the port of Yokohama. Perry had been sent to Japan to carry a message: open the country to foreign trade or prepare for war. Less than one year before, he had led four ships into harbour at Edo and left a week later with the promise to return for an answer: ‘[t]he undersigned, as an evidence of his friendly intentions, has brought but four of the smaller [ships], designing, should it become necessary, to return […] in the ensuing spring with a much larger force’ (Perry 1856, p. 259).

The threat posed by the American ‘black ships’ led to the signing of a series of treaties by which Japan opened to commerce, lost tariff autonomy over several ports, and exempted US citizens from the jurisdiction of domestic laws. Shortly afterwards, similar agreements were made with Britain, Russia, and France. The subservience of the Shogun’s government to the demands of foreign countries exacerbated discontent among the samurai class, leading to a revolutionary crisis which culminated in the ousting of the Tokugawa regime in 1868. This was followed by a period of rapid and unprecedented change. An imperial government was established in Edo, renamed Tokyo, and many shogunal institutions were replaced by new, western-style ones: a constitution was adopted and the hierarchical class structure was abolished; a conscription law was passed; western equipment, techniques, and skilled labour were imported from abroad; schools and universities were founded throughout the country.

The fall of the shogunate and the subsequent process of social transformation were driven by a radical change in Japan’s self-perceived status. Until Perry’s arrival, most Japanese citizens had been unaware of their country’s backwardness relative to the West. Foreign pressures sparked fear and revealed how far Japan lagged behind the industrialised nations. An anonymous kyōka poem of the time goes:

Taihei no
nemuri wo samasu

Awoken from sleep
of a peaceful quiet world
The verses have a double entendre, the four cups of jōkisen tea referring to Perry’s four ships. An alternative interpretation is ‘awakened from sleep / of a peaceful quiet world / by steamships; / a mere four ships / and one loses sleep at night.’

Japan’s ignorance of the outside world stemmed from two sources: the country’s geographic isolation as an island and the sakoku (seclusion) policy pursued by the Tokugawa shogunate since the 1630s. Political and commercial relations had been strictly limited: the sole connection with other countries was the Dutch trading post at Dejima, a small artificial island off the coast of Nagasaki. Aside from a small number of Chinese and Dutch merchants and diplomats, foreigners had been barred from entering the country, and Japanese citizens were not allowed to leave it. The lack of interactions had nearly halted the flow of information.

The implementation of reforms by the new Meiji leaders was not without issues. Many problems arose from the need to legitimise the reforms and shape the national culture to ease the adoption of foreign technologies and institutions. To achieve these goals, the government pursued a twofold strategy. First, it consolidated the new order through its ritualisation. Mandatory education was used to bring Confucianism to the population and elevate it to the status of a state religion, thus strengthening the loyalty of the people to the Tennō (celestial emperor). Second, it sought to change the Japanese ideas of identity and tradition. This was done by reinterpreting certain elements of the national culture in ways that suited the needs of the time, as suggested by the slogan wakon yōsai (Japanese spirit, western techniques). For example, until the late nineteenth century, merchants had lain at the bottom of the social class hierarchy and money lending had been viewed poorly. In order to free merchants from the social stigma associated with their trade, the Meiji government sought to promote

the ‘myth’ of the ‘unique’ nationalist entrepreneur. Recognising the need to rehabilitate the validity of economic activity as part of the national project, [...] the state [...] employed a powerful rhetoric to stress the extent to which entrepreneurship, investment and profit-making were just as valid expressions of nationalism and patriotism as political or military service (Hunter 2006, p. 61).

The idea of ‘enterprise for the sake of the country’ was thus imbued with

1 The ideograms that comprise the word jōkisen, a variety of green tea, are homophonic to those used to write ‘steamship,’ and shihai can also be understood as ‘four ships.’
a symbolic meaning that transcended simple, selfish profit-seeking. A leading Italian orientalist, Franco Mazzei, commented that eventually the Japanese entrepreneur came to see themself as ‘a loyal subject of the Tennō at least as much as his Calvinist counterpart considered himself a good servant of God’ (Mazzei 1982, p. 30, author’s translation).

**Information, perception, and the structure of social interactions**

In the introduction to his *Principles*, Alfred Marshall (1890, p. 1) claimed economics to be ‘on the one side a study of wealth; and on the other, and more important side, a part of the study of man.’ Nevertheless, the theoretical apparatus which Marshall helped develop was largely inadequate for this purpose. Traditionally, economic agents have been modelled as perfectly rational optimisers and abstracted from the social context in which they act. The focus of analysis has long been on allocative decisions by self-regarding individuals in situations of complete certainty about the world, or where uncertainty is represented by a probability distribution over a set of well-defined events. This has precluded many important factors from entering into discussions about human behaviour.

A related point can be made with reference to positive theories of decision making, where the behaviourist notion of revealed preference continues to be widely employed. Gul and Pesendorfer (2008) seek to explain the rationale for a ‘mindless economics,’ stressing that economic theories should only concern themselves with people’s observable choices. Perception, motivations, and other psychological variables are dismissed as futile because evidence about choice behaviour cannot discriminate among them. Nevertheless, mindless approaches do not come without limitations, as Amartya Sen’s (1993) famous example of dining etiquette demonstrates. Given the choice between eating the last remaining apple in a fruit basket and having nothing, a polite dinner guest — who, if alone, would definitely choose to eat the apple — decides to have nothing. This choice cannot be rationalised by a preference ordering over the available options, but it by no means indicates a lack of rationality. The guest’s behaviour may be based upon norms of good manners, aversion to social disapproval, or other factors that are relevant to them at the time the decision is made (Dietrich and List 2016).

It is only relatively recently that economists have started to grapple with the inability of the aforementioned paradigms to explain behaviour. Increasingly, there have been explorations of concepts that had been commonly thought to belong to the domain of other disciplines. Pioneering works by Herbert Simon, and more recently by Gerd Gigerenzer and co-
authors, have rejected the idea of maximising behaviour in favour of ‘satisficing’ and ‘fast and frugal’ heuristics (Simon 1956; Gigerenzer and Selten 2001; Gigerenzer, Hertwig, et al. 2011). Behavioural economics has documented that people often have inconsistent preferences and are influenced by a variety of cognitive biases (Kahneman and Tversky 2000; Thaler and Sunstein 2008). Research in game theory has recognised the importance of epistemic and psychological conditions in determining the outcome of strategic interactions (Geanakoplos et al. 1989; Aumann and Brandenburger 1995). Evolutionary models have been employed to investigate the long-term social outcomes of myopic decisions and the emergence and interplay of people’s beliefs and preferences (Fudenberg and Levine 1998; Young 1998; Bowles and Gintis 2011).

Among these lines of research are those concerned with how perception and information processing, culture, and the social structure relate to one another and shape social interactions. The main features and implications of each of these factors are worthy of specific mention.

I. Perception and information. The Meiji revolution exemplifies what Douglass North (2005a, p. 4) described as a ‘story of perceived reality → beliefs → institutions → policies → altered perceived reality and on and on,’ whose key element is ‘the way beliefs are altered by feedback from changed perceived reality.’ Japan’s industrialisation and westernisation originated from a sudden awareness of the country’s irrelevance on the world stage, and these goals were achieved by, among other things, changing perceptions and ideological constructs about the role of business activities.

There is now a large body of evidence that beliefs and preferences are context-dependent (Tversky and Simonson 1993; Fehr and Hoff 2011). Changes in the social environment can activate different mental frames, inducing shifts in individual preferences, and experience (i.e., the acquisition of new information) can influence behaviour via changes in an agent’s interpretation of a situation. For Bicchieri (2006, pp. 57, 59), ‘it is precisely the mapping from context to interpretation, and thus to beliefs and expectations, that elicits a preference for conformity [to a norm of behaviour].’ To say that a social norm applies in a given situation is to say that individuals ‘infer from some situational cues what the appropriate behavior is, what they should expect others to do, and what they are expected to do themselves, and act upon those cues.’

Differences in how information spreads and is perceived can have important consequences on the outcome of interactions, as the case of the Folk Theorem — one of the mainstays of the theory of repeated games — shows. The theorem states that, if individuals are sufficiently far-sighted
and there is perfect public information concerning the strategy chosen by each player, then social cooperation can be achieved and sustained over time. However, things get complicated as soon as private information and perceptual errors enter the picture. As errors are not perfectly correlated across individuals, a cooperator may be mistakenly recognised by some players as having defected. This undermines the usefulness of trigger strategies, because an erroneously perceived defection signal leads to a break-down of cooperation. Sekiguchi (1997) and Bhaskar and Obara (2002) prove that there exists an equilibrium where all players cooperate with less than unit probability in the first period, and then react to defection signals (both correct and incorrect) by defecting themselves. Results by Gintis (2009) and Bowles and Gintis (2011), however, show that for plausible values of the game’s parameters, this equilibrium is highly inefficient.\footnote{Repeated games with private information can also admit mixed strategy equilibria in which players randomise over their actions in each period (e.g., Mailath and Morris 2006). Bowles and Gintis (2011) dismiss these equilibria as ‘evolutionary irrelevant,’ objecting that there is no reason to believe that agents would ever coordinate on a particular randomisation.}

II. Social structure. Behaviour and the flow of information crucially depend on the underlying structure of interactions. Recent works in this area have related the topology of social relations to such things as the spread of fake news (Del Vicario et al. 2016), the conformity to deviant group norms (Patacchini and Zenou 2012), and the success of microfinance programmes (Banerjee et al. 2013). The structure of social and economic networks also plays an important role in the emergence of power and inequality (e.g., Cook and Emerson 1978; Bowles 2009; DiMaggio and Garip 2012), as well as in determining the willingness of a society for political collective action (Barberà and Jackson 2018). Returning to the case of Japan, the efforts by the Tokugawa to prevent natives and Westerners from interacting with one another were mainly aimed at preempting revolts by impeding the teaching and practice of Christianity. The sakoku was a means of political and social control, introduced to enforce the ban on a religion which the shoguns saw as a potential threat to their rule.

Network formation processes involve a combination of random events and strategic considerations (see Jackson et al. 2016 and Pin and Rogers 2016 for overviews). People form relationships based on costs and benefits, both internal and external, and tend to associate disproportionately with others who have similar characteristics (Lazarsfeld and Merton 1954). Homophilic preferences have been documented with respect to a variety
of individual traits (McPherson et al. 2001), and most often with respect to ethnicity (Currarini, Jackson, et al. 2009, 2010). Importantly, like-with-like relationships can promote the emerge of cooperation in social dilemmas: if individuals interact assortatively, the risk of cooperating may be compensated for by a higher probability of interacting with another cooperator (Bergstrom 2003; Cooney et al. 2016). On the other hand, however, the tendency of similar people to clump together may hamper social learning; segregation can cause ideas and information to spread rapidly within groups but to move slowly, if not at all, across them (Golub and Jackson 2012).

III. Culture and identity. Cultural closeness, commonality of social and ethnic backgrounds, and group identity all influence preferences and beliefs, resulting in a higher propensity to engage in prosocial behaviours. A number of scholars (including Tuomela and K. Miller 1988; Searle 1990; Bratman 1992; Bacharach 2006; Gold and Sugden 2007; Akerlof 2016) have sought to extend decision making models in ways that account for cases where agents reason in terms of ‘we’ rather than ‘I.’ Team reasoning, or we-thinking, applies to those situations where the sharing of a common identity causes individuals to internalise the group’s goals as their own. This can allow members to avoid coordination failures without a need for communication.

Experimental evidence confirms that individuals are more altruistic towards people who belong to the same group as them (Bernhard et al. 2006). This is true even in those cases where group identity is induced artificially in the laboratory (Chen and Li 2009; Guala et al. 2013). Similarly, ethnic diversity is generally found to reduce cooperation (Glaeser et al. 2000; Fershtman and Gneezy 2001; Habyarimana et al. 2007) — a finding that is consistent with the remark by Robert Putnam (2007) that greater ethnic heterogeneity tends to be associated with lower social trust. Bowles and Gintis (2004) stress that when centralised decision making institutions are lacking, impersonal exchanges can be sustained by cultural in-group versus out-group distinctions and exclusionary practices towards outsiders.

In a related vein, Clist and Verschoor (2017) have conducted a series of experiments on a bilingual population in Uganda to test whether the language in which a public good game is played affects individual contributions. Contributions were found to be significantly higher when the game was played in the national language, Luganda, rather than in Lugisu. Further analysis revealed that this result was entirely driven by those participants who were most closely associated with — but did not necessarily belong to — the Gisu culture, which highly values self-reliance.
and individual autonomy. According to the authors, this suggests that language does ‘affect cooperation, but only for individuals who both have different latent norms and for whom language activates these norms’ (p. 47).

**Scope and organisation of the dissertation**

This dissertation consists of three essays on the determinants and consequences of social interactions. The chapters are self-contained and can be read independently of one another.

Chapter 1 deals with the perceptual and reasoning processes that underpin regularities in behaviour, and draws upon David Lewis’ (1969) *Convention* and the literature it inspired. An assumption common to several Lewisian models is that the sustaining of behavioural regularities requires some symmetry in the agents’ reasonings. We stress that this need not be the case, and we suggest weaker sufficient conditions for the entrenchment of conventions. A distinction is made between the world as it is and the ways agents frame it when facing a decision problem; different frames can stem from differences in culture, experience, and personality, as well as from other context-specific factors. We show that consistency between reasoning and experience does not preclude individuals from understanding the same situation differently, and that the beliefs that sustain a convention may well be false. Regularities in behaviour may emerge as the outcome of repeated self-confirming interactions, in which no action that contradicts the agents’ beliefs is ever observed.

Chapter 2 discusses the long-term consequences of interactions based on pairwise comparisons. We study the evolution of play in a population of agents who repeatedly engage in a Stag Hunt game, which is seen as a prototypical representation of the social contract. A key feature of the model is the use of pairwise adaptive rules of the imitate-if-better kind; a revising agent, $i$, compares themself with another individual, $j$, and copies $j$’s action if and only if it did better than their own action. This is a reasonable assumption for many situations where agents rely on intuitive, cognitively undemanding heuristics, or where it is difficult to obtain information about the behaviour of other people. Results for two different imitative rules are contrasted with each other and with those from a third, non-imitative protocol, namely, best response to the current state of play. Depending on payoffs and on how players meet, best response can select either the Pareto superior all-cooperate equilibrium or the inefficient all-defect equilibrium as the long term convention. Pairwise imitation, on the contrary, always selects the Pareto inferior equilibrium;
when behaviour is driven by myopic pairwise comparisons, it becomes very difficult for a society to settle upon a social contract. This finding is robust to alternative specifications of the model, and most notably to the introduction of assortative matching.

Chapter 3, co-authored with Chiara Rapallini and Aldo Rustichini, turns to the determinants of group formation. The chapter makes a case for studying homophily from a multidimensional perspective and shows that, in explaining peer selection, characteristics such as religious creed and normative beliefs cannot be reduced to ethnicity. We propose a model of network formation that accounts for both multiple dimensions of homophily and group composition, and we estimate it using a two-wave panel survey of secondary school students in four European countries. To try and address the endogeneity concerns that arise when asking which individual characteristics are most relevant for peer assortment, we identify instruments by exploiting the structure of classroom networks and using the characteristics of an individual’s indirect friends. We find evidence of deep segregation and preferences for similarity with respect to gender, nationality, religion, socio-economic status, and academic achievements. Additionally, we provide evidence that normative beliefs and other culture-related traits all play distinct and non-negligible roles in determining friendship relations.
1. Frames, Reasoning, and the Emergence of Conventions

1.1 Introduction

Cognitive processes have long been recognised as an important component of social interactions, and their study has been providing valuable insights into the nature and causes of human behaviour. Yet social sciences continue to reflect an ambivalent stance towards perceptual and cognitive assumptions. Although most people agree that individuals inaccurately perceive and interpret the environment within which they act, decision and game theoretic models frequently posit that agents represent situations to themselves in a systematically correct way. This causes subjective appraisals to coincide with reality, and makes uncertainty boil down to a probability distribution over states of the world.

Early attempts by economists to understand how knowledge is acquired and transmitted trace back to the works of Friedrich Hayek (1937, 1952) and Herbert Simon (1955). For Hayek (1937, pp. 36, 46), the information upon which an individual bases his decisions are ‘things as they are known to (or believed by) him to exist, and not in any sense objective facts.’ Therefore, the study of social processes ‘must necessarily run in terms of assertions about causal connections, about how experience creates knowledge.’ The differences with more recent epistemic frameworks (see, e.g., Brandenburger 2014 and the textbooks by Gintis 2009 and Perea 2012) are not merely a matter of jargon. The latter provide powerful tools for representing beliefs, conditional to the occurrence of some events, in situations that are unambiguously defined, but do not venture to explain how agents form their mental models (Sugden 2011a).

Possibly in response to this, Lewisian models — often advanced to elucidate the functioning of institutions and social norms — have been receiving increasing attention (Aoki 2011, 2017; Sillari 2013; Hédoin 2014, 2017; Hindriks and Guala 2015). All build on David Lewis’ (1969) Convention, which seeks to capture the role of reasoning in the emergence and sustaining of regularities in behaviour. For Lewis, a pattern in the
behaviour of a group of individuals represents a convention if and only if it is customary, expected, and satisfies a mutual optimality condition. Moreover, these properties must be common knowledge (in a sense that shall be clarified below) in the group to which the convention applies.

Notwithstanding its relevance, the line of research initiated by Lewis contains an element of ambiguity as to whether — and if so, how far — stable conventions require reasoning to be symmetric across individuals. Some scholars have maintained that, for a convention to be in place, agents must have common background information and inductive standards; others have suggested that this might not be the case, although without discussing this point much further. Intuitively, to say that, in a certain situation, a group of individuals shares the same inductive standards is to say that: (i) agents have a common view or understanding of the relevant features of that situation; (ii) taking this view as a premise, they all reach the same conclusion — which, for instance, may consist of a proposition of the kind ‘I ought to do $x$’; and (iii) they all have reason to believe that others have made the same inference as they did.

Depending on context, the assumptions of shared and symmetric reasoning may or may not be appropriate. To better see why, consider the following examples.

**Example 1 (Meeting at the bar).** Alice and Bob meet accidentally in a bar. For both of them, making the acquaintance of someone unrelated to work is a breath of fresh air. They enjoy their time together, and, without any prior agreement, they start meeting at the bar every week. However, their motives and understandings of the situation differ considerably. Alice sees Bob as a new friend to share stories with, and nothing in Bob’s behaviour gives her reason to believe that he may view their relationship any differently. Bob, however, is immediately infatuated with her. Although aware that Alice sees him just as a friend, he hangs on and continues to meet her, hoping that someday he will succeed in making her reciprocate his feelings. After all, Bob believes, good things come to those who wait...

**Example 2 (Delinquency and violence in juvenile gangs).** When privately interviewed, gang members often reveal a deep uneasiness with their behaviour (Matza 1964). Yet, since they do not express their misgivings publicly, they give the impression of being genuinely committed. Each member believes that other members confer a higher status on those who engage in violent acts, and some members actually regard delinquency and violence as the proper way to express their identity. No one, on the other hand, seems to consider the possibility that many of their peers
behave violently not for the sake of it, but because they believe that other members believe that gangs should behave that way, they believe that others believe that they themselves believe that gangs should behave that way, and so on. Conformity signals endorsement and confirms the false belief that the group as a whole supports the gang’s conduct (Bicchieri and Fukui 1999).

The examples illustrate two related points. First, some components of an individual’s view of a situation may well not be shared by other people. Bob has romantic feelings for Alice and thinks of their meetings as occasions to win her love, whereas Alice only likes Bob as a friend; similarly, some gang members take pride in their behaviour, while others conform out of a fear of being rejected or punished by the group. Second, and importantly, the agents’ beliefs about others’ motives and reasoning may be imprecise or false. Alice is wrongly convinced that Bob thinks the same about her as she does him, and each gang member wrongly believes that all other members endorse delinquency. In both examples, false beliefs about other people’s views co-exist with adherence to a regularity in behaviour. Clearly, in each case there are some beliefs that are correct and common to all agents (for instance, each gang member correctly believes that each other believes that all other members approve the group’s behaviour.) Nevertheless, to limit discussion to correct beliefs only, and to think of the above conventions as being underpinned by symmetric reasoning processes, would be misleading.

This chapter seeks weaker, non-symmetric conditions for the emergence of Lewisian conventions. It does so by making a clear distinction between the world as it is, or as seen from the external viewpoint of the social scientist, and the ways individuals perceive and interpret it. The exercise draws from two separate bodies of literature, one related to the concept of self-confirming behaviour (see Section 1.4 and the references therein) and the other concerned with the framing of situations (Bacharach 1993, 2003, 2006; Turner 2001). ‘Frames’ are bundles of concepts that are employed in an individual’s subjective representation of objects and events. Reasoning processes operate upon these representations, and can vary considerably depending on which concepts are used to frame a situation. Similarity in the agents’ frames, moreover, can be thought of as a stylised fact about their culture, as human mental processes operate over culturally developed assemblies of experience (Bacharach 1993; Gintis 2017).

The chapter aims to contribute to the literature in two ways. First, it stresses that for a convention to emerge and persist over time, the agents’ reasonings need not be symmetric. More specifically, it shows that
the development by individuals of consistent higher-order beliefs does not preclude their views of a situation from being different in relevant respects. Second, it allows for cases where conventions are based upon false but self-confirming beliefs. This requires for the conventional course of behaviour to never prove the agents wrong; in the words of Hayek (1937, pp. 39-40), ‘the subjective data, given to the different persons, and the individual plans, which necessarily follow from them, are in agreement,’ and ‘there is consequently a conceivable set of external events which will allow all people to carry out their plans and not cause any disappointments.’

The chapter unfolds as follows. Section 1.2 introduces Lewis’ theory and discusses the assumption of symmetric reasoning. Section 1.3 presents our frame-dependent model. Section 1.4 gives an example drawn from history. Section 1.5 reviews some related literature. Section 1.6 concludes and suggests possible directions for future research.

1.2  Conventions and reason to believe

To appreciate the depth of Lewis’ thought, two points need to be stressed. First, his analysis was originally meant to apply to coordination problems only. However, in almost no case does Lewis rely on properties of an action other than its being optimal under an agent’s beliefs. Therefore, the same framework can be applied to a much broader set of interactions.

Second, although Lewis is commonly credited as the first to introduce the concept of common knowledge, his work actually concerns beliefs rather than knowledge (Cubitt and Sugden 2003; Sillari 2005, 2008). More precisely, Lewis’ interest lay in assessing the modes of reasoning that underpin human beliefs and behaviour. The framework is that of epistemic rationality: by saying that agent $i$ has reason to believe proposition $x$, we mean that $i$ is justified in believing that $x$ is true by virtue of the reasoning they endorse and the evidence they possess. Eventually, Lewis himself admitted that the term ‘common knowledge’ used in his definition of convention ‘was unfortunate, since there is no assurance that it will be knowledge, or even that it will be true’ (Lewis 1978, p. 44, cit. in Sillari 2008). Therefore, to avoid confusion, hereafter we shall use the term ‘common reason to believe’ (CRTB).

1 That is, the idea that all agents know that each one knows that all other agents know . . . and so on ad infinitum (cf. Aumann 1976). As a matter of interest, it is worth mentioning that an even earlier definition of common knowledge was given by American sociologist Morris Friedell (1967, 1969). For a discussion of the roles of common knowledge and common belief in the development of game theory, see Perea (2014).
Lewis’ (1969) definition of CRtB goes as follows. Members of a population $P$ have common reason to believe a proposition $x$ if and only if some state of affairs $A$ holds such that:

(i) everyone in $P$ has reason to believe that $A$ holds;
(ii) $A$ indicates to everyone in $P$ that everyone in $P$ has reason to believe that $A$ holds;
(iii) $A$ indicates to everyone in $P$ that $x$.

As suggested by Cubitt and Sugden (2003), states of affairs can be thought of as corresponding to Savage’s states of the world — ‘description[s] of the world, leaving no relevant aspect undefined’ (Savage 1954, p. 9). When conditions (i)–(iii) are met, $A$ is said to be a basis for common reason to believe. This, together with what Lewis refers to as ‘suitable ancillary premises’ about the agents’ rationality and their beliefs about each other’s reasoning, allows the development of higher-order reason to believe.

The mutual consistency of individual expectations is of crucial importance for the sustaining of conventions. For expectations to be consistent, each individual must have reason to believe that everyone else will adhere to a regularity in behaviour. If agents have a conditional preference for conformity, then each one is motivated to behave in ways which, in turn, confirm other people’s expectations. The challenge, therefore, is to identify the conditions that allow the development of mutually consistent expectations about future behaviour. An obvious way of producing first- and higher-order reason to believe, Lewis observes, consists in explicitly agreeing to stick to an action or course of action. Another important factor influencing beliefs and behaviour is salience, that is, the perception that some elements of one’s view of a situation stand out from the others. Finally, in the case of repeated interactions, precedence may emerge as a special kind of salience, resulting in a tendency to take those actions which have proven successful in the past.\footnote{Sugden (2011b) uses the concept of salience to explain how and why a behavioural regularity may come about. The role of perceptions of similarity with past instances of a decision problem has been discussed, among others, by Gilboa and Schmeidler (1995) and Alberti et al. (2012).}

1.2.1 On symmetric reasoning

What individuals have reason to believe depends on both their modes of reasoning and on the information upon which they assess situations. In particular, Lewis’ common reason to believe can only obtain when all
members of a group have reason to believe that they all share the same inductive standards and background information. CRtB, therefore, is not to be understood from a purely individualist perspective; rather, it requires people to endorse lines of reasoning of the kind ‘I have reason to believe \( x \). Thus, so may others.’

Lewisian models, however, often go beyond this and assume inductive standards to be actually shared among agents. For example, Hédoin (2014) explicitly postulates that individuals commonly know that they share the same modes of reasoning with respect to a given set of situations. Lewis’ original framework, too, is generally interpreted as requiring common inductive standards. For Sillari (2005, p. 391), ‘in some cases inductive standards are shared by groups of agents, as, for example, in those cases in which common reason to believe comes about,’ while Bicchieri (2006, p. 37) observes that ‘Lewis’ argument is crucially dependent on assuming shared inductive standards.’ Similarly, Vanderschraaf (1998, p. 362) claims that ‘a crucial assumption in Lewis’ analysis of common knowledge is that agents know they share the same rationality, inductive standards, and background information with respect to a state of affairs.’

Lewis’ definition of CRtB does not necessarily require reasoning to be identical across agents. For instance, it allows for individuals to reach the same conclusion from the same premise through somewhat different paths. Also, it does not require all inductive rules to be shared; the mutual ascription of common inductive standards only applies to a particular set of propositions which are relevant to the situation being considered. However, Lewis’ definition does require individuals to reason symmetrically: if a proposition (‘\( A \) holds’) indicates another proposition (\( x \)) to a member of \( P \), then it does the same to all other members of the population (Hédoin 2014, 2017; Vanderschraaf and Sillari 2014). This condition is often seen as controversial: Sillari (2008, p. 31) acknowledges it as being ‘far from innocuous,’ and for Aoki (2011, p. 27) ‘it is admittedly strong and may often fail to apply.’

This symmetry in the agents’ reasoning arises in Lewis’ framework as a consequence of the fact that no attempt is made to distinguish between the world as it is and the world as individuals see it. Lewis’ and Lewisian models implicitly assume that states of affairs are public and perfectly self-revealing (Cubitt and Sugden 2003). If \( A \) obtains, then each agent has reason to believe that \( A \) holds; this, in turn, leads everyone to infer that \( x \). A graphical representation of the perception-inference process underlying Lewis’ CRtB is shown in the left-hand panel of Figure 1.1

---

3 Credit goes to Herbert Gintis for making this point.
When is an event public and self-revealing? As an example, consider the coordination problem discussed by Lewis (1969, p. 52) to introduce conventions by agreement:

[s]uppose the following state of affairs — call it $A$ — holds: you and I have met, we have been talking together, you must leave before our business is done; so you say you will return to the same place tomorrow. Imagine the case. Clearly, I will expect you to return. You will expect me to expect you to return. I will expect you to expect me to expect you to return.

This state of affairs can reasonably be seen as being capable of generating higher-order expectations through roughly common lines of thought: $(i)$ my view of $A$ is the same as yours in all relevant respects (you had to leave and you said you will return to the same place tomorrow), and both you and me have reason to believe that this is indeed the case; $(ii)$ $A$ indicates to both that both have reason to believe that $A$ holds; and $(iii)$ $A$ indicates to both that you will return. Therefore, $A$ represents a basis for CRTB in Lewis’ sense.

In other cases, and particularly when individuals are driven by different motives or interpret the same situation differently, the symmetry-in-reasoning condition may not be met. Yet agents may still form consistent higher-order beliefs. Cubitt and Sugden (2003) offer useful insights in this regard by introducing the concept of ‘distributed reason to believe.’ In discussing why American drivers drive on the right side of the road, they observe that no driver can have direct knowledge of other people’s experiences and driving habits, and that each agent’s decisions are made...
based on the experience of a particular sample of American drivers. This makes it natural to think of the drivers’ expectations as being licensed by different mental constructs. In a related vein, Cubitt and Sugden (2014) introduce a model where each agent endorses a private ‘reasoning scheme.’ Reasoning schemes may differ from one another, but are assumed to share a core of ‘common reason.’ The latter is made of those modes of reasoning that are shared across individuals and that each agent attributes to everyone in the population. Our model builds on these ideas by introducing elements of Bacharach’s Variable Frame Theory into Lewis’ framework.

1.3 The model

1.3.1 Some formal apparatus

Reason to believe and indication. The notation follows that of Cubitt and Sugden (2003) and adapts it to a possible worlds setting. Let $P = \{1, \ldots, n\}$ be a finite population of agents and let $\mathcal{A}$ be a non-empty set of possible worlds or states of affairs. A proposition is a subset $x \subseteq \mathcal{A}$, and the set of all possible propositions is given by the power set $\mathcal{P}(\mathcal{A})$. For each $i \in P$, let $\mathcal{R}_i$ be a modal operator from $\mathcal{P}(\mathcal{A})$ to itself. $\mathcal{R}_i(x)$ contains those worlds in which agent $i$ has reason to believe that $x$. Reason to believe can be nested to arbitrary depths; we write $\mathcal{R}_i[\mathcal{R}_j(x)]$ to represent the proposition that $i$ has reason to believe that $j$ has reason to believe that $x$, and we write $\mathcal{R}^P(x)$ to mean that $\mathcal{R}_i[\mathcal{R}_j[\mathcal{R}_k[\ldots[\mathcal{R}_n(x)]\ldots]]]$ is true for all finite sequences $i, j, k, \ldots, n \in P$.

Lewis’ indication relation is captured by a two-place modal operator, $\text{ind}_i$, where $x \text{ ind}_i y$ is the proposition that $x$ indicates to $i$ that $y$. The relation $A \in (x \text{ ind}_i y)$ means that, in world $A \in \mathcal{A}$, having reason to believe that $x$ provides agent $i$ with reason to believe that $y$. However, this may not hold for any other $j \neq i \in P$ and $A' \neq A \in \mathcal{A}$, meaning that indication has no universal validity. We take $\text{ind}_i$ to be reflexive: every proposition $x$ always indicates to $i$ that $x$, that is, $x \text{ ind}_i x$ holds tautologically.

The following axioms are assumed. Let $x$, $y$, and $z$ be propositions; for any pair of distinct individuals $i, j \in P$:

$$\mathcal{R}_i(x) \cap (x \text{ ind}_i y) \subseteq \mathcal{R}_i(y), \quad (1.1)$$

$$\quad (x \text{ ind}_i y) \cap (y \text{ ind}_i z) \subseteq (x \text{ ind}_i z), \quad (1.2)$$

$$\quad (x \text{ ind}_i \mathcal{R}_j(y)) \cap \mathcal{R}_i(y \text{ ind}_j z) \subseteq [x \text{ ind}_i \mathcal{R}_j(z)]. \quad (1.3)$$
Rule (1.1) states that if \( x \) indicates to agent \( i \) that \( y \), then reason to believe \( x \) implies reason to believe \( y \). Rule (1.2) requires indication to be transitive: if \( x \) indicates to agent \( i \) that \( y \), and if \( y \) indicates to them that \( z \), then \( x \) indicates to them that \( z \). Rule (1.3) says that if \( x \) indicates to \( i \) that \( j \) has reason to believe that \( y \), and if \( i \) has reason to believe that \( y \) indicates to \( j \) that \( z \), then \( x \) indicates to \( i \) that \( j \) has reason to believe that \( z \). For example, let \( x \), \( y \), and \( z \) be the propositions ‘the thunders are getting louder,’ ‘it might rain soon,’ and ‘it is a good idea to take an umbrella when leaving the house,’ respectively. According to (1.3), if the thunders indicate to \( i \) that \( j \) has reason to believe that it might rain soon, and if \( i \) has reason to believe that the possibility of showers indicates to \( j \) that it is a good idea to take an umbrella, then having reason to believe that the thunders are getting louder indicates to \( i \) that \( j \) has reason to believe that they should leave the house with an umbrella.

**Framing structures.** A ‘frame’ is composed of bundles of concepts that are employed in an individual’s contextual representation of a situation. As Turner (2001, pp. 13, 101) observes, ‘basic mental operations operate over [...] frames,’ and ‘it is natural to assume that decision making in any specific situation will depend on what frames are used by the decision maker as conceptual inputs.’ Moreover, interacting agents typically make decisions while taking into account other people’s views — or, better to say, what they think other people’s views might be. Based on these observations, let a framing structure be a tuple:

\[
\mathcal{F} := \langle P, \mathcal{A}, \{f_i, f_{i;j}\}_{j \neq i \in P} \rangle,
\]

where, for each \( i \in P \), the function \( f_i : \mathcal{A} \to \mathcal{P}(\mathcal{A}) \) maps world \( A \in \mathcal{A} \) to a proposition, \( f_i(A) \subset \mathcal{P}(\mathcal{A}) \), which represents \( i \)'s view of \( A \). Different frames may reflect differences in culture, cognitive abilities, or experience. In like fashion, for any pair of individuals \( i \) and \( j \), the function \( f_{i;j} : \mathcal{A} \to \mathcal{P}(\mathcal{A}) \) maps \( A \) to a proposition, \( f_{i;j}(A) \), which represents what \( i \) conjectures \( j \)'s frame of \( A \) to be.\(^4\)

\(^4\) The idea that individuals rationalise their actions by attributing thoughts and beliefs to other people is well supported in the literature. In his *Theory of Moral Sentiments*, Adam Smith argues that:

\[\text{[as] we have no immediate experience of what other men feel, we can form no idea of the manner in which they are affected, but by conceiving what we ourselves should feel in the like situation. [...] it is by the imagination only that we can form any conception of what are his sensations (A. Smith 1759, p. 34).}\]

Similarly, Lewis (1969, p. 27) notes that ‘we may acquire [...] expectations, or correct or corroborate whatever expectations we already have, by putting ourselves in the
Several points are worth making here. First, if the equality \( f_i(A) = f_i(A') \) holds for any distinct \( A, A' \in \mathcal{A} \), then the two worlds are frame-equivalent, meaning that \( i \) cannot discriminate \( A \) from \( A' \). Second, in order for all members of \( P \) to hold the same view of \( A \), it must be true that \( f_i(A) = f_j(A) \) for all \( i, j \in P \). Third, in the particular case where \( f_i(A) = f_{i;j}(A) = A \) for all \( i \) and \( j \), our frame-dependent model reduces to the symmetric Lewisian model. As we have seen, this may be a very strong condition to impose; in general, \( i \)'s and \( j \)'s frames of \( A \) will be similar in some (perhaps many) respects and different in some others. Likewise, \( i \)'s frame and \( i \)'s conjecture of \( j \)'s frame will typically be correlated with one another, but may also differ in some important respects (for instance, in our earlier example, some gang members view their group’s behaviour as morally wrong but conjecture that other members take pride in violence).

Fourth, ‘framing is logically prior to believing’ (Bacharach 2003, p. 66): if \( f_i(A) \) and \( f_{i;j}(A) \) do not contain proposition \( y \), then agent \( i \) cannot have beliefs about \( y \) when representing \( A \) to themself. Moreover, \( i \) remains unaware of this fact because \( y \) has not come to their mind at all.

Finally, the components of an individual’s frame may be seen as being characterised by different degrees of salience and choice relevance. The salience of a proposition is related to the frequency with which that proposition is employed in the agents’ representations of a situation (Bacharach 2003, 2006), whereas choice relevance refers to the ability to influence decisions (Gold 2012). Both salience and choice relevance are typically time- and context-dependent.

1.3.2 Breaking the symmetry

A more general definition of basis for CRtB can now be given.

**Definition 1** (Basis for common reason to believe). Let \( P \) be a population and \( x \) be a proposition. A state of affairs \( A \in \mathcal{A} \) is a basis for common reason to believe in \( P \) that \( x \) if and only if:

\[
\forall i \in P : \ A \in \mathcal{R}_i [f_i(A)], \quad (1.4)
\]

\[
\forall i \in P : \ A \in [f_i(A) \ \text{ind}_i x], \quad (1.5)
\]

and

\[
\forall i, j \in P : \ A \in [f_i(A) \ \text{ind}_i \mathcal{R}_j [f_{i;j}(A)]]. \quad (1.6)
\]

other fellow’s shoes, to the best of our ability.’ More recent contributions on folk psychology generally refer to this as Simulation Theory (R. M. Gordon 1986).
Condition (1.4) states that if world $A \in \mathcal{A}$ obtains, then each $i \in P$ frames it as $f_i(A)$ and has reason to believe $f_i(A)$ to hold. Thus, $A$ is neither perfectly nor identically perceived by members of $P$. Note that it may be the case that $A \notin f_i(A)$ (so that agent $i$ completely misperceives $A$) and yet $A \in \mathcal{R}_i[f_i(A)]$ (meaning that $i$ has reason to believe in their view of $A$). This is because, for any proposition $y$, the $\mathcal{R}_i$ operator need not satisfy the truth axiom $\mathcal{R}_i(y) \Rightarrow y$. Condition (1.5) says that, in world $A$, $f_i(A)$ indicates to each $i$ that $x$. The perception-inference process corresponding to conditions (1.4) and (1.5) is represented in the right-hand panel of Figure 1.1. If condition (1.6) holds for some $A$, then for each pair of individuals $i, j \in P$, having reason to believe that $f_i(A)$ holds indicates to $i$ that $j$ has reason to believe $f_{i;j}(A)$ to hold — the latter being $i$’s conjecture of $j$’s view of $A$, which may differ to a large extent from $j$’s actual view.

Sometimes (as in the example of Section 1.4) agents might have reason to believe that other members of $P$ have the same view of $A$ as themselves, or at least that they do so in all important respects. If this is the case, then $f_i(A) = f_{i;j}(A)$ for all $i, j \in P$, and (1.6) reduces to:

$$\forall i, j \in P : A \in [f_i(A) \text{ ind}_i \mathcal{R}_j[f_i(A)]] \quad (1.6')$$

which reflects a mutual ascription of common views. The condition in (1.6’) states that, in world $A$, $f_i(A)$ indicates to each $i$ that each $j \neq i$, too, has reason to believe that $f_i(A)$ holds. Again, this by no means implies an actual commonality of views.

Our ancillary condition for the development of common reason to believe is given by the following definition.

**Definition 2** (Ascription of mutually consistent inductive standards). Let $y$ be a proposition. In world $A \in \mathcal{A}$, members of $P$ ascribe mutually consistent inductive standards to each other if and only if:

$$\forall i, j \in P, \forall y : A \in [f_i(A) \text{ ind}_i y] \Rightarrow A \in \mathcal{R}_i[f_{i;j}(A) \text{ ind}_j y]. \quad (1.7)$$

The intuition behind (1.7) is that, although recognising that other agents’ views may differ from one’s own, an individual may have reason to believe that these views all license an inference to proposition $y$. For instance, in our earlier example: (i) Bob’s view indicates to him that the meetings with Alice are not going to stop any soon; (ii) Bob is aware that Alice sees him as a friend, i.e., that Alice’s view differs from his own; and yet (iii) Bob has reason to believe that Alice’s view (as he believes it to be) indicates to her that the weekly meetings are going to continue.
More precisely, to say that (1.7) holds for some $A$ is to say that, for every $i, j \in P$, if $f_i(A)$ indicates to $i$ any proposition $y$, then $i$ has reason to believe that $j$’s view of $A$ (as $i$ conjectures it to be!) indicates to $j$ that $y$. As in Cubitt and Sugden (2003, p. 189, 192), this condition is stronger than necessary but allows to keep technicalities from dominating the exposition. In the following we will not require the material implication to hold for all $y$, but only for a limited set of propositions that are required for the development of higher-order reason to believe.

Definition 2 encompasses those cases where agents have reason to believe that others share their same inductive standards, that is, where:

\[ \forall i, j \in P, \forall y: \ A \in [f_i(A) \ \text{ind}_i \ y] \Rightarrow A \in R_i [f_i(A) \ \text{ind}_j \ y]. \quad (1.7') \]

Condition (1.7’) states that if $f_i(A)$ indicates any proposition $y$ to $i$, then $i$ has reason to believe the same to hold true for any other person in $P$. This condition corresponds to Lewis’ ‘suitable ancillary premise’ about the agents’ reasoning, and dovetails with one provided by Perea (2007) to discuss the epistemic foundations of Nash equilibrium. As Perea shows, in games with at least three players, Nash equilibrium requires for each player to have ‘projective beliefs.’ Simply put, if $i$ has projective beliefs, then their belief about $j$’s choice is the same as their belief about $k$’s belief about $j$’s choice.

Conditions (1.4)–(1.7), together with axioms (1.1)–(1.3), yield the following proposition, which is a frame-dependent version of Lewis’ main result. If a state of affairs $A$ represents a basis for common reason to believe in $P$ that $x$, and if members of $P$ ascribe mutually consistent inductive standards to each other, then each $i \in P$ can form higher-order beliefs about $x$.

**Proposition 1.** If the conditions in Definitions 1 and 2 hold for some $A \in \mathcal{A}$, then $A \in R^P(x)$.

**Proof.** The proposition follows straightforwardly by applying our definitions to the chain of logical implications used by Cubitt and Sugden (2003, Appendix 1) to formalise Lewis’ CRtB.

\begin{align*}
\forall i \in P: & \quad A \in R_i [f_i(A)] \quad \text{from (1.4)} & (P1.1) \\
\forall i, j \in P: & \quad A \in [f_i(A) \ \text{ind}_i \ R_j [f_{ij}(A)]] \quad \text{from (1.6)} & (P1.2) \\
\forall i \in P: & \quad A \in [f_i(A) \ \text{ind}_i \ x] \quad \text{from (1.5)} & (P1.3) \\
\forall i \in P: & \quad A \in R_i (x) \quad \text{from (P1.1), (P1.3), and (1.1)} & (P1.4)
\end{align*}
\[
\forall i, j \in P : \quad A \in R_i[f_{i,j}(A) \text{ ind}_j x] \quad \text{from (P1.3)}
\]
and (1.7) \hspace{1cm} (P1.5)

\[
\forall i, j \in P : \quad A \in [f_i(A) \text{ ind}_i R_j(x)] \quad \text{from (P1.2), (P1.5),}
\]
and (1.3) \hspace{1cm} (P1.6)

\[
\forall i, j \in P : \quad A \in R_i[R_j(x)] \quad \text{from (P1.1), (P1.6),}
\]
and (1.1) \hspace{1cm} (P1.7)

\[
\forall i, j, k \in P : \quad A \in R_i[f_{i,j}(A) \text{ ind}_j R_k(x)] \quad \text{from (P1.6)}
\]
and (1.7) \hspace{1cm} (P1.8)

\[
\forall i, j, k \in P : \quad A \in [f_i(A) \text{ ind}_i R_j[R_k(x)]] \quad \text{from (P1.2),}
\]
(P1.8), and (1.3) \hspace{1cm} (P1.9)

\[
\forall i, j, k \in P : \quad A \in R_i[R_j[R_k(x)]] \quad \text{from (P1.1), (P1.9),}
\]
and (1.1) \hspace{1cm} (P1.10)

\ldots \text{and so on, up to } A \in R^P(x). \quad \square

Line (P1.8) follows from the fact that (P1.6) holds for any pair of individuals in P. Thus, having reason to believe that \( f_i(A) \) holds indicates to i that \( R_k(x) \). Lines (P1.4), (P1.7), and (P1.10), together with the subsequent steps, establish the result.

The proposition relaxes Lewis’ result by providing weaker sufficient conditions for the development of higher-order reason to believe. The distinction between a state of affairs, an individual’s private view, and an individual’s conjecture of other agents’ views allows us to distinguish between two different types of heterogeneity in reasoning. First, different individuals may frame \( A \) in ways that differ considerably from one another. Second, an individual may have beliefs about other agents’ beliefs that are at odds with these agents’ actual beliefs. In principle, there might even be cases where \( f_i(A) \cap f_j(A) \) and \( f_{i,j}(A) \cap f_j(A) \) are both equal to the empty set. Moreover, as long as the course of action is consistent with the agents’ mental models, players do not receive any evidence that their views are inaccurate and their beliefs false. The following example illustrates this point.

1.4 An extended example

The understanding of the American Revolution put forward by Jack Rakove, Barry Weingast, and co-authors (Rakove 1996; Rakove et al. 2005; de Figueiredo et al. 2006) presents several interesting similarities with our framework. As historians have long recognised, the revolutionary
war ‘becomes comprehensible only when the mental framework [...] into which the Americans fitted the events of the 1760s and 1770s, is known’ (Wood 1966, p. 162, cit. in Rakove 1996). The century of Anglo-American cooperation that preceded the 1764-1776 crisis, moreover, offers an example of regularity in behaviour that is conventional (i.e., customary, expected by each party, and characterised by a conditional preference for conformity) but cannot be adequately explained by symmetric Lewisian models.

Underlying cooperation were profound disagreements over the nature of the rule of law and the division of authority between motherland and colonies, of which neither Americans nor British were aware. Long-established practices had been based on misunderstandings that did not emerge until the British parliamentary intervention in colonial affairs became a matter of paramount importance.

Colonists understood the constitution of the empire as a system of common law rules built on precedent (Reid 1995). Between the late seventeenth and the mid-eighteenth centuries, American control over domestic affairs had continued to grow towards what Greene (1991) calls ‘negotiated authority’: the motherland had retained power over issues related to security and international trade, while colonies had come to administer internal affairs that ranged from religion to taxation and the enforcement of contracts. As North (2005b, p. 1006) put it, ‘to the extent that Americans thought about the metropole, it was that Britain was a benign, if remote, presence.’ Moreover, in the eyes of the Americans, well-established customs and self-government practices had to be given a pre-eminent legal status. The fact that the British had long ceased to directly intervene in colonial domestic matters was seen as the foundation for a constitutional claim; sovereignty within the empire, colonists thought, was divided, and the motherland lacked the authority to regulate colonial domestic affairs.

However, on the other side of the Atlantic, a different view of common law constitutionalism had matured. By the middle of the eighteenth century, legislative bodies had considerably extended their control over the executive authority of the Crown. The King-in-Parliament principle had been firmly established, and parliamentary sovereignty had emerged as the core of the British constitution; the Houses had been vested with the power to create, amend, or abolish any law. Furthermore, the British saw the division of authority between motherland and colonies as being motivated by mere expediency. For them, the structure of the empire was a matter of policy choice having no constitutional status. Allowing the colonists control over their domestic affairs was a privilege,
not a right. [...] Moreover, the British had no reason to believe that the colonists saw these issues any differently (de Figueiredo et al. 2006, p. 389).

The British view had therefore little to do with the American ideas of shared sovereignty and self-government practices as acquired rights which restricted parliamentary authority. British policy-makers saw sovereignty as indivisible, and the Houses as having supreme legal power in both motherland and colonial domains.

For decades, Anglo-American relations remained peaceful because of a mutual unawareness of these different, and difficult to reconcile, views. This caused both parties to rely on incorrect but self-confirming beliefs. For the British, what underpinned cooperation was the threat posed by France, both in Europe and overseas. Colonists feared the French and their native allies, too, but an equally compelling reason to cooperate was provided by their being fine with the political independence they had gained. No evidence emerged to reveal that the two parties’ constitutional ideas were contradictory and their beliefs inaccurate.

Figure 1.2 represents Anglo-American cooperation as the equilibrium of a coordination game. Upper and lower case letters represent the choices of, and the payoffs to, the British and the Americans, respectively. $B$, $S$, $T$, and $W$ stand for best, second best, third best, and worst; $C$ and $D$ denote cooperation and defection. For the British, to cooperate means not intervening in American domestic affairs. For the Americans, it means not revolting against the British. Cooperative relations are recognised by both players as yielding the highest payoff, and each player prefers defecting when the opponent cooperates rather then the other way around. Therefore, the players’ preference orderings over the possible outcomes are given by $(C,c) \succ_{UK} (D,c) \succ_{UK} (D,d) \succ_{UK} (C,d)$ and $(C,c) \succ_{US} (C,d) \succ_{US} (D,d) \succ_{US} (D,c)$, respectively. The game admits two Nash equilibria in pure strategies, $(C,c)$ and $(D,d)$.

Figure 1.2. The Anglo-American cooperation game.

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>c</td>
<td>$B$; $b$</td>
<td>$W$; $s$</td>
</tr>
<tr>
<td>d</td>
<td>$S$; $w$</td>
<td>$T$; $t$</td>
</tr>
</tbody>
</table>

This representation reveals very little about both the players’ view and their (mis)conjectures about the opponent’s view. To apply our framework, let the state of affairs $A$ denote Anglo-American relations
as described above and as seen by an omniscient external observer) at some point in time between the 1690s and the 1760. Also, let \( f_{US}(A) \) and \( f_{UK}(A) \) denote the contextual features of \( A \) that the Americans and the British considered as relevant. For ease of exposition, and coherently with Rakove’s narrative, suppose that conditions (1.6’) and (1.7’) hold; this reflects the fact that each party had no reason to expect the other country to frame \( A \) in a significantly different way from themself. The American and British views of \( A \) can be thought of as:

\[
f_{US}(A) = \{ \text{common law constitutionalism built on precedent;} \]
\[
\text{colonial self-governance as an acquired right;}
\]
\[
\text{negotiated authority as a satisfying status quo} \}
\]

\[
f_{UK}(A) = \{ \text{parliamentary supremacy;}
\]
\[
\text{colonial self-governance as a matter of policy choice;}
\]
\[
\text{negotiated authority necessary under the French threat} \}
\]

Finally, let \( x \) correspond to the proposition ‘the existing status quo should be maintained.’ Table 1.1 summarises the modes of reasoning that underlay the two parties’ behaviour. Lines one, two, and three show that, notwithstanding the significant differences in the British and American views of \( A \), the latter represented a basis for common reason to believe that \( x \). Line four represents the auxiliary condition used to allow higher-order expectations about cooperation to come about.

Table 1.1. The reasoning schemas underpinning Anglo-American relations.

<table>
<thead>
<tr>
<th>British point of view</th>
<th>American point of view</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1.4) ( A \in R_{UK}[f_{UK}(A)] )</td>
<td>( A \in R_{US}[f_{US}(A)] )</td>
</tr>
<tr>
<td>(1.5) ( A \in [f_{UK}(A) ; \text{ind}_{UK} x] )</td>
<td>( A \in [f_{US}(A) ; \text{ind}_{US} x] )</td>
</tr>
<tr>
<td>(1.6’) ( A \in [f_{UK}(A) ; \text{ind}<em>{UK} R</em>{US}[f_{UK}(A)]] )</td>
<td>( A \in [f_{US}(A) ; \text{ind}<em>{US} R</em>{UK}[f_{US}(A)]] )</td>
</tr>
<tr>
<td>(1.7’) ( A \in R_{UK}[f_{UK}(A) ; \text{ind}_{US} x] )</td>
<td>( A \in R_{US}[f_{US}(A) ; \text{ind}_{UK} x] )</td>
</tr>
</tbody>
</table>

France’s defeat in the Seven Years’ War produced a major environmental shock, which induced a change in the two parties’ willingness to cooperate and eventually led to an open conflict. For the first time in years, in order to cover the costs of war, Britain sought to intervene in colonial domestic affairs. Most laws, such as the 1764 Sugar Act and the 1765 Stamp Act, introduced relatively minor taxes. Yet they astounded Americans, because they were inconsistent with the colonists’ long-held beliefs. The American view of common law constitutionalism caused colonists to interpret the British interventions as unlawful; the British, in
turn, interpreted the American discontent as fading loyalty towards the empire and reacted by suspending the New York Colonial Assembly. As events unfolded and the revolutionary crisis escalated, the cooperative equilibrium rapidly fell apart.

1.5 Related literature

The case of pre-1760s Anglo-American relations can be recognised as an example of ‘spurious unanimity,’ that is, as a situation characterised by unanimity of preferences without unanimous reasons (Mongin 2016). It also represents a case of self-confirming behaviour, a concept introduced into the economics literature by the aforementioned work by Hayek (1937, p. 51): ‘we may [. . .] very well have a position of equilibrium only because some people have no chance of learning about facts which, if they knew them, would induce them to alter their plans.’ Along this same line of thought is Hahn (1977), who describes an economy where agents formulate conjectures about market conditions from the private signals they receive. The agents’ hypotheses may be imprecise and yet lead to a ‘conjectural equilibrium,’ i.e., a set of mutually consistent signals and individual actions that confirm and induce each other.

In game theory, the concept of self-confirming equilibrium was introduced by Fudenberg and Levine (1993) as a coarsening of Nash equilibrium such that no player ever observes a play that contradicts her beliefs — which need not be correct at off-equilibrium paths. Battigalli et al. (2015) study self-confirming equilibria in situations where players do not know the probabilistic model underlying the variables affecting their choices. If the game is played repeatedly in a stationary environment, then agents can learn the distribution of payoffs associated with observed strategy profiles. However, ambiguity about unchosen strategies persists and makes them less appealing. Schipper (2018) introduces a notion of self-confirming equilibrium that applies to games with unawareness, that is, games in which individuals may not be aware of some of the choices that can be made by other players. He shows that rational play can enrich the players’ information sets, and formalises the representation of a game as the endogenous result of strategic interaction and learning. An epistemic game theoretic approach to reasoning in games with unawareness has been proposed by Perea (2017), who studies the concept of common belief in rationality in that setting.

The role of the agents’ views is also central to Greenberg et al.’s (2009) analysis of mutually acceptable behaviours. The authors show that, although viewing the same interaction problem as different extensive form
games, players may follow a common course of action. Kaneko and Atsui (1999) and Kaneko and Kline (2008) propose an inductive framework to explain how individuals develop and modify their views of the strategic environment. In their models, agents do not have a priori knowledge of the structure of the game they play, and gather information by means of occasional random moves.

Another means of studying conventions is through evolutionary game theoretic models, which rely on random perturbations to introduce variations in the frequencies with which strategies are played by individuals in a population (e.g., Young 1998; Skyrms 2014). Information processing is assumed to be costly, and players myopically adapt to the behaviour of other agents. This allows to investigate the roles of mutation and adaptation in determining long-run behavioural patterns, but also results in a downplay of reasoning. This makes the evolutionary approach a valuable complement to, but not a substitute for, Lewisian models of behaviour.

1.6 Concluding remarks

Ideas from psychology and cognitive sciences have recently started to permeate theories of decision making, resulting in a variety of models that emphasize the role of mental states as drivers of behaviour. In a related field of inquiry, Lewisian models have been investigating how experience and reasoning can allow the emergence of behavioural regularities.

Building on an ambiguity in previous research about whether or not conventions require symmetric reasoning, this chapter has sought to expand Lewis’ framework by integrating it with the concept of ‘frame.’ We have stressed that different agents can understand the same situation differently, and that individuals may believe others’ beliefs to differ from their own. This diversity pervades our lives, as shown through examples ranging from day-to-day activities to deviant behaviours and international relations. The main contribution of the chapter has been to show that individuals can develop consistent higher-order expectations even when their modes of reasoning differ considerably from one another. This has also allowed us to show that beliefs about other people’ mental models may not correspond to truth. As a result, self-confirming patterns of behaviour may emerge that are sustained by false beliefs.

Needless to say, many of the aforementioned points would benefit from further study. For example, the topological properties of framing structures have been largely left unspecified, and a formal definition of when agents perceive particular features of a situation as salient
or choice relevant is still missing. Finally, embedding the model in a
dynamic framework would allow us to elucidate how the revision of
frames and beliefs works. Research along these lines can provide important
new insights into the analysis of the reasoning processes that underlie
behaviour.
2. Pairwise Imitation and the Evolution of the Social Contract

2.1 Introduction

As Jean-Jacques Rousseau (1762, p. 172) famously argued, ‘[t]here is often a great deal of difference between the will of all and the general will.’ The distinction recalls, to some extent, that between social welfare and individual well-being. Rousseau’s general will concerns the common interest aspect of interactions: ‘it is always right,’ and aims at the good of all people. Runciman and Sen (1965) view it as always fulfilling the conditions of Pareto optimality. The will of all, in contrast, results from the aggregation of private, particular interests. Despite being aware of this tension, Rousseau had a strong faith that individuals could implement the common good: ‘if, when the people, being furnished with adequate information, held its deliberations, the citizens had no communication one with another, the grand total of the small differences would always give the general will, and the decision would always be good’ (p. 173).\footnote{Regrettably, this is not necessarily the case. See Camerer (2003) and Henrich et al. (2004) for game theoretic arguments and experimental evidence.}

A less triumphant, yet perhaps more sound, claim can be made with respect to the opposite case: when decisions are made on the basis of inadequate information, bad outcomes are to be expected. This chapter develops this latter perspective, exploring how myopic pairwise comparisons can shape long-run patterns of behaviour.

The distinction between the general will and the will of all can also be framed by contrasting mutual advantage and individual risk. The choice, to be taken independently of other people and without prior knowledge of their decisions, is between risky cooperation on the one hand and safe defection on the other. The cooperative action generously rewards those who succeed in coordinating on it, whereas defection performs relatively better in case of miscoordination. The game that is used to represent situations of this type is the Stag Hunt — after a story, again by Rousseau, in which two hunters have to join forces to kill a stag but are tempted to
abandon the hunt and catch a hare by themselves instead. The decision to hunt a stag involves a risk of going back home empty handed, since there is a possibility that the other hunter will go for a hare. To catch a hare means to assume no such risk but also to forsake the potential gains from a successful stag hunt. Both mutual hare hunting and cooperating by jointly hunting a stag are Nash equilibria, as each action is a best response to itself.

Following Binmore (1994) and Skyrms (2004, 2014), one may recognise the Stag Hunt as a prototypical representation of the social contract, where the state of nature corresponds to the everyone-for-themself equilibrium, or to a nonequilibrium state, and the social contract takes the form of a Pareto improving reform. Therefore, the implementation of the contract poses a problem of equilibrium selection:

you can either devote energy to instituting the new social contract or not. If everyone takes the first course, the social contract equilibrium is achieved; if everyone takes the second course, the state of nature equilibrium results. But the second course carries no risk, while the first does (Skyrms 2004, p. 9).

We study the evolution of behaviour in a population of agents who repeatedly engage in a Stag Hunt game. Interactions are modelled as pairwise random encounters, and players are occasionally called to revise their actions. The resulting long-run conventions are derived by applying stochastic stability theory and studying which equilibria are easiest to flow into when players make mistakes with a small probability (Foster and Young 1990; Kandori, Mailath, et al. 1993; Young 1993). Importantly, we focus mainly on adaptive rules based on imitation, which has long been recognised as a common form of social learning (Hurley and Chater 2005a,b; Rendell et al. 2010). Imitative heuristics can serve multiple purposes: they are cognitively and informationally parsimonious (Gigerenzer and Gaissmaier 2011), allow to free ride on the knowledge of other people, and can be employed as signalling devices (Cho and Kreps 1987). The tendency of individuals to engage in social comparisons and imitative behaviours may stem also from a drive for self-evaluation and a desire to be as good as their peers (Festinger 1954).

A key feature of our model consists in using an interact-and-imitate revision protocol of the kind studied by Bilancini, Boncinelli, and Campigotto (2019). Revising players compare their payoff with the payoff of their last opponent, and copy the opponent’s action if and only if it did better than their own action. We contrast this protocol with two other rules, namely, imitation of a randomly selected individual and best response to the current state of play. Owing to its complexity, a
best response to a state is seen as an instance of informed, deliberative decision making. Contrariwise, pairwise imitative protocols are relatively effortless, rely on information at the local level only, and meet needs for quick action.

Depending on payoffs and on how encounters take place, best response can select either the rewarding All Stag equilibrium or the riskless All Hare equilibrium as the long-run stochastically stable outcome. Pairwise imitation, on the other hand, selects the Pareto inferior equilibrium; when decisions are driven by short-sighted pairwise comparisons, and particularly by comparisons with those whom individuals interact with, it is very difficult for a society to settle upon a social contract. This result is robust to the introduction of assortment, which is often thought of as one of the main forces underlying the emergence of cooperation, and to allowing for some heterogeneity in the behavioural rules used by players.

The chapter is structured as follows. Section 2.2 reviews the relevant literature and discusses some aspects of the revision protocols that we consider. Section 2.3 introduces the model. Section 2.4 presents our main results and assesses their robustness to alternative specifications. Section 2.5 contains some concluding remarks.

2.2 Related literature

Adaptive rules, or revision protocols, can be classified according to several criteria. For example, a distinction can be made with respect to what drives behaviour; rules of the kind ‘copy the first person you meet’ make actions depend solely on their frequency distribution, whereas other protocols assume decisions to be a function of observed payoffs. Differences can also be observed in how much information is needed for decisions to be made. Some rules require players to observe a single individual or a small subset of the population; other rules impose stronger requirements.

A taxonomy of payoff-dependent revision protocols has been proposed by Sandholm (2010). The less demanding class only requires for individuals to know the payoff to their (or to another player’s) strategy. For instance, Bjrönerstedt and Weibull (1996) examine the case where players change strategy depending on the difference between their payoff and some aspiration level. Next come protocols that require players to observe both their payoff and the payoff of another individual. This class of rules includes the imitate-if-better protocols considered in this chapter, which prescribe copying another player if and only if that player’s payoff exceeds one’s own payoff. Finally, at the opposite end of the taxonomy are those protocols that assume players to know either the average payoffs to all
strategies over the entire population or the single pieces of information necessary to compute average payoffs (e.g., Robson and Vega-Redondo 1996). Rules of this kind can allow agents to work out the potential risks and benefits of a decision. However, the information requirements of these protocols become harder to meet as a population grows.

An important feature of many rules based on pairwise comparisons is that they assume random sampling of candidate strategies. Whenever a player receives a revision opportunity, they select another player at random for comparison purposes and copies that player’s behaviour with a probability that depends on the difference in their payoffs (Schlag 1998, 1999; Alós-Ferrer and Schlag 2009; Sandholm 2010; S. S. Izquierdo and L. R. Izquierdo 2013; Arieli and Young 2016). The random sampling assumption induces a separation between the structure of interactions, which specifies how players meet to carry out their business, and the agents’ reference groups, which consist of the individuals with whom they compare themselves. However, this approach is not always representative of reality; often there are situations where agents can only observe, and act upon, the behaviour of those with whom they interact. The revision protocol considered in this chapter, named interact-and-imitate, yields a class of evolutionary game dynamics in which interaction and observation overlap, that is, in which players use their respective opponents as a reference for comparison and imitation. The same idea can be found in the literature on games on networks, where individuals only play against, and adapt to, their nearest or second nearest neighbours (Alós-Ferrer and Weidenholzer 2008, 2014; Fosco and Mengel 2011; Tsakas 2014).

As for the structure of interactions, evolutionary and learning models of the Stag Hunt often consider settings in which players are randomly matched in a pairwise fashion (e.g., Kandori, Mailath, et al. 1993; Young 1993; Kandori and Rob 1995; Rankin et al. 2000). Bilancini and Boncinelli (2018) consider the case where each pair of interacting agents can break up with positive probability after every round of play. They show that if players follow a myopic best response rule, and if mistakes become less likely as realised payoffs increase, then unstable interactions cause inefficient equilibria to emerge. Other models assume that agents play against a fixed set of neighbours (Ellison 1993; Skyrms 2004). Depending on the underlying topology of interactions, the long-run convention can correspond to either the payoff dominant equilibrium or the risk dominant equilibrium, the latter being the equilibrium that minimises the product of the players’ losses from unilateral deviations (Harsanyi and Selten 1988). Yet another approach to modelling interactions is to assume that players strategically choose which action to take and whom to play with (Goyal
and Vega-Redondo 2005; Santos et al. 2006; Staudigl and Weidenholzer 2014). When the number of individuals with whom a player can interact is bounded above by a sufficiently small number, the Pareto superior equilibrium can emerge as the long-run convention.

Finally, models featuring random encounters can either exhibit assort-ment or not. Assortative matching reflects the tendency of similar people to group together, and has been shown to be an important factor in the evolution of cooperation (Eshel and Cavalli-Sforza 1982; Bergstrom 2003; Allen and Nowak 2015; Cooney et al. 2016). Intuitively, when matching is assortative, the risk of cooperating in a social dilemma can be offset by a higher probability of playing against other cooperators. Assortment may be driven by the sharing of cultural, socio-economic, and other individual characteristics (Alger and Weibull 2013, 2016), by similarity in behaviour (Bilancini, Boncinelli, and Wu 2018), or both.

2.3 The model

Consider a finite population \( P = \{1, 2, \ldots, n\} \), with \( n \) an even number. Members of \( P \) are repeatedly matched in pairs to play a stage game. Time unfolds discretely and is indexed by \( t = 0, 1, 2, \ldots \). Assume uniform random matching of agents and denote player \( i \)'s opponent at time \( t \) by \( \zeta(i, t) \in P \setminus \{i\} \). Therefore, for every \( j \neq i \in P \):

$$\Pr[\zeta(i, t) = j] = \frac{1}{n-1}. $$

In every period, players choose an action \( a \in A = \{S, H\} \). Let \( \Delta \) be the simplex of probability distributions over \( A \), and let the function \( \pi : A \times A \rightarrow \mathbb{R} \) specify the payoff \( \pi(a, a') \) earned by an agent playing \( a \) against an opponent playing \( a' \).

Figure 2.1. The stage game.

\[
\begin{array}{c|cc}
S & H \\
S & \alpha, \alpha & \beta, \gamma \\
H & \gamma, \beta & \delta, \delta \\
\end{array}
\]

The stage game, represented in Figure 2.1, is a two-player Stag Hunt. Throughout the chapter it is assumed that \( \alpha > \gamma, \delta > \beta \), implying: (i) that \((S, S)\) and \((H, H)\) are both strict Nash equilibria; (ii) that \((S, S)\) is Pareto superior to \((H, H)\); and (iii) that \( H \) is the maximin action, that is, the action that maximises the minimum payoff a player could possibly
receive. Assume also that $\alpha + \beta \neq \gamma + \delta$ so that either $S$ or $H$ is risk dominant. If $\alpha + \beta > \gamma + \delta$, then $S$ is both payoff dominant and risk dominant. When the opposite inequality holds, the risk dominant action is $H$.

Let $n_a(t) \in [0, n]$ be the number of individuals playing action $a \in A$ at time $t$. The aggregate behaviour of the population in each period is described by a state variable $x(t) := (x_S(t), x_H(t)) \in X \subseteq \Delta$, where $x_a(t) = n_a(t)/n$ and $X$ are, respectively, the proportion of individuals playing $a$ and the state space. By definition, $\sum_{a \in A} x_a(t) = x_S(t) + x_H(t) = 1$. We shall refer to the states $(0, 1)$ and $(1, 0)$ as the maximin convention and the payoff dominant convention, respectively, or as the state of nature and the social contract. The risk dominant convention can be either All Hare or All Stag depending on which is the risk dominant action. We take the initial state, $x(0)$, to be exogenous and study the dynamics induced by different adaptive rules.

### 2.3.1 Revision protocols

After each round of play, each player has an independent probability $p \in (0, 1)$ of being selected for action revision. A revision protocol is a rule that maps the information available to an individual into the set of possible actions. We consider three protocols: best response to the current state of play (BR), pairwise imitation of a randomly sampled individual (or pairwise random imitation, PRI), and pairwise interaction-and-imitation (PII). These rules all feature bounded memory — they base decisions solely on observations made in the last round — but differ significantly with respect to the information and computational abilities needed to make choices.

**Definition 3** (best response). The action chosen at time $t+1$ by a revising player, $i$, who best responds to the state of play at time $t$ is:

$$a_i^{t+1} \in B_i(x(t)),$$

where

$$B_i(x(t)) := \left\{ \arg \max_{a \in A} \sum_{a' \in A} \pi_i(a, a') x_{a'}(t) \right\} \subseteq A.$$  

In case of indifference between actions, $a_i^{t+1}$ is determined by tossing a fair coin.

Best responding to a state can be a difficult matter. It requires revising agents to reason at a global (i.e., population) level, to know how many
times each action was taken by members of \( P \), and to compute expected payoffs according to the current distribution of play. Compared to this, pairwise imitative rules place a considerably less heavy burden upon decision makers. Protocols based on pairwise comparisons simply require individuals to observe and possibly copy another player. As such, they are most useful in situations where agents have limited cognitive abilities and lack strategic sophistication, or where it is difficult to gather information on other players. Another way of interpreting the difference between pairwise imitation and best response is to think of them as stylised forms of intuitive and deliberative thinking, respectively (Kahneman 2003, 2011; Sloman 1996). The former is highly accessible, considerably faster, and computationally undemanding, whereas the latter involves a more thorough evaluation of the available options.

Following Alós-Ferrer and Schlag (2009), a pairwise imitative rule can be expressed as a mixed action

\[
\Psi (a, \pi, a', \pi') \in \Delta,
\]

with \( \Psi (a, \pi, a', \pi')_{a'} \) the conditional probability of choosing action \( a' \) after having played action \( a \), having earned payoff \( \pi \), and having observed another agent playing action \( a' \) and receiving payoff \( \pi' \). Let \( \xi (i, t) \in P \setminus \{ i \} \) denote the player to whom agent \( i \) compares themself at time \( t \) — henceforth, agent \( i \)'s reference. Different assumptions about who can be chosen as a reference correspond to different adaptive rules. For example, under uniform random sampling of references we have \( \Pr [\xi (i, t) = j] = 1/(n - 1) \) for every \( j \neq i \in P \). If player \( i \)'s opponent and player \( i \)'s reference are drawn independently of one another, then the probability of \( \zeta (i, t) \) and \( \xi (i, t) \) being the same player \( j \) is:

\[
\Pr [\zeta (i, t) = \xi (i, t) = j] = \frac{1}{(n - 1)^2}.
\]

Using subscripts and superscripts to denote individuals and time, respectively, we can emphasise the distinction between player \( i \)'s opponent and reference by writing the imitative rule as:

\[
\Psi \left[ a^t_i, \pi_i \left( a^t_i, a^\zeta_{\xi(i,t)} t \right), a^\zeta_{\xi(i,t)} t, \pi_{\xi(i,t)} \left( a^t_{\xi(i,t)}, a^t_{\xi(\xi(i,t), t)} \right) \right], \tag{2.2}
\]

where \( a^t_i \) and \( \pi_i (\cdot) \) are the action chosen and the payoff earned by player \( i \) at time \( t \) against an opponent — \( \zeta (i, t) \) — playing \( a^\zeta_{\xi(i,t)} t \); while \( a^t_{\xi(i,t)} \) and \( \pi_{\xi(i,t)} (\cdot) \) are the action chosen and the payoff earned by player \( i \)'s reference — \( \xi (i, t) \) — against an opponent — \( \zeta (\xi (i, t), t) \) — playing
We write \( \Psi[.] \) to indicate the conditional probability that \( i \) has of switching from \( a^t_i \) to \( a^t_\xi(i,t) \), and we assume for simplicity that:

\[
\Psi[a^t_i, \pi_i(\cdot), a^t_\xi(i,t), \pi_\xi(i,t)(\cdot)]_{a^t_\xi(i,t)} = \begin{cases} 1 & \text{if } \pi_\xi(i,t)(\cdot) > \pi_i(\cdot) \\ 0 & \text{otherwise} \end{cases}.
\]

Thus, \( i \) chooses action \( a^t_\xi(i,t) \) only if \( \xi(i,t) \)'s payoff is higher than their own payoff. This gives the following definition.

**Definition 4** (pairwise random imitation). Under PRI, the action chosen at time \( t + 1 \) by a revising player, \( i \), is:

\[
a^t_{i+1} = \begin{cases} a^t_\xi(i,t) & \text{if } \pi_i(a^t_i, a^t_\zeta(i,t)) < \pi_\xi(i,t)(a^t_\xi(i,t), a^t_\zeta(i,t), \pi_\zeta(i,t)(a^t_\zeta(i,t), a^t_i)) \\ a^t_i & \text{otherwise} \end{cases},
\]

where \( \xi(i,t) \) is sampled, independently of \( \zeta(i,t) \), according to a uniform probability distribution over \( P \).

PRI requires a player to ignore, either voluntarily or not, how did they perform relative to their match, and to select a third party as a reference for comparison instead. As we have noted, however, often people only observe or care about the behaviour of those with whom they interact. If this is the case, then \( \xi(i,t) = \zeta(i,t) \) and \( \zeta(\xi(i,t), t) = i \), and the rule in (2.2) can be rewritten as:

\[
\Psi[a^t_i, \pi_i(a^t_i, a^t_\zeta(i,t)), a^t_\zeta(i,t), \pi_\zeta(i,t)(a^t_\zeta(i,t), a^t_i)].
\]

Revising players who focus exclusively on how well they do vis-à-vis their opponents are said to follow a pairwise interact-and-imitate rule.

**Definition 5** (pairwise interaction-and-imitation). Under PII, the action chosen at time \( t + 1 \) by a revising player, \( i \), is:

\[
a^t_{i+1} = \begin{cases} a^t_\zeta(i,t) & \text{if } \pi_i(a^t_i, a^t_\zeta(i,t)) < \pi_\zeta(i,t)(a^t_\zeta(i,t), a^t_i) \\ a^t_i & \text{otherwise} \end{cases},
\]

where \( \zeta(i,t) \) is at once \( i \)'s reference for comparison and \( i \)'s opponent at time \( t \).

For every \( x, x' \in X \) and every \( \kappa \in K = \{ \text{BR, PRI, PII} \} \), let \( T^\kappa_{x,x'} \in [0,1] \) be the probability of moving from state \( x \) to state \( x' \) in one period under revision protocol \( \kappa \), and let \( T^\kappa \) be the matrix with typical element
Transition probabilities are assumed to be time-homogeneous. Together, \(X\) and \(T^x\) define a discrete-time, unperturbed Markov chain on the state space, denoted by \((X, T^x)\). A recurrent class (or absorbing set) is the smallest set \(R \subseteq X\) such that any state in \(R\) is accessible from any other state in \(R\) — meaning that there is a positive probability of moving from any one state to any other in a finite number of steps — whereas no state outside \(R\) is accessible from any state inside it. A singleton recurrent class is called an absorbing state. Conversely, a state that does not belong to any recurrent class is said to be transient.

2.4 Results

2.4.1 Unperturbed dynamics

Prior to discussing the main results, it is useful to identify which switches can occur under each of the revision protocols being considered. Under best response to a state of play, switching to Stag or to Hare depends on how confident a player is that their opponent will choose the same action. For every \(x \in X\) such that \(\alpha x_S + \beta (1 - x_S) > \gamma x_S + \delta (1 - x_S)\), the expected payoff from cooperating exceeds that from defecting. This implies that if \(x_S > (\delta - \beta) / (\alpha - \beta - \gamma + \delta)\), then a revising agent will best respond by playing Stag. If the opposite inequality holds, the player will choose Hare.

Under pairwise random imitation, a cooperator who receives a revision opportunity will switch to Hare if and only if their last opponent and their reference both played \(H\). Contrarily, a switch from Hare to Stag requires for a defector to select a cooperator who received a payoff of \(\pi(S, S) = \alpha\) as a reference. For this to be the case, \(S\) must be the action chosen by both the revising player’s reference and by the reference’s opponent. Finally, pairwise interaction-and-imitation only admits switches from Stag to Hare. Such a switch occurs whenever a cooperator is matched against, and then copies, a defector. On the other hand, no hare hunter will ever choose to switch to Stag because the maximin action always yields a payoff that is greater or equal to that of the opponent.

The peculiar dynamics of switches under PII leads to a clear-cut result: with the exception of the case where everyone cooperates from the very beginning, play will converge to the maximin convention no matter what the initial state is.

**Lemma 1.** Under pairwise interaction-and-imitation, the unperturbed process converges with probability one to \((0, 1)\) from every \(x(t) \in X \setminus \{(1, 0)\} \).
Chapter 2

Proof. See text above. □

The following lemma characterises the asymptotic behaviour of the process described by each revision protocol.

Lemma 2. For all \( \kappa \in \{ \text{BR}, \text{PRI}, \text{PII} \} \), the only absorbing states of the unperturbed process \((X, T^\kappa)\) are \((0, 1)\) and \((1, 0)\).

Proof. Consider any action \( a \in A \) and suppose that \( x_a(t) = 1 \), meaning that no agent in \( P \) played action \( a' \neq a \) at time \( t \). Since \( \pi(a, a) > \pi(a', a) \), a best responder who receives a revision opportunity will always stick to \( a \). Likewise, a revising agent who follows a pairwise imitative rule will continue to play \( a \) because it is impossible to observe (and copy) a reference who played \( a' \). This proves that, under all three revision protocols, \((0, 1)\) and \((1, 0)\) are both absorbing states.

We next verify that no other absorbing state exists. This requires showing that all states but All Stag and All Hare are transient, meaning that either the payoff dominant convention or the maximin convention, or both, are accessible from all states in \( X \setminus \{(1, 0) ; (0, 1)\} \). To establish the result, it suffices to show that there exists a positive probability path from every such state to All Stag or All Hare. We consider each revision protocol in turn.

BR. Suppose that \( a \) is a best response to the state of play at time \( t \) and consider a player, \( i \), who is playing \( a' \neq a \). The probability that \( i \) has of being the only player selected for action revision is \( p (1 - p)^{n-1} \), and the conditional probability of \( i \) switching to action \( a \) is either one (if \( a \) is the unique best response) or \( 1/2 \) (in case of indiffERENCE between \( a \) and \( a' \)). If the switch occurs, the system moves to a state, \( x(t + 1) \), in which \( x_a(t + 1) = \frac{n_a(t) + 1}{n} \) and \( x_{a'}(t + 1) = \frac{n_{a'}(t) - 1}{n} \). If \( n_a(t) + 1 = n \), then either \((1, 0)\) or \((0, 1)\) has been reached. If not, then a recursive application of this argument shows that a state in which \( x_a = 1 \) can be reached with non-zero probability in a finite number of periods.

PRI. We first show that All Hare is accessible from every state in \( \{ x \in X : n_H > 0 \} \). Let \( m(t) \geq 0 \) denote the number of cooperators who are matched against a defector at time \( t \). The probability that a cooperator who received a payoff of \( \pi(S, H) = \beta \) has of being the only player selected for action revision is \( m(t) p (1 - p)^{n-1} \). Furthermore, with conditional probability \( n_H(t) / (n - 1) \), this player will select a defector (whose payoff is either \( \gamma \) or \( \delta \)) as a reference. Since \( \beta < \min \{ \gamma, \delta \} \), the revising agent will switch from Stag to Hare with
probability one. The system will then move to a state, $x(t + 1)$, in which $x_H(t + 1) = \lfloor n_H(t) + 1 \rfloor / n$ and $x_S(t + 1) = \lfloor n_S(t) - 1 \rfloor / n$. A recursive application of this argument proves that All Hare can be reached with positive probability.

Similarly, All Stag is accessible from all states in $\{x \in X : n_S \geq 2\}$. The probability that a defector, who received a payoff of either $\gamma$ or $\delta$, has of being the only player chosen for action revision is $n_H(t) p (1 - p)^{n - 1}$. Moreover, with conditional probability $[n_S(t) - m(t)] / (n - 1)$, this player will select a cooperator who received a payoff of $\pi(S, S) = \alpha$ as a reference. Since $\alpha > \max\{\gamma, \delta\}$, the revising agent will switch from Hare to Stag with probability one and the system will move to a state, $x(t + 1)$, in which $x_S(t + 1) = \lfloor n_S(t) + 1 \rfloor / n$ and $x_H(t + 1) = \lfloor n_S(t) - 1 \rfloor / n$. A recursive application of this argument shows that All Stag can be reached with positive probability whenever $n_S(t) \geq 2$. By contrast, if $n_S(t) = 1$, All Stag is inaccessible because it is impossible to ever observe a stag hunter who received a payoff of $\alpha$.

The result follows from Lemma 1. Since no switch from Hare to Stag can occur, contagion dynamics will drive the process to All Hare with probability one.

Lemma 2 shows that the three unperturbed processes are all nonergodic; eventually they will end up and remain in All Stag or All Hare, but which of the two will be reached depends on the initial state of the system.

### 2.4.2 Stochastically stable states

We proceed by allowing for the possibility of mutations. Suppose that, whenever a player receives a revision opportunity, there is a probability that she will make a mistake or decide to experiment with a new strategy. We make the standard assumption that, with probability $\varepsilon$, a revising agent chooses the action that is not prescribed by the protocol being considered. Conversely, with probability $1 - \varepsilon$ action revision takes place as described in Section 2.3.1.

Denote by $T^{\kappa, \varepsilon}_{x, x'}$ the probability of moving from $x$ to $x'$ in one period under revision protocol $\kappa$ when the system is perturbed by random mutations. Let $T^{\kappa, \varepsilon}$ be the corresponding transition matrix, and let $(X, T^{\kappa, \varepsilon})$ denote the resulting Markov chain. For every $\varepsilon > 0$ this Markov chain is irreducible and regular — meaning that it has exactly one absorbing set consisting of the whole state space and that there exists
v ∈ N such that \((T^{\kappa, \varepsilon})^v\) has all strictly positive entries, respectively. Since the state space \(X\) is finite, regularity of transition probabilities implies that there exists a unique invariant distribution over the action profiles, denoted by \(\mu^{\kappa, \varepsilon}\), which describes the asymptotic behaviour of the process. Irreducibility and regularity, moreover, both imply that the process is ergodic, i.e., that its long-run average behaviour is independent of the initial state. Following Young (1993), we say that a state \(x\) is stochastically stable if \(\lim_{\varepsilon \to 0} \mu^{\kappa, \varepsilon}(x) > 0\). Simply put, a process that is subject to small but persistent random perturbations will spend most of its time in the stochastically stable states.

If action revision is based on best response, the perturbed process exhibits a well-known property: when the probability of mistakes approaches zero, stochastic stability selects the risk dominant action — which, depending on payoffs, may or may not coincide with the payoff dominant action — as the long-run conventional way of playing the game.

**Proposition 2.** Under best response to the current state of play, the only stochastically stable state of the perturbed process is the risk dominant convention.

**Proof.** For every \(x, x' \in X\), define the resistance of a transition from \(x\) to \(x'\) as the minimum number of mutations needed for the process to move from state \(x\) to state \(x'\) in one period. As \(\varepsilon\) tends to zero, \(T^{\kappa, \varepsilon}\) approaches \(T^\kappa\) and \(T^{\kappa, \varepsilon}_{x, x'}\) is in the order of \((p \varepsilon)^r(x, x')\), with \(p\) the probability of receiving a revision opportunity and \(r(x, x') \in \mathbb{N}\) the resistance of the transition. Also, denote by \(\mathcal{R}\) the set of all recurrent classes of the unperturbed process. For any distinct \(\mathcal{R}\) and \(\mathcal{R}'\) in \(\mathcal{R}\), define a \(\mathcal{R}\)-to-\(\mathcal{R}'\) path as a sequence of states that begins in \(\mathcal{R}\) and ends in \(\mathcal{R}'\). The resistance of such a path is the sum of the resistances of all points in the sequence, and the stochastic potential of \(\mathcal{R}'\) is the least resistance over all possible \(\mathcal{R}\)-to-\(\mathcal{R}'\) paths. As shown by Young (1993), a state is stochastically stable if and only if it belongs to a recurrent class with minimum stochastic potential. In our model, \(\mathcal{R} = \{(1, 0), (0, 1)\}\). The stochastic potential of \((1, 0)\) is the minimum resistance over all paths that go from All Hare to All Stag, and the stochastic potential of \((0, 1)\) is the minimum resistance over all paths that go from All Stag to All Hare.

Finally, let the basin of attraction of \(\mathcal{R}\) be a set \(B_\mathcal{R} \subseteq X\) such that, in the unperturbed process, there is a positive probability of moving from any state in \(B_\mathcal{R}\) to a state in \(\mathcal{R}\) in a finite number of periods. Note that the least resistance to moving from \(\mathcal{R}\) to \(\mathcal{R}'\) equals the least resistance to moving from \(\mathcal{R}\) to \(B_\mathcal{R}'\), since no mutation is required to reach \(\mathcal{R}'\) after entering its basin of attraction.
The least resistant path from All Hare to the basin of attraction of All Stag is found as follows. Suppose that the process is in \((0,1)\). For revising agents to switch from Hare to Stag, it must be the case that \(x_S > (\delta - \beta) / (\alpha - \beta - \gamma + \delta) \equiv \varpi\) so that the expected payoff from playing \(S\) exceeds that from playing \(H\). This will happen if and only if \([\varpi n]\) mutations occur, where \([-\cdot]\) denotes the ceiling function. The probability of this event is in the order of \((p\varepsilon)^{[\varpi n]}\). As shown in the proof of Lemma 2, from this point onwards there is a positive probability of reaching \((1,0)\) without further mistakes.

A similar reasoning shows that the minimum resistance to moving from All Stag to the basin of attraction of All Hare is \([((1 - \varpi) n]\). By Young’s theorem, All Stag is stochastically stable if and only if \(\varpi < (1 - \varpi)\). A simple algebraic manipulation reveals that this is equivalent to \(\alpha + \beta > \gamma + \delta\), which is the condition for Stag to be risk dominant. Conversely, All Hare is stochastically stable if and only if \(\alpha + \beta < \gamma + \delta\), that is, if and only if Hare is the risk dominant action. Together, these two cases yield the proposition.

We contrast Proposition 2 with the following result.

**Proposition 3.** Under pairwise random imitation and pairwise interaction-and-imitation, the only stochastically stable state of the perturbed process is the maximin convention.

**Proof.** We consider each imitative protocol in turn.

**PRI.** Recall from the proof of Lemma 2 that the basin of attraction of \((0,1)\) is \(\{x \in X : n_H > 0\}\). In order to reach this set starting from All Stag, a single mutation suffices. Thus, the least resistance to moving from All Stag to All Hare is one. On the other hand, the basin of attraction of \((0,1)\) is \(\{x \in X : n_S \geq 2\}\). The most likely transition from All Hare to this set requires two mistakes, each having a probability of \(p\varepsilon\). Therefore, the least resistance to reaching All Stag from All Hare is 2. (If a single mutation occurs, the process will move back into All Hare because of the impossibility of switching from Stag to Hare.) This proves that All Hare is the only stochastically stable state.

**PII.** As for the case of pairwise random imitation, the basin of attraction of All Hare is \(\{x \in X : n_H > 0\}\). Starting from All Stag, this set is reached after a single mutation. Contagion dynamics will then drive the system to \((0,1)\) without any need for further mistakes. By contrast, a transition from All Hare to All Stag requires no less than
$n$ mutations. Again we have that All Hare is the unique stochastically stable state.

Proposition 2 shows that if individuals are calculating enough to behave as expected payoff maximisers, then we get the familiar result of evolution favouring risk dominance. In intuitive terms, if in the state $(\frac{1}{2}, \frac{1}{2})$ players recognise the expected reward from a successful stag hunt as being sufficiently high to justify the risk of miscoordination, then All Stag will emerge as the long-run convention. The worse news, as shown in Proposition 3, is that pairwise comparisons favour the emergence of the maximin convention independently of whether or not it is risk dominant to cooperate. This can produce significant inefficiencies; when agents employ pairwise imitative rules, the perturbed process spends most of its time in All Hare regardless of how rewarding social coordination is, and transitions to the social contract occur with very low probabilities.

While it is true that pairwise random imitation and pairwise interaction-and-imitation give the same limit result, a noteworthy difference exists between the processes defined by the two protocols. In short, reaching the payoff superior convention is much less difficult under PRI than under PII. As shown in Figure 2.2, the minimum resistance from $(1,0)$ to $(0,1)$, denoted by $r_{SH}$, is one under both rules — meaning that a single defector can trigger a contagion-like phenomenon with non-zero probability. However, consider a state in which everyone but a single player cooperates. Under PRI, there is a positive (and possibly very high) probability that this defector will receive a revision opportunity, select any of the $n-2$ cooperators who earned a payoff of $\alpha$ as a reference, and then switch to Stag. Under PII, on the contrary, the only way for the process to revert back to $(1,0)$ is for the defector to make a mistake and play Stag. As the probability of mutation goes to zero, the probability of this event approaches zero as well. An even greater difference can be observed when considering the minimum resistance from $(0,1)$ to $(1,0)$, denoted by $r_{HS}$. Under PII, moving from the maximin convention to the payoff dominant convention requires that at least $n$ agents make a mistake. PRI makes this transition relatively easier because it allows to reach All Stag after 2 mutations only.

Figure 2.3 shows the results from a simulation of 50,000 rounds of play.\textsuperscript{2} We assume that $\alpha = 10$, $\beta = 0$, and $\gamma = \delta = 3$, meaning that $S$ is both payoff and risk dominant. The probability of receiving a revision

\textsuperscript{2} The MATLAB code is given in the appendix.
Figure 2.2. Least resistant paths under PRI and PII.

\[
\begin{align*}
\text{All Stag} &\quad r_{SH} = 1 \\
\text{All Hare} &\quad r_{HS} = 2 \\
\end{align*}
\]

(PRI)

\[
\begin{align*}
\text{All Stag} &\quad r_{SH} = 1 \\
\text{All Hare} &\quad r_{HS} = n \\
\end{align*}
\]

(PII)

Figure 2.3. Simulation run — \( n = 20, p = 0.50, \) and \( \varepsilon = 0.05 \).

opportunity and the probability of mutation are set to 0.5 and 0.05, respectively. The population consists of 20 agents, and the initial state is \( \left( \frac{1}{2}, \frac{1}{2} \right) \). The figure shows the evolution of the proportion of cooperators, \( x_S \), under the two imitative rules. From the initial state, both processes converge rapidly to a situation in which all non-mutants choose the same action. Under PRI, play can converge either to All Stag or (as is the case in the figure) to All Hare. From that point onwards the process exhibits a punctuated equilibrium-like behaviour, spending the majority of its time in All Hare and occasionally jumping into All Stag. This
pattern is due to the accumulation of stochastic mutations, which can dislodge the process from the ruling convention and tip it into the other convention. The tipping episodes become more frequent as $\varepsilon$ increases: Figure 2.4 shows an example with $\varepsilon = 0.10$. This alternating pattern is never observed when players interact-and-imitate, because PII makes the maximin convention much more robust to random shocks as compared to PRI.

Figure 2.4. Simulation run — $n = 20$, $p = 0.5$, and $\varepsilon = 0.10$.

2.4.3 Assortative pairing

What if encounters were not uniformly random? Depending on context, it might be more appropriate to assume that individuals meet with one another according to some other rule, which may reflect, among other things, geographic proximity or homophilic group selection. Following Bergstrom (2003, 2013), let $\rho(x)$ be the conditional probability that a stag hunter has of being matched against another stag hunter, and $\tilde{\rho}(x)$ be the conditional probability that a hare hunter has of being matched against a stag hunter. Thus, the conditional probability for a stag hunter to be matched against a hare hunter and the conditional probability for a hare hunter to be matched against another hare hunter are, respectively, $1 - \rho(x)$ and $1 - \tilde{\rho}(x)$. $\rho(x)$ and $1 - \tilde{\rho}(x)$ reflect the extent to which
those who engage in similar behaviours are more likely than chance to interact.

The expected proportion of matches in which a stag hunter plays against a hare hunter is \(x_S [1 - \rho(x)]\). Since this corresponds to the expected proportion of matches in which a hare hunter plays against a stag hunter, \(x_H \tilde{\rho}(x) = (1 - x_S) \hat{\rho}(x)\), we have that:

\[
x_S [1 - \rho(x)] = (1 - x_S) \hat{\rho}(x) .
\] (2.5)

Bergstrom’s index of assortment, \(\phi(x)\), is defined as the difference between the probability that an individual playing \(a\) has to meet another agent playing \(a\) and the probability that an individual playing \(a' \neq a\) has to meet an agent playing \(a\):

\[
\phi(x) := \rho(x) - \tilde{\rho}(x) .
\] (2.6)

Equation (2.6) expresses \(\phi(x)\) as the difference between the probability that a cooperator meets another cooperator and the probability that a defector meets a cooperator. Noting that \(\rho(x) - \tilde{\rho}(x) = [1 - \hat{\rho}(x)] - [1 - \rho(x)]\), it is easy to see that this corresponds to the difference between the probabilities that a defector meets another defector and that a cooperator meets a defector. When \(\phi(x) > 0\), matching is positively assortative.

From Equation (2.5) we can write \(\tilde{\rho}(x) = x_S [1 - (\rho(x) - \tilde{\rho}(x))]\). Substituting this into (2.6) and rearranging terms, we obtain:

\[
\tilde{\rho}(x) = x_S [1 - \phi(x)] ,
\] (2.7)

and from (2.6) and (2.7):

\[
\rho(x) = \phi(x) + \tilde{\rho}(x)
= \phi(x) + x_S [1 - \phi(x)] .
\] (2.8)

As previously mentioned, best response causes switches from \(a\) to \(a'\) to occur with positive probability if and only if the expected payoff from playing \(a\) is greater or equal to the expected payoff from playing \(a'\). The resulting process is said to be payoff monotonic, meaning that in every period the proportions of agents playing Stag and Hare grow at rates that are ordered in the same way as the expected payoffs from the two actions (Weibull 1995). Under assortative pairing, the expected payoff of a stag hunter is:

\[
\alpha \rho(x) + \beta [1 - \rho(x)] = \beta + (\alpha - \beta) \rho(x)
= \beta + (\alpha - \beta) [\phi(x) + x_S [1 - \phi(x)]]
= \beta + (\alpha - \beta) \phi(x) + (\alpha - \beta) x_S [1 - \phi(x)] ,
\] (2.9)
while the expected payoff of a hare hunter is:

\[
\gamma \tilde{\rho} (x) + \delta \left[ 1 - \tilde{\rho} (x) \right] = \delta + (\gamma - \delta) \tilde{\rho} (x)
= \delta + (\gamma - \delta) x_S \left[ 1 - \phi (x) \right].
\] (2.10)

Subtraction of (2.10) from (2.9) gives the difference in the expected return from cooperating versus defecting when the population state is \( x \), denoted by \( \omega (x) \):

\[
\omega (x) = \beta - \delta + (\alpha - \beta) \phi (x) + (\alpha - \beta - \gamma + \delta) x_S \left[ 1 - \phi (x) \right].
\] (2.11)

For every \( x \in X \setminus \{(1, 0), (0, 1)\} \), the rate of growth of \( x_S \) and \( \omega (x) \) are of the same sign. In the commonly studied case where the index of assortment is constant with respect to the state variable (e.g., Allen and Nowak 2015; Bear and Rand 2016), \( \phi (x) = \phi \) and switches from Hare to Stag occur only if:

\[
x_S \geq \frac{\delta - \beta - (\alpha - \beta) \phi}{(\alpha - \beta - \gamma + \delta) (1 - \phi)} \equiv \overline{\omega},
\] (2.12)

where \( \overline{\omega} \) is smaller than \( \frac{\delta - \beta}{\alpha - \beta - \gamma + \delta} = \omega \) for every \( \phi > 0 \). In every state in which \( \overline{\omega} < x_S < \omega \), assortative pairing causes revising agents to best respond by playing Stag, whereas uniform random encounters induce best responders to defect. Differently put, assortment promotes the emergence of the payoff dominant convention by reducing the stochastic potential of All Stag from \( \lceil \omega n \rceil \) to \( \lceil \overline{\omega} n \rceil \).

When decisions are based upon pairwise comparisons, this result does not hold. Our imitative dynamics fail to satisfy payoff monotonicity because switches depend only on the ranking of observed payoffs.\(^3\) For example, suppose that the process was in a state where a single agent cooperates and \( n - 1 \) agents defect. If the population was sufficiently small, or the gains from cooperation sufficiently high, the expected payoff from cooperating would still exceed that from defecting. Yet imitation would drive the process towards All Hare, because switches from H to S cannot occur when \( n_S = 1 \). Keeping in mind that it takes two to catch a Stag, it is easy to see that the result of Proposition 3 can be extended to the case of assortative encounters. Under pairwise imitation, the stochastic potential of All Hare is one regardless of the matching rule being considered. Similarly, the stochastic potential of All Stag is 2

\(^3\) A more realistic assumption would be to consider the probability of a switch from \( a \) to \( a' \) as a decreasing function of \( \pi (a, a') - \pi (a', a) \). This, however, would not affect the gist of our argument.
under PRI and \( n \) under PII. Therefore, \((0, 1)\) continues to be the only stochastically stable state.

One feature that differentiates the perturbed imitative processes with assortment from those with uniform random encounters is the speed of adjustment from one convention to another. If cooperators are more likely than chance to meet other cooperators, then the accumulation of random shocks can cause niches of stag hunters to emerge and persist over some time in situations where the prevailing norm is to defect. Similarly, if defectors are very likely to meet other defectors, then niches of hare hunters can persist even if it is customary to cooperate. It is also worth noting that (notwithstanding that stochastic stability selects the maximin convention) in the case of pairwise random imitation, assortment does facilitate cooperation as compared to the case of uniform random encounters. To see this, recall that a revising player who follows the PRI protocol will switch from Stag to Hare if and only if they interacts with a defector and then selects another defector as a reference. Everything else being equal, the probability of this event is always lower under assortative matching than under uniform random matching. A switch from Hare to Stag, on the other hand, requires for a defector to select a cooperator who received a payoff of \( \pi(S, S) \) as a reference — an event that is more likely to happen if individuals meet assortatively than if they do not. Thus, assortment increases the probability of a switch from Hare to Stag and reduces the probability of a switch from Stag to Hare. This property does not carry over to the case where agents compare their payoff with the payoff of their opponent, since PII always precludes defectors from switching to Stag.

### 2.4.4 Heterogeneous revision protocols

So far we have considered the case where all members of \( P \) follow the same revision protocol. However, agents may have different learning opportunities, or may decide to follow different rules at different times. One way to account for this is to suppose that, whenever a player receives a revision opportunity, she has a probability \( q^K \) of following rule \( K \), with \( q^K > 0 \) for every \( K = \{\text{BR, PRI, PII}\} \) and \( \sum_{K \in K} q^K = 1 \). For simplicity, we restrict ourselves to the case of uniform random encounters. We write \( T \) and \((X, T)\) to denote the transition matrix under heterogeneous revision protocols and the resulting unperturbed Markov chain, respectively, and we write \( T^e \) and \((X, T^e)\) to indicate their perturbed counterparts.

Generalising to the case of heterogeneous rules gives the following result.
Proposition 4. If $H$ is the risk dominant action or $n > 2 \left\lfloor \frac{\alpha - \beta - \gamma + \delta}{2(\delta - \beta)} \right\rfloor$, then the only stochastically stable state of the heterogeneous-population perturbed process is the maximin convention. If $S$ is risk dominant and $n \leq 2 \left\lfloor \frac{\alpha - \beta - \gamma + \delta}{2(\delta - \beta)} \right\rfloor$, then both the maximin convention and the payoff dominant convention are stochastically stable.

Proof. Let $B_a^\kappa$ denote the basin of attraction of action $a$ under revision protocol $\kappa$. Since $q^\kappa > 0$ for all $\kappa$, there is a non-zero probability that $(X, T)$ will evolve precisely as $(X, T^{BR})$, $(X, T^{PRI})$, or $(X, T^{PII})$ from every possible initial state. This implies that, in the absence of mutations, All Hare is accessible from every state in the set:

$$B_H = \bigcup_{\kappa \in K} B_H^\kappa = \{x \in X : n_H > 0\},$$

whereas All Stag is accessible from every state in:

$$B_S = \bigcup_{\kappa \in K} B_S^\kappa = \left\{x \in X : n_S \geq \min \left\{ \left\lfloor \frac{n (\delta - \beta)}{\alpha - \beta - \gamma + \delta} \right\rfloor, 2 \right\} \right\}.$$

Starting from All Stag, $B_H$ is reached after a single mutation. From that point it is possible to move into All Hare in a finite number of periods without any further mistake, meaning that the least resistance to moving from $(1, 0)$ to $(0, 1)$ is:

$$r_{SH} = \min \{(r_{SH}^\kappa)_{\kappa \in K}\} = 1.$$

Similarly, the least number of mutations required to move from $(0, 1)$ to $(1, 0)$ is:

$$r_{HS} = \min \{(r_{HS}^\kappa)_{\kappa \in K}\} = \min \left\{ \left\lfloor \frac{n (\delta - \beta)}{\alpha - \beta - \gamma + \delta} \right\rfloor, 2 \right\},$$

which, since $n$ is assumed to be even, equals one if and only if:

$$n \leq 2 \left\lfloor \frac{\alpha - \beta - \gamma + \delta}{2(\delta - \beta)} \right\rfloor,$$  \hspace{1cm} (2.13)

that is, if and only if $n$ is smaller or equal than the largest multiple of 2 that is not greater than $(\alpha - \beta - \gamma + \delta) / (\delta - \beta)$. If the inequality in (2.13) does not hold, then $r_{HS} = 2 > r_{SH} = 1$ and $(0, 1)$ is the unique stochastically stable state. Moreover, if $\alpha + \beta < \gamma + \delta$ (so that $H$ is risk dominant), then $(\delta - \beta) / (\alpha - \beta - \gamma + \delta) > 1/2$ and the condition in (2.13) cannot be satisfied by any positive value of $n$. Therefore, in order for the stochastic potential of All Stag to equal one, $S$ must be the risk dominant action. If the condition in (2.13) holds and Stag risk dominates Hare, then $r_{HS} = r_{SH} = 1$ and both $(1, 0)$ and $(0, 1)$ are stochastically stable. □
Proposition 4 states that if $H$ is risk dominant or the population is sufficiently large, then All Hare is singled out as the unique stochastically stable state of the perturbed process $(X, T^c)$. Conversely, if $S$ is risk dominant and the population is small enough, then both All Stag and All Hare are stochastically stable. The intuition behind the proof is straightforward. Every time a player receives a revision opportunity, they has a positive probability of following a pairwise imitative protocol. Hence, since the stochastic potential of All Hare is one under both PRI and PII, there is a possibility (though by no means the certainty) that the process will transit from the payoff superior convention to the maximin convention after a single mutation. This is true regardless of which is the risk dominant action. On the contrary, the least number of mutations required to move from All Hare to All Stag depends on both payoffs and on the population size. In particular, for a single mutation to induce best responders to cooperate, it must be true that $1/n > (\delta - \beta) / (\alpha - \beta - \gamma + \delta)$. It follows that in order for All Stag to be stochastically stable along All Hare, the population cannot exceed the value given in Equation (2.13). If Hare risk dominates Stag, however, this condition cannot be satisfied. Therefore, when $H$ is risk dominant, the only stochastically stable state is the maximin convention.

To make the ideas concrete and show how great the effect of pairwise imitation is, suppose that $\alpha = 10, \beta = 0, \gamma = \delta = 3$. For All Stag and All Hare to both be stochastically stable, the expected payoff from cooperating when $n_S = 1$ must be greater than the expected payoff from defecting. This, in turn, requires the population to consist of 2 agents only! *Ceteris paribus*, the bound on the number of players becomes less tight as the gains from cooperation increase — or, to put it another way, the Pareto superior convention can emerge in large populations only if the payoff from a successful stag hunt exceeds the payoff from defection by a considerable extent. For example, if $\alpha = 100$, then the population must consist of no more than 32 agents.

2.5 Concluding Remarks

Achievements and behaviours are often assessed based on how they compare to those of other people. Building on this observation, this chapter has presented a simple evolutionary model of the Stag Hunt where individuals revise their actions by engaging in pairwise comparisons. The long-run conventions of the game have been derived using stochastic stability techniques.

We have considered two different imitate-if-better protocols: pairwise
interaction-and-imitation and pairwise random imitation. The former focusses exclusively on how a player performs relative to their opponent. The latter prescribes to imitate one individual drawn at random from the population; this, for example, might be the case if interactions do not require people to meet in person, so that it is not possible to observe the opponent’s payoff. The resulting adaptive processes have been contrasted with each other and with that observed when agents best respond to the current state of play. Best response relies on information at the population level and can be computationally demanding; pairwise imitation is less effortful and more appropriate in situations where it is difficult for individuals to obtain information about the evolution of play.

Under best response, players choose to cooperate only if they find it profitable to do so, that is, only if the expected gains from cooperation exceed those from defection. As a consequence, evolutionary pressure favours the risk dominant convention. Under pairwise imitative protocols, this result does not hold. As we have seen, pairwise interaction-and-imitation only admits switches from Stag to Hare, causing the Pareto inferior equilibrium to be singled out as the long-run convention regardless of how rewarding social coordination is. In contrast, pairwise random imitation allows agents to observe pieces of information from outside their interaction neighbourhood. This makes it possible for defectors to compare themselves with successful stag hunters, and enlarges the basin of attraction of the Pareto superior convention as compared to the case of interaction-and-imitation. However, when the probability of mistakes falls towards zero, the inefficient maximin convention continues to be the unique stochastically stable state.

The stochastic stability of All Hare under pairwise imitative rules, and particularly under pairwise interaction-and-imitation, has been shown to be robust to the introduction of heterogeneous revision protocols and assortative encounters. Assortment, nevertheless, concentrates interactions among cliques or subpopulations and can allow a behaviour to take root within a subset of individuals who interact intensely with each other. Additionally, it is perhaps worth mentioning that our findings are insensitive to assuming asynchronous learning (i.e., a one-agent-per-period revision process) rather than independent revision opportunities.

Our results are in line with the result by Alós-Ferrer and Weidenholzer (2014) that if information is limited to an agent’s neighbourhood and if there are no information spillovers, then imitation causes inefficient conventions to emerge independently of the network structure. They are also consistent with the experimental findings by Proto et al. (2017) that cognitive abilities and personality traits affect strategic behaviour and
the willingness to cooperate in repeated interactions. As we have seen, if individuals rely on cognitively inexpensive, unsophisticated heuristics based on the imitation of another player, it is very difficult to sustain cooperation. This is especially so in the case of pairwise interaction-and-imitation; when individuals see their opponents’ successes as their own failures, and take actions to reduce the difference in their payoffs, the road to establishing a social contract becomes an extremely steep road to take.
In recent years, European countries have witnessed a substantial increase in migration flows. In 2016, an estimated 2 million people migrated to the EU-28 from non-member countries, while 1.3 million moved from one Member State to another (Eurostat 2018). Overall, foreign-born residents account for approximately 11 percent of the European population, with 36.9 million people born outside of the Union living in EU Member States as of January 2017, and 20.4 million born in a Member State different from the one they reside in. The largest volumes of non-national residents are registered in Germany (9.2 million persons), the United Kingdom (6.1 million), and Italy (5.0 million). In proportional terms, the foreign-born presence is particularly relevant in Austria (18.7 percent of the population), Sweden (17.1 percent), Ireland (16.8 percent), Norway (14.8 percent), and Germany (14.2 percent).

These changes have brought about new economic and social challenges, both because of the extent of the phenomenon and because of the very short period over which it has occurred. The study of the effects of migration flows has been pursued in multiple fields that include labour (e.g., D’Amuri and Peri 2014; Dustmann et al. 2013), public finances and welfare (Bratsberg et al. 2010, 2014; Preston 2014), and education (Ballatore et al. 2018; Lüdemann and Schwerdt 2013; Maestri 2017).

From a more general perspective, several studies have been investigating — mostly with reference to the US context — the consequences of migration and ethnic diversity on trust and social cohesion (Putnam 2007; Stolle et al. 2008). As stressed by Bisin et al. (2016), two opposing views characterise the scholarly debate on the issue. The body of literature that stems from assimilation and contact theories (Allport 1954; M. M. Gordon 1964) sees the structure of social interactions as being driven by a preference for conformity and inclusiveness, and positive out-group attitudes among individuals as being favored by cross-ethnic relationships.
The contrasting view (Blumer 1958; Glazer and Moynihan 1970) recognises group identity as the key component of social relations and maintains that ethnic diversity can end up strengthening in-group vs. out-group distinctions.¹

The school environment provides an ideal context for studying integration and segregation processes, both from a substantive viewpoint — education institutions play a fundamental role in shaping civic life — and from a methodological one. School- and class-level analyses, in fact, allow researchers to reconstruct the relevant network of relations and gather detailed information for each node. This is not generally true when considering networks in neighbourhoods and local communities, since complete information on both links and individual characteristics are available only for a subset of nodes in these cases. Often, data on non-interviewed subjects are only partially retrievable and can be obtained only through self-reports from those who do take part in the interviews. For these reasons, relations among schoolmates have been widely examined by scholars interested in social cohesion and homophily, that is, the tendency of individuals to interact and establish relationships with those who are similar to them (Lazarsfeld and Merton 1954).

Homophilic preferences have been documented with respect to many individual characteristics (McPherson et al. 2001), and ethnicity is the most frequently investigated (Currrarini, Jackson, et al. 2009, 2010; Currarini, Matheson, et al. 2016). The role that specific individual traits play in network formation, however, is typically considered in isolation, that is, ignoring the concurrent influence of other attributes. This point, made by Block and Grund (2014) and Rapallini and Rustichini (2016), is central to the evaluation of homophilic interactions. Indeed, ‘despite consensus on the importance of homophily for social relationships as well as on the conceptualization of individuals as multidimensional beings, little is known about how both these combine. Is there a qualitative difference in social relationships when individuals have more than one attribute in

¹ The empirical evidence on the matter remains mixed, and an exhaustive review falls outside the scope of this work. As an aside, it is worth mentioning that it is often perceived ethnic diversity rather than actual diversity that influences attitudes towards minority groups (Koopmans and Schaeffer 2016; Piekut and Valentine 2016). Moreover, there is general agreement on the importance of the spatial unit being considered. The conclusion shared by recent meta-analyses (Meer and Tolsma 2014; Schaeffer 2014) is that the documented effects of ethnic diversity are likely to depend on the size of the unit of observation: studies that find no evidence of negative diversity effects tend to rely on aggregate (e.g., national) levels of analysis, whereas studies that consider smaller spatial units generally find that diversity negatively impacts social cohesion (see, among others, Algan et al. 2016).
common? The literature is surprisingly silent on such multidimensional homophily’ (Block and Grund 2014, p. 192). Studies carried out under the lens of one-dimensional homophily fail to take into account all the individual attributes that intersect, but do not overlap, with the one they are examining. Simply put, similarity in characteristic \( x \) may result in a relationship between individuals that differ in terms of characteristic \( y \).

Assessing the role of characteristics that might be influenced by individuals — such as preferences and abilities — poses the empirical challenge of distinguishing between peer selection and influence. Similarity among people might result from either or both of two mechanisms: while it is true that ‘similarity breeds fellowship’ (McPherson et al. 2001, p. 428), individuals may also end up holding particular values, or engaging in certain behaviours because of socialisation. This issue, which so far has not been adequately addressed by studies on homophilic preferences over potentially endogenous attributes (Kandel 1978; Knoke 1990; Rapallini and Rustichini 2016; S. Smith et al. 2014), is related to the well-known Manski reflection problem (Manski 1993). The possible simultaneity of these forces makes it difficult to disentangle assortative pairing from the influence of one’s peers and the correlated effects induced by exposure to a common environment.

This chapter has two aims. The first is to evaluate how much bonds among individuals grow out of a common ethnic background and how much they grow out of other common individual traits. More specifically, the chapter assesses the relative strength of various dimensions of homophily by simultaneously considering a rich array of individual characteristics, including ethnicity, gender, socio-economic status, and language as well as analytic abilities. The second aim is to test whether or not characteristics such as religion, normative beliefs, and other cultural attributes can be reduced to ethnicity to explain how people choose friends. Indeed, if ethnicity represented other individual traits, the role played by ethnic backgrounds in shaping relationships would be misinterpreted. Assessing bonds by looking at ethnicity and not considering religion, for example, may cause one to infer evidence in support of ethnic homophily; instead, these relationships may be mainly due to sharing the same creed and religious practices.

We begin by introducing a simple model of friendship formation that accounts for both multiple dimensions of homophily and group composition, so as to take into consideration the concurrent influences of individual preferences and mixing opportunities. The model is then estimated using two-wave panel data from a survey of secondary school students in four European countries that are among the more ethnically diverse: England,
Germany, the Netherlands, and Sweden. To address the endogeneity concerns related to the distinction between selection and peer influence, we adopt an empirical strategy that draws upon Bramoullé et al. (2009). As detailed in Section 3.1.3, we identify instruments by exploiting the structure of classroom networks and using the characteristics of an individual’s indirect friends. Consider, for example, an intransitive triad where agent \( i \) is linked to \( j \), who in turn is linked to \( k \). When controlling for correlated effects, agent \( i \)’s characteristics can be seen as being influenced by \( k \) only through \( j \) and do not depend on the peer-specific factors that affect the group consisting of \( j \) and \( k \) (Patacchini, Rainone, et al. 2017). The characteristics of dyad \( ik \) are therefore used to instrument those of dyad \( ij \).

This chapter aims to contribute to the literature on social cohesion along several lines. First, we investigate segregation and friendship patterns among young individuals as they are going through a developmental phase that is crucial for shaping their identity. Second, we adopt a perspective that allows us to assess the relative importance of multiple dimensions of homophily. Third, we make progress in separating the effects of selection and socialisation. The instrumental variable strategy proposed here seems to serve the purpose of evaluating the role and extent of homophilic preferences while taking into account possible influences from an individual’s peers.

The remainder of the chapter is organized as follows. Section 3.1 introduces the model and the estimation strategy. Section 3.2 presents the data and descriptive statistics. Section 3.3 discusses the main results. Section 3.4 provides our conclusions.

3.1 Theory

We model how friendship links among individuals in a group are formed depending on the characteristics of those individuals. Members of the group (e.g., students in a class) are indexed by \( i = 1, \ldots, n \). Each individual is described by a vector \( \theta^i \equiv (\theta^i_k)_{k=1,...,K} \) of characteristics. For example, the first coordinate may describe whether individual \( i \) is female or male, the second may describe whether she is foreign born or not, and so on. The utility of having a link for an individual depends on her characteristics and on the characteristics of the match. Assuming a simple additive form, the value for \( i \) of having a friendship link with \( j \) is:

\[
 w^ij = \sum_{k=1}^{K} \left( \lambda_k I_{\theta^i_k = \theta^j_k} + \mu_k I_{\theta^i_k \neq \theta^j_k} \right),
\]
where \( \lambda \) and \( \mu \) are real-valued parameters, and \( \mathbb{1} \) denote the indicator function. Letting \( \sigma^k_{ij} \equiv \mathbb{1}_{\theta^i_k = \theta^j_k} \) and \( \gamma_k \equiv \lambda_k - \mu_k \), the expression can be rewritten as:

\[
 w^{ij} = \sum_{k=1}^{K} \mu_k + \sum_{k=1}^{K} \gamma_k \sigma^k_{ij},
\]

(3.1)

and assuming preferences to be homophilic, then for every characteristic \( k \) we have that:

\[
 \lambda_k \geq \mu_k \geq 0,
\]

implying \( \gamma_k \geq 0 \). If we indicate with \( C \) the set of individuals that are linked to agent \( i \), then the utility of \( i \) is:

\[
 U^i(C) = \sum_{j \in C} w^{ij} - \#(C)^\alpha,
\]

(3.2)

with \( \alpha > 1 \) and where \( \#(C) \) is the cardinality of set \( C \). The convex cost associated with a bigger \( C \) represents the cost (network management, transportation, and so on) from having a larger group of friends.

3.1.1 Matching

Suppose that individuals have information on the distribution of characteristics in the group. Each agent then sends a link to a set of peers to maximise the utility defined in equation (3.2). This choice is made simultaneously for all individuals, and the links are sent to all agents that one would like to have as friend. If we order the vector of values in a decreasing order of the index, then the optimal size \( c^*_i \) of the group of links for agent \( i \) is the solution of the maximisation problem:

\[
 c^*_i = \max \{ c : w^{ic} \geq (c + 1)^\alpha - (c)^\alpha \},
\]

(3.3)

that is, agent \( c^*_i \) is the last for whom the marginal benefit from adding \( c^*_i \) to the list, \( w^{ic} \), is larger than the incremental cost of adding another friend to the list of friends, \( (c^* + 1)^\alpha - (c^*)^\alpha \). The optimal \( c^* \) exists and is unique: this follows from the definition in equation (3.3), the fact that the marginal cost is only dependent on the size \( c \), and the ordering convention for the values.

3.1.2 Estimation

Let \( L_{ij} \) be a dichotomous variable equal to one when \( i \) chooses to send a link to \( j \) and zero otherwise. The agent chooses to send the link if the value to her of establishing the link with \( j \), \( w^{ij} \), is larger than a
threshold \( w^{i,th} \) that is group and agent dependent. A simple statistics of the distribution of values in the group to agent \( i \) is the average value:

\[
\bar{w}^i \equiv \frac{1}{n - 1} \sum_{l \neq i} w^{il}
\]

\[
= \sum_k \mu_k + \sum_k \gamma_k \left( \frac{1}{n - 1} \sum_{l \neq i} \sigma^k_{il} \right)
\]

\[
= \sum_k \mu_k + \sum_k \gamma_k \bar{\sigma}^k_i,
\]

(3.4)

where \( \bar{\sigma}^k_i \equiv \frac{1}{n-1} \sum_{l \neq i} \sigma^k_{il} \) indicates how many agents in the group have characteristic \( k \) equal to \( i \). The value \( \bar{\sigma}^k_i \) depends on both subject and group, and it can be easily computed from the data. In view of equation (3.3) determining the optimal size of the group of friends, the threshold \( w^{i,th} \) is increasing if the vector of values increases point-wise. We adopt the assumption that the threshold \( w^{i,th} \) is increasing in the average value \( \bar{w}^i \), and even more simply that:

\[
w^{i,th} = \eta \bar{w}^i,
\]

(3.5)

with \( \eta > 0 \). So, an agent \( i \) sends a link to \( j \) if and only if \( w^{ij} > \eta w^{i,th} \).

Finally, the parameters \((\alpha, (\beta \gamma_k)_{k=1, \ldots, K})\) can be determined by specifying a logit model for the conditional probability of sending a link:

\[
Pr \{ L_{ij} = 1 \mid \theta^1, \ldots, \theta^n \} = \Lambda \left[ \alpha + \beta \left( w^{ij} - w^{i,th} \right) \right]
\]

\[
= \Lambda \left[ \alpha + \beta (1 - \eta) \sum_k \mu_k + \beta \sum_k \gamma_k \left( \sigma^k_{ij} - \eta \bar{\sigma}^k_i \right) \right],
\]

(3.6)

with \( \Lambda(\cdot) = \exp(\cdot) / [1 + \exp(\cdot)] \), \( \beta \gamma_k \) the coefficient for the \( j \)-specific similarity \( \sigma^k_{ij} \) and \( \beta \gamma_k \eta \) for the average similarity, that is, the fraction of individuals in the group similar to \( i \) in characteristic \( k \).

### 3.1.3 Empirical strategy

The model of friendship formation yields a simple equation amenable to estimation:

\[
Pr \{ Y_{ij, l, t} = 1 \mid \cdot \} = \Lambda \left( \beta_0 + X'_{ij, l, t} \theta_1 + \bar{X}'_{ij, l, t} \theta_2 + \varepsilon_{ij, l, t} \right),
\]

(3.7)
where the subscripts $i$ and $j$, $l$, and $t$ denote individuals, class, and time, respectively. The dependent variable, $Y_{ij,l,t} = Y_{ji,l,t}$, is a dichotomous variable that assumes value 1 if student $i$ (Ego) indicates classmate $j$ (Alter) to be a friend at time $t$ and vice versa, and 0 otherwise. We thus focus on bilateral, or undirected, friendship relations. The vector of covariates $X_{ij,l,t}$ captures similarities and differences in the characteristics of members of dyad $ij$ (same gender, same nationality, and so on), whereas $X'_{ij,l,t} ≡ [X'_{i,l,t}, X'_{j,l,t}]$ specifies the shares of students in class $l$ similar to $i$ and $j$ in each characteristic at time $t$ — $X_{i,l,t}$ and $X_{j,l,t}$, respectively. This approach allows us to account for how common a given characteristic is within a class, that is, for differences in mixing opportunities.

A known feature of friendship relations, and of social networks in general, is cliquishness. For each triad $ijm$, the probability of a link between $i$ and $j$ is not independent of the presence of links between $i$ and $m$ and between $j$ and $m$. With everything else equal, the likelihood of a link between a pair of individuals is higher if both belong to the same clique. To capture this non-stochastic component of peer selection, we always control for the number of friends that $i$ and $j$ share.

The study of social interactions can also pose a simultaneity problem. To try and separate the role of selection from the roles of peer influence and unobservable variables, we refer to the approach advanced by Bramoullé et al. (2009) for the estimation of peer effects (see also Calvó-Armengol et al. 2009; de Melo 2014; Lin 2010; Patacchini, Rainone, et al. 2017). The main idea is that each individual’s reference group is composed of her best friends. Then, whenever the structure of the social network is known and correlated effects affect all members of the group, the characteristics of the friends of an individual’s friends, who are not her friends themselves, can serve to instrument the behaviour of her own friends. The crucial assumption is that relevant unobservable characteristics can be seen as group specific. As observed by Patacchini, Rainone, et al. (2017), this assumption is most reasonable when networks are quite small — which is true in our case. Correlated effects at the group level, which are caused by exposure to a common environment, can be related to the influence of good or bad teachers, the availability and quality of school facilities, and participation in education or vocational programs. Their presence is accounted for through the inclusion of group fixed effects.

We draw upon this method to fit our needs. Let $Z$ denote the set of students in a class, and for each dyad $ij$, let $Z_{i}^{ij}_{ind}$ and $Z_{j}^{ji}_{ind}$ be the sets of $i$’s indirect friends through $j$ and $j$’s indirect friends through $i$, respectively.
respectively:

\[ Z^{i;j}_{\text{ind}} = \{ m \in \mathbb{Z} \setminus \{i, j\} \mid Y_{im} = 0 \text{ and } Y_{jm} = 1 \}, \]

\[ Z^{j;i}_{\text{ind}} = \{ m' \in \mathbb{Z} \setminus \{i, j\} \mid Y_{im'} = 1 \text{ and } Y_{jm'} = 0 \}. \]

(For simplicity, the class and time subscripts are omitted.) To make the ideas concrete, consider the friendship network depicted in Figure 3.1. Here, for \((i, j) = (2, 3)\), we have \(Z^{2;3}_{\text{ind}} = \{8\}\) and \(Z^{3;2}_{\text{ind}} = \{1, 4, 5\}\). Note that the label ‘indirect friends’ entails a slight abuse of terminology because the definitions of \(Z^{i;j}_{\text{ind}}\) and \(Z^{j;i}_{\text{ind}}\) do not require \(i\) and \(j\) to be friends. In the case of individuals 1 and 9, for example, we have \(Z^{1;9}_{\text{ind}} = \{10\}\) and \(Z^{9;1}_{\text{ind}} = \{2, 4, 5, 7\}\), but \(Y_{1,9} = 0\). We made this choice for ease of exposition, and we are confident that it will cause no confusion.

Figure 3.1. An undirected friendship network

Unobserved group-level heterogeneity is addressed through a within transformation, expressing the model in deviations from class average values. For each dyad in class \(l\), instruments are then identified by extracting and combining information from Alter’s (Ego’s) friends who are not friends of Ego (Alter). By exploiting the symmetry that characterises bilateral friendship relations, it is possible to identify a number of instruments that are twice that of endogenous variables: each endogenous regressor \(x_{ij} \in X_{ij}\) is instrumented by measures of how the characteristics of \(i\) and \(j\) differ from the characteristics of their respective indirect friends. The underlying claim is that, on average, the influence exerted by Ego on Alter when a friendship link exists carries over to Ego’s indirect friends only in a residual — possibly negligible — manner. In contrast, when
$Y_{ij} = 0$, no influence between Ego and Alter is to be expected, since neither of them belongs to the other’s reference group.

Formally, the real-valued variable $x_{ij} = |x_i - x_j|$ is instrumented by:

$$z_{ij} = \left| x_i - \frac{1}{\#(Z_{ind}^{i:j})} \sum_{m \in Z_{ind}^{i:j}} x_m \right|$$

and

$$z'_{ij} = \left| x_j - \frac{1}{\#(Z_{ind}^{j:i})} \sum_{m' \in Z_{ind}^{j:i}} x_{m'} \right|,$$

while the dichotomous variable

$$\tilde{x}_{ij} = \begin{cases} 1 & \text{if } \theta^i_k = \theta^j_k = \tilde{\theta}_k \\ 0 & \text{otherwise} \end{cases}$$

is instrumented by:

$$\tilde{z}_{ij} = \begin{cases} 1 & \text{if } \theta^i_k = \tilde{\theta}_k \text{ and } \sum_{m \in Z_{ind}^{i:j}} I_{\theta^m_k = \tilde{\theta}_k} > \frac{\#(Z_{ind}^{i:j})}{2} \\ 0 & \text{otherwise} \end{cases}$$

and

$$\tilde{z}'_{ij} = \begin{cases} 1 & \text{if } \theta^j_k = \tilde{\theta}_k \text{ and } \sum_{m' \in Z_{ind}^{j:i}} I_{\theta^m'_k = \tilde{\theta}_k} > \frac{\#(Z_{ind}^{j:i})}{2} \\ 0 & \text{otherwise} \end{cases},$$

with the last definitions simply requiring binary instruments to assume a value of 1 if and only if both Ego (Alter) and the majority of her indirect friends through Alter (Ego) share characteristic $\tilde{\theta}_k$. The endogeneity of covariates and the validity of instruments are assessed in Section 3.3 through the usual tests.

As for inference, paired observations pose challenges related to the complex pattern of error correlations. Errors for dyad $ij$, for example, are correlated with errors for any other dyad composed of either $i$ or $j$, including both those for $im$ and $in$ and those for $mj$ and $nj$ (Cameron and D. L. Miller 2014; Tabord-Meehan 2018). This makes simple two-way clustering inadequate. A dyadic-robust variance estimator has been proposed by Fafchamps and Gubert (2007), but it relies on the key assumption that errors for dyads that do not share a member are always uncorrelated. While this assumption can hold for country-pair data on
introduction to class-level friendship relations. Dyadic clustering, moreover, is easily recognisable as being nested into class clustering. We therefore adopt the standard approach of clustering at the higher (class) level of aggregation (Cameron, Gelbach, et al. 2011).

3.2 Data

Homophilic preferences for peer assortment are evaluated using the first two waves of the CILS4EU (Children of Immigrants Longitudinal Study for Four European Countries, Kalter et al. 2016) data set, which allowed us to reconstruct friendship networks in nationally representative samples of secondary schools from four European countries: England, Germany, the Netherlands, and Sweden. This is convenient for our purposes, since all these countries are among the most ethnically heterogeneous in the EU.

The first wave of surveys was administered during academic year 2011-12 when students were aged 15-16, and the second wave was conducted the following year. There were 11,920 and 9,839 respondents, respectively. Since not all subjects answered all of the questions, the number of observations tends to decline as controls increase. The descriptive statistics and demographic characteristics of respondents are summarised in Table 3.1, while the questionnaire items are given in the appendix.

Information on social status, academic performance, religion, and cultural attributes are used as covariates. We describe them in turn. Socio-economic status is assessed by the number of books in an individual’s house, the size of the house measured by the number of rooms, and the education level of the parents. In particular, we consider whether or not at least one of a student’s parents obtained a university degree. Data on these three variables were collected only in wave 1, but they can reasonably be assumed not to have changed over one year. Academic performance is measured by school grades. We focus on achievements in mathematics and the survey country’s language (SCL), which can be thought of as reflecting analytical and communication abilities. The grades of English and Swedish students are available only for wave 2.

Finally, we consider normative beliefs about gender roles, attitudes towards cultural assimilation, and subjective integration. These traits are intended to reflect aspects of individual and group identity that are related to culture but are not entirely reducible to ethnicity or religion. Two questions that asked ‘in a family, who should do the following?’ capture gender-role orientations with respect to earning money and cleaning. We
Table 3.1. Descriptive statistics

<table>
<thead>
<tr>
<th>wave</th>
<th>EN</th>
<th>GE</th>
<th>NL</th>
<th>SW</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2,655</td>
<td>3,079</td>
<td>3,289</td>
<td>2,897</td>
<td>11,920</td>
</tr>
<tr>
<td>%</td>
<td>22.27</td>
<td>25.83</td>
<td>27.59</td>
<td>24.30</td>
<td>100.00</td>
</tr>
<tr>
<td>2</td>
<td>2,154</td>
<td>2,578</td>
<td>2,655</td>
<td>2,452</td>
<td>9,839</td>
</tr>
<tr>
<td>%</td>
<td>21.89</td>
<td>26.20</td>
<td>26.98</td>
<td>24.92</td>
<td>100.00</td>
</tr>
<tr>
<td>N</td>
<td>2,154</td>
<td>2,578</td>
<td>2,655</td>
<td>2,452</td>
<td>9,839</td>
</tr>
<tr>
<td>%</td>
<td>21.89</td>
<td>26.20</td>
<td>26.98</td>
<td>24.92</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Schools

| 1 | 107 | 135 | 100 | 129 | 471 |
| 2 | 97  | 134 | 99  | 127 | 457 |

Classes

| 1 | 207 | 250 | 222 | 251 | 930 |
| 2 | 187 | 242 | 220 | 247 | 896 |

Class size (avg±s.d.)

| 1 | 23.54±6.59 | 21.05±5.27 | 22.36±5.32 | 21.82±4.33 | 22.15±5.47 |
| 2 | 20.61±6.33 | 19.20±5.90 | 18.97±5.28 | 19.63±4.33 | 19.55±5.52 |

% Males

| 1 | 48.15 | 50.95 | 48.45 | 48.20 | 48.97 |
| 2 | 48.39 | 49.90 | 48.27 | 48.55 | 48.80 |

Age (avg±s.d.)

| 1 | 15.34±.49 | 15.79±.74 | 15.54±.63 | 15.02±.25 | 15.43±.63 |
| 2 | 16.34±.48 | 16.75±.72 | 16.50±.61 | 16.01±.24 | 16.41±.61 |

% Natives

| 1 | 48.01 | 45.58 | 60.20 | 46.49 | 50.38 |
| 2 | 46.73 | 46.75 | 61.67 | 47.51 | 50.97 |

% 1st gen. immigrants

| 1 | 12.22 | 8.71 | 5.94 | 10.99 | 9.28 |
| 2 | 12.42 | 8.21 | 5.62 | 10.36 | 8.96 |

% 2nd gen. immigrants

| 1 | 25.69 | 34.50 | 23.17 | 29.93 | 28.30 |
| 2 | 26.70 | 33.55 | 21.71 | 29.45 | 27.83 |

% 3rd gen. immigrants

| 1 | 14.08 | 11.22 | 10.69 | 12.58 | 12.04 |
| 2 | 14.15 | 11.48 | 11.01 | 12.68 | 12.24 |

% Christians‡

| 1 | 38.95 | - | - | 49.67 | 43.77§ |
| 2 | 39.93 | - | - | 49.76 | 44.27§ |

% Catholics

| 1 | - | 31.02 | 14.96 | - | - |
| 2 | - | 31.77 | 14.76 | - | - |

% Protestants

| 1 | - | 31.67 | 9.79 | - | - |
| 2 | - | 30.80 | 10.21 | - | - |

% Muslims

| 1 | 12.09 | 20.69 | 14.26 | 15.91 | 15.84 |
| 2 | 13.32 | 19.12 | 12.84 | 14.40 | 14.98 |

% Atheists/agnostics

| 1 | 38.79 | 11.85 | 53.85 | 30.86 | 34.06 |
| 2 | 38.67 | 12.99 | 55.93 | 32.18 | 34.98 |

% Others

| 1 | 10.17 | 4.77 | 7.14 | 3.55 | 6.33 |
| 2 | 8.08  | 5.32 | 6.26 | 3.66 | 5.76 |

Most represented minorities

<table>
<thead>
<tr>
<th>Pakistan</th>
<th>Turkish</th>
<th>Turkish</th>
<th>ex-Yugoslav</th>
</tr>
</thead>
<tbody>
<tr>
<td>(~7%)</td>
<td>(~15.5%)</td>
<td>(~6%)</td>
<td>(~6%)</td>
</tr>
<tr>
<td>Indian</td>
<td>Russian</td>
<td>Moroccan</td>
<td>Finnish</td>
</tr>
<tr>
<td>(~7%)</td>
<td>(~6%)</td>
<td>(~5.5%)</td>
<td>(~6%)</td>
</tr>
<tr>
<td>Irish</td>
<td>Polish</td>
<td>Surinamese</td>
<td>Iraqi</td>
</tr>
<tr>
<td>(~3.5%)</td>
<td>(~5.5%)</td>
<td>(~5%)</td>
<td>(~4%)</td>
</tr>
</tbody>
</table>

‡: Christian denominations were not assessed in England and Sweden.
§: includes Catholics and Protestants.
are particularly interested in attitudes indicative of more traditionalist views, that is, women should clean the house and men should be the breadwinners. Normative beliefs about cultural assimilation and diversity, instead, are measured by agreement or disagreement with two statements: ‘natives should be open to immigrants’ customs and traditions’ and ‘immigrants should do all they can to keep their customs.’ The answers to both questions assume values on a five-point scale ranging from ‘strongly agree’ to ‘strongly disagree.’ Finally, the strength of the bond with the country of current residence is measured on a four-point scale going from ‘not at all strong’ to ‘very strong.’ Answers can be interpreted as reflecting subjective integration or, alternatively, identification with one’s country.

Table 3.2. Normative beliefs and subjective integration

<table>
<thead>
<tr>
<th>In a family...who should earn money?</th>
<th>natives</th>
<th>1st gen. immigrants</th>
<th>2nd gen. immigrants</th>
<th>3rd gen. immigrants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mostly the man (‘conservative’)</td>
<td>31.76</td>
<td>40.45</td>
<td>38.93</td>
<td>30.02</td>
</tr>
<tr>
<td>Mostly the woman</td>
<td>1.01</td>
<td>0.98</td>
<td>1.33</td>
<td>0.81</td>
</tr>
<tr>
<td>Both about the same</td>
<td>67.23</td>
<td>58.57</td>
<td>59.74</td>
<td>69.18</td>
</tr>
</tbody>
</table>

| ...who should clean the house?      |         |                     |                     |                     |
| Mosty the woman (‘conservative’)   | 34.62   | 37.81               | 39.82               | 29.92               |
| Mostly the man                      | 0.74    | 1.69                | 0.98                | 0.77                |
| Both about the same                 | 64.64   | 60.49               | 59.20               | 69.32               |

| [Survey country members] should be open to immigrants’ customs and traditions |         |                     |                     |                     |
| Strongly agree                      | 13.46   | 31.54               | 33.68               | 17.07               |
| Agree                               | 37.21   | 40.95               | 40.01               | 40.52               |
| Neither agree nor disagree          | 31.38   | 21.72               | 19.40               | 28.85               |
| Disagree                            | 11.68   | 3.41                | 4.53                | 8.54                |
| Strongly disagree                   | 6.26    | 2.38                | 2.38                | 5.02                |

| Immigrants should do all they can to keep their customs and traditions |         |                     |                     |                     |
| Strongly agree                     | 5.62    | 24.48               | 24.95               | 6.66                |
| Agree                              | 21.78   | 35.12               | 33.09               | 24.96               |
| Neither agree nor disagree         | 45.11   | 31.46               | 31.55               | 46.05               |
| Disagree                           | 18.65   | 6.30                | 7.48                | 15.25               |
| Strongly disagree                  | 8.84    | 2.63                | 2.93                | 7.08                |

| How strongly do you feel [survey country member]? |         |                     |                     |                     |
| Very strongly                       | 71.40   | 14.23               | 24.49               | 55.70               |
| Fairly strongly                     | 24.70   | 44.82               | 48.18               | 35.60               |
| Not very strongly                   | 3.05    | 25.75               | 18.66               | 7.12                |
| Not at all strongly                 | 0.85    | 15.20               | 8.67                | 1.57                |

The answers are summarised in Table 3.2. Normative beliefs of third-generation immigrants roughly match those of natives, and the figures for first-generation immigrants are similar to those for second-generation immigrants. This similarity should not be understood as something that necessarily occurs as the time spent by non-native families in a foreign
country increases, though. Our data refer to the offspring of individuals who migrated several decades, ago and reflect a particular social context. Under different social and economic conditions (e.g., in the presence of substantially higher migration flows and more ethnically diverse populations), the same outcomes may not occur. Table 3.3 reports inter-variable correlations, which always remain within acceptable levels and ease concerns about collinearity.

Table 3.3. Inter-variable correlations

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Native</th>
<th>1st gen. immigrant</th>
<th>Catholic</th>
<th>Protestant</th>
<th>Muslim</th>
<th>Atheist/agnostic</th>
<th>Parent with univ. degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Integration: weak</td>
<td>.01*</td>
<td>-.30*</td>
<td>.23*</td>
<td>-.02*</td>
<td>-.06*</td>
<td>.27*</td>
<td>-.15*</td>
<td>-.02*</td>
</tr>
<tr>
<td>Money: conservative</td>
<td>.21*</td>
<td>-.06*</td>
<td>.04*</td>
<td>.03*</td>
<td>.04*</td>
<td>.14*</td>
<td>-.07*</td>
<td>-.11*</td>
</tr>
<tr>
<td>Clean: conservative</td>
<td>.16*</td>
<td>-.02*</td>
<td>.01</td>
<td>.04*</td>
<td>.03*</td>
<td>.10*</td>
<td>-.03*</td>
<td>-.13*</td>
</tr>
<tr>
<td>Natives should be open to immigrants' customs</td>
<td>-.02*</td>
<td>-.19*</td>
<td>.08*</td>
<td>-.04*</td>
<td>-.05*</td>
<td>.22*</td>
<td>-.13*</td>
<td>.08*</td>
</tr>
<tr>
<td>Immigrants should do all to keep their customs</td>
<td>.04*</td>
<td>-.22*</td>
<td>.11*</td>
<td>-.04*</td>
<td>-.07*</td>
<td>.31*</td>
<td>-.15*</td>
<td>.01</td>
</tr>
<tr>
<td>Grade: mathematics</td>
<td>.03*</td>
<td>.05*</td>
<td>-.02*</td>
<td>.05*</td>
<td>.05*</td>
<td>-.06*</td>
<td>.00</td>
<td>.06*</td>
</tr>
<tr>
<td>Grade: SC language</td>
<td>-.17*</td>
<td>.04*</td>
<td>-.06*</td>
<td>.05*</td>
<td>.06*</td>
<td>-.06*</td>
<td>-.02*</td>
<td>.05*</td>
</tr>
</tbody>
</table>

*: significant at 5 percent level or better.

From the perspective of network formation, the most relevant feature of the dataset is the availability of information on within-class friendships. Students were asked to identify their five best friends in the class. An undirected friendship link exists whenever a nomination by Ego is reciprocated by Alter. The ceiling imposed on the number of nominations represents a possible limitation of the data, since it may prevent the identification of one’s entire reference group. About 6 and 3 percent of students were linked to five best friends in wave 1 and wave 2, respectively, while 38 and 31 percent sent five nominations.

3.2.1 Descriptive Statistics

Information on directed and undirected links are summarised in Table 3.4. Male students tended to nominate more classmates as best friends, but the number of undirected links does not significantly differ by gender. On average, the number of links for natives slightly exceeds that for pupils with a migration background.

About 88 percent of friendship nominations for both males and females were sent to students of the same gender, and figures grow up to 92 percent when considering undirected links. Calculations do not take into account mixing opportunities (e.g., single-sex or highly unbalanced classes), but nevertheless suggest the presence of strong homophilic preferences over
gender. Similar, although less pronounced, patterns can be observed with respect to religion and generational status. Roughly 60 percent of nominations sent by Christian pupils were directed to other Christians, whereas same-creed nominations of Muslims and atheists amounted to 54 and 56 percent, respectively. More than 75 percent of nominations made by natives and third-generation immigrants were directed to other students with little or no migration background, whereas first- and second-generation immigrants nominated other first- and second-generation immigrants 61 percent of the time. Similarities can be observed from the perspectives of cultural traits and school achievements as well. As shown in Table 3.5, the percentages of dyads that share potentially endogenous characteristics are always higher for friends than for non-friends. Tests of proportions reveal that, even when small, the differences are statistically significant.

Table 3.4. Friendship nominations and links

<table>
<thead>
<tr>
<th></th>
<th>wave 1 (avg±s.d.)</th>
<th>wave 2 (avg±s.d.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ego-to-Alter nominations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>overall</td>
<td>3.35±1.59</td>
<td>2.44±1.77</td>
</tr>
<tr>
<td>males</td>
<td>3.49±1.58</td>
<td>2.51±1.83</td>
</tr>
<tr>
<td>natives</td>
<td>3.43±1.55</td>
<td>2.52±1.74</td>
</tr>
<tr>
<td>1st gen. immigrants</td>
<td>3.10±1.68</td>
<td>2.25±1.83</td>
</tr>
<tr>
<td>Undirected links</td>
<td></td>
<td></td>
</tr>
<tr>
<td>overall</td>
<td>2.23±1.46</td>
<td>1.63±1.43</td>
</tr>
<tr>
<td>males</td>
<td>2.22±1.48</td>
<td>1.60±1.45</td>
</tr>
<tr>
<td>natives</td>
<td>2.30±1.45</td>
<td>1.73±1.43</td>
</tr>
<tr>
<td>1st gen. immigrants</td>
<td>1.96±1.47</td>
<td>1.43±1.41</td>
</tr>
<tr>
<td>reciprocity index</td>
<td>.65±.14</td>
<td>.67±.15</td>
</tr>
</tbody>
</table>

3.3 Determinants of link formation

We now turn to an analysis of the factors affecting the establishment of friendship links. We begin by considering only those covariates that are treated as exogenous throughout the chapter.

3.3.1 Gender, ethnicity, and religion

Table 3.6 reports maximum likelihood estimates for the whole set of observations and for selected subsamples. In column 1, the probability of a link is regressed against nationality alone. Even including the shares of
Table 3.5. % of dyads with similar characteristics and tests of proportions

<table>
<thead>
<tr>
<th></th>
<th>$Y_{ij} = 0$ (not friends)</th>
<th>$Y_{ij} = 1$ (friends)</th>
<th>Difference significant*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subj. integration: both strong</td>
<td>28.78</td>
<td>33.03</td>
<td>✓</td>
</tr>
<tr>
<td>Subj. integration: both weak</td>
<td>3.13</td>
<td>4.34</td>
<td>✓</td>
</tr>
<tr>
<td>Men should be the breadwinners</td>
<td>13.04</td>
<td>15.18</td>
<td>✓</td>
</tr>
<tr>
<td>Women should clean the house</td>
<td>14.22</td>
<td>16.44</td>
<td>✓</td>
</tr>
<tr>
<td>Natives: open to others’ customs</td>
<td>5.38</td>
<td>6.04</td>
<td>✓</td>
</tr>
<tr>
<td>Immigrants: keep own customs</td>
<td>2.40</td>
<td>3.14</td>
<td>✓</td>
</tr>
<tr>
<td>Math grade higher than median</td>
<td>12.94</td>
<td>14.21</td>
<td>✓</td>
</tr>
<tr>
<td>Math grade lower than median</td>
<td>12.48</td>
<td>14.01</td>
<td>✓</td>
</tr>
<tr>
<td>SCL grade higher than median</td>
<td>10.29</td>
<td>12.06</td>
<td>✓</td>
</tr>
<tr>
<td>SCL grade lower than median</td>
<td>10.19</td>
<td>11.03</td>
<td>✓</td>
</tr>
</tbody>
</table>

*: 5 percent level or better.

students in class that match Ego and Alter in terms of nationalities, the explanatory power of the model is limited. The estimates shown in column 2 are obtained by introducing covariates related to generational status, religion, age, and socio-economic characteristics as well as the full set of controls for confounding factors (described below). The improvement in terms of goodness-of-fit is considerable, and several interesting points can be made. First, the coefficient on nationality remains highly significant but drops by more than fifty percent. This indicates that the single-variable model estimated in column 1 picks up the effects of all variables correlated with ethnicity. Second, gender proves to be by far the most relevant predictor of friendship — a well-established result in the literature on relations among adolescents, which confirms the observation made in the previous section. Third, generational status appears to have independent predictive power for first- and second-generation immigrants, but not for third-generation ones. The interaction term between nationality and generational status is also found to be nonsignificant. Fourth, there is significant evidence of homophily with respect to social status, but the coefficients are smaller than those for nationality and religion. These findings carry over when restricting the analysis to dyads in which Ego is a third-generation immigrant or has no migration background but differ for dyads in which Ego is a first-generation immigrant. The estimates are reported in columns 3 and 4 of Table 3.6, respectively. The coefficient on nationality is greater for first-generation immigrants, who, on the other hand, show no evidence of homophily with respect to socio-economic conditions.
Table 3.6. Benchmark estimates

<table>
<thead>
<tr>
<th>Dyad characteristics</th>
<th>All dyads</th>
<th>All dyads</th>
<th>Natives and 3rd gen. immigrants</th>
<th>1st gen. immigrants</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Same nationality</td>
<td>.054***</td>
<td>.026***</td>
<td>.017***</td>
<td>.035***</td>
</tr>
<tr>
<td></td>
<td>(.004)</td>
<td>(.003)</td>
<td>(.005)</td>
<td>(.014)</td>
</tr>
<tr>
<td>Same gender</td>
<td>.130***</td>
<td>.129***</td>
<td>.164***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.003)</td>
<td>(.004)</td>
<td>(.004)</td>
<td>(.013)</td>
</tr>
<tr>
<td>Both 1st gen. immigrants</td>
<td>.019**</td>
<td>.026**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.009)</td>
<td>(.012)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Both 2nd gen. immigrants</td>
<td>.013***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.004)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Both 3rd gen. immigrants</td>
<td>-.002</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.008)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Same nationality × both 1st gen. immigrants</td>
<td>.023</td>
<td>.018</td>
<td>.024</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.016)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Same religion</td>
<td>.018***</td>
<td>.013***</td>
<td>.022**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.002)</td>
<td>(.003)</td>
<td>(.009)</td>
<td></td>
</tr>
<tr>
<td>Both have a parent with a university degree</td>
<td>.005**</td>
<td>.007**</td>
<td>.004</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.002)</td>
<td>(.003)</td>
<td>(.009)</td>
<td></td>
</tr>
<tr>
<td>Difference: number of books</td>
<td>-.004***</td>
<td>-.004***</td>
<td>.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.001)</td>
<td>(.001)</td>
<td>(.005)</td>
<td></td>
</tr>
<tr>
<td>Difference: number of rooms</td>
<td>-.001**</td>
<td>-.001</td>
<td>.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.001)</td>
<td>(.001)</td>
<td>(.002)</td>
<td></td>
</tr>
<tr>
<td>Difference: age</td>
<td>-.005***</td>
<td>-.005**</td>
<td>-.003</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.002)</td>
<td>(.002)</td>
<td>(.006)</td>
<td></td>
</tr>
<tr>
<td>Class fixed effects</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Shares of students with same characteristics</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Number of common friends</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Five minutes distance</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Socio-economic characteristics jointly nonsignificant (p-value)</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>.951</td>
<td></td>
</tr>
<tr>
<td>N (dyads)</td>
<td>108,193</td>
<td>108,193</td>
<td>70,367</td>
<td>6,960</td>
</tr>
<tr>
<td>Pseudo-$R^2$</td>
<td>.005</td>
<td>.376</td>
<td>.377</td>
<td>.427</td>
</tr>
</tbody>
</table>

One, two, and three asterisks indicate significance at the 10, 5, and 1 percent level, respectively.
Estimates were obtained while controlling for mixing opportunities, cliquishness, and unobserved heterogeneity at the class level. Additionally, we controlled for geographical proximity by checking whether or not Ego and Alter live within a five-minute walking distance. The number of observations for each specification is \( \sum_{l=1}^{L} (\kappa_l (\kappa_l - 1))/2 \), with \( L \) and \( \kappa_l \) being the number of classes and the number of children in class \( l \) who filled in the relevant sections of the questionnaire, respectively. The size of the coefficients is in line with other studies on bilateral network formation (e.g., Comola and Fafchamps 2013; Comola and Mendola 2015).

Overall, the results indicate a high degree of segregation and point to a plausible conclusion: recently migrated individuals bond mostly on the basis of ethnicity (possibly because of a need for communication) and religion, whereas those who have lived longer in a country exhibit homophilic preferences over a broader set of dimensions.

### 3.3.2 Cultural traits and school achievements

We next turn to cultural traits. Table 3.7 shows that the endogeneity test rejects the null hypothesis of exogeneity for all regressors (we chose a conservative significance level of 20 percent). The predictive power of normative beliefs and subjective integration is therefore estimated using IV-GMM techniques (Baum et al. 2003, 2007). First stage \( F \)-statistics for all specifications exceed the rule-of-thumb level of 10, which indicates that weakness of instruments is not a concern (Staiger and Stock 1997). The null hypothesis of the clustering-robust underidentification test (Kleibergen and Paap 2006), which checks whether excluded instruments are correlated with the endogenous regressors, is always rejected, meaning that equations are identified and instruments are relevant. The overidentification of models, moreover, allows us to test for the validity of instruments via the Sargan-Hansen J test of overidentifying restrictions. This is most important for our purposes because it permits us to assess whether excluded instruments are orthogonal to residuals from the estimated equations. The null of the test is not rejected for all specifications, which suggests that instruments are valid, i.e., uncorrelated with the error terms.

Estimated coefficients on endogenous variables are based on first-stage predictions, and standard errors are larger than those for other covariates. This makes it difficult to assess the relative predictive power of instrumented regressors with precision. The coefficients and significances, nevertheless, suggest that culture plays an important role in shaping friendship patterns and possibly has as much influence as ethnicity and
Table 3.7. Estimates (cultural traits)

<table>
<thead>
<tr>
<th>Other covariates included</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>in benchmark specification</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Same nationality</td>
<td>.017*** (.003)</td>
<td>.017*** (.003)</td>
<td>.018*** (.003)</td>
<td>.018*** (.003)</td>
<td>.017*** (.003)</td>
<td>.017*** (.003)</td>
</tr>
<tr>
<td>Same religion</td>
<td>.013*** (.002)</td>
<td>.012*** (.002)</td>
<td>.013*** (.002)</td>
<td>.013*** (.002)</td>
<td>.012*** (.002)</td>
<td>.012*** (.002)</td>
</tr>
<tr>
<td>Subjective integration:</td>
<td>.005 (.006)</td>
<td>.028* (.014)</td>
<td>.019*** (.007)</td>
<td>.015** (.006)</td>
<td>.035** (.014)</td>
<td>.040** (.021)</td>
</tr>
<tr>
<td>both strong§</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>both weak§</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Both think men should</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>be the breadwinners§</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Both think women should</td>
<td>.015** (.006)</td>
<td>.019*** (.007)</td>
<td>.015** (.006)</td>
<td>.035** (.014)</td>
<td>.040** (.021)</td>
<td>.040** (.021)</td>
</tr>
<tr>
<td>clean the house§</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Both think immigrants</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>should keep their customs§</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Both think natives should</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>open to immigrants’ customs§</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All controls for confoundings</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>First stage $F$-statistic</td>
<td>393.63</td>
<td>149.47</td>
<td>592.695</td>
<td>677.929</td>
<td>257.900</td>
<td>117.409</td>
</tr>
<tr>
<td>Sargan-Hansen J test (p-value)</td>
<td>.825</td>
<td>.952</td>
<td>.732</td>
<td>.799</td>
<td>.497</td>
<td>.331</td>
</tr>
<tr>
<td>Underidentification test (p-value)</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Endogeneity test (p-value)</td>
<td>.071</td>
<td>.109</td>
<td>.028</td>
<td>.104</td>
<td>.015</td>
<td>.040</td>
</tr>
<tr>
<td>N (dyads)</td>
<td>72,937</td>
<td>72,937</td>
<td>73,253</td>
<td>73,118</td>
<td>72,304</td>
<td>72,492</td>
</tr>
</tbody>
</table>

One, two, and three asterisks indicate significance at the 10, 5, and 1 percent level, respectively.

§: instrumented.

religion. This is especially true for attitudes towards immigrant cultures. A weak bond with the country Ego and Alter reside in was also found to matter, whereas strong subjective integration alone appears not to. The latter result lends support to the hypothesis of causation running also the other way around, that is, that friendship relations strengthen one’s subjective integration.

Contrary to what was found for cultural traits, the null hypothesis of the endogeneity test is never rejected when considering school achievements. The coefficients reported in Table 3.8 are therefore computed by standard maximum likelihood methods. Note that this does not necessarily rule out the possibility of peer effects in education, which are generally intended as influence-related spillovers from group average outcomes and not just from Alter to Ego. Estimates reveal that proficiency in both mathematics
and the country’s language carry significant predictive power. In contrast, 
the evidence is mixed for grades below the class median. Some caution, 
however, is warranted in interpreting coefficients, since they do not nec-
essarily reflect a preference for interacting with peers who are similar in 
terms of abilities. Instead, what matters could be affinity in terms of the 
unobservable determinants of grades, such as effort.

Table 3.8. Estimates (school achievements)

<table>
<thead>
<tr>
<th>Logit estimates - Marginal effects, clustered SEs in parentheses</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) (2) (3) (4) (5) (6)</td>
</tr>
<tr>
<td>Other covariates included in benchmark model</td>
</tr>
<tr>
<td>Same nationality</td>
</tr>
<tr>
<td>Same religion</td>
</tr>
<tr>
<td>Math grade: both higher than class median</td>
</tr>
<tr>
<td>Math grade: both lower than class median</td>
</tr>
<tr>
<td>SCL grade: both higher than class median</td>
</tr>
<tr>
<td>SCL grade: both lower than class median</td>
</tr>
<tr>
<td>All controls for confoundings</td>
</tr>
<tr>
<td>Endogeneity test (p-value)</td>
</tr>
<tr>
<td>Grades jointly nonsignificant (p-value)</td>
</tr>
<tr>
<td>N (dyads)</td>
</tr>
</tbody>
</table>

One, two, and three asterisks indicate significance at the 10, 5, and 1 percent levels, respectively. 
EN and SW: wave 2 only.

Finally, one can verify that no such relations exist for dyads where 
Ego is a first-generation immigrant. Table 3.10 reveals that, even without 
examining the direction of causation through instrumental variables 
methods, no significant correlation exists between cultural traits, academic 
achievements, and friendship links. The only exceptions, both reasonable, 
amer in terms of the unobservable determinants of grades, such as effort.

3.3.3 A consistency check

The CILS4EU dataset also provides information on the pupils’ parents, 
some of whom were interviewed during the first wave to assess cultural
assimilation. Ideally, this could allow one to use parental cultural traits as additional instruments based on the assumption that Ego’s (Alter’s) opinions do not influence those of Alter’s (Ego’s) parents. The data on parents, however, have several limitations. First, as already mentioned, they were collected only in wave 1 and only for approximately fifty percent of students. Second, and more importantly, identification of respondents was not random but voluntary. This leads to self-selection problems related to major differences between pupils and parents. The sample population, for example, is very over-representative of mothers (79 percent) and natives (81 percent).

When considering the strength of the bond with one’s country of residence, however, the distribution of answers for parents is similar to that for the students, and weakness of instruments is of no concern. Estimates for wave 1 data are reported in Table 3.9 and confirm the results that were previously obtained. The coefficient on weak integration remains significant at the 10 percent level, whereas that on strong integration continues to not be significant. First stage $F$-statistics and the outcomes of the underidentification and Sargan-Hansen tests indicate that the instruments are relevant and orthogonal to the error terms.

<table>
<thead>
<tr>
<th>Table 3.9. Consistency check</th>
</tr>
</thead>
<tbody>
<tr>
<td>IV-GMM second stage estimates · Clustered SEs in parentheses</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Other covariates included</td>
</tr>
<tr>
<td>in benchmark specification</td>
</tr>
<tr>
<td>Same nationality</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Same religion</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Subjective integration:</td>
</tr>
<tr>
<td>both strong§</td>
</tr>
<tr>
<td>Subjective integration:</td>
</tr>
<tr>
<td>both weak§</td>
</tr>
<tr>
<td>All controls for confoundings</td>
</tr>
<tr>
<td>First stage $F$-statistic</td>
</tr>
<tr>
<td>Sargan-Hansen J test (p-value)</td>
</tr>
<tr>
<td>Underidentification test (p-value)</td>
</tr>
<tr>
<td>N (dyads)</td>
</tr>
</tbody>
</table>

One, two, and three asterisks indicate significance at the 10, 5, and 1 percent level, respectively.

§: instrumented.
Table 3.10. Estimates (cultural traits and school achievements, Ego is a first-generation immigrant)

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)*</th>
<th>(8)*</th>
<th>(9)*</th>
<th>(10)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logit estimates · Marginal effects, clustered SEs in parentheses</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other covariates included in benchmark specification</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Same nationality</td>
<td>.037***</td>
<td>.036**</td>
<td>.036**</td>
<td>.035**</td>
<td>.033**</td>
<td>.035**</td>
<td>.037**</td>
<td>.038**</td>
<td>.037**</td>
<td>.037***</td>
</tr>
<tr>
<td></td>
<td>(.014)</td>
<td>(.014)</td>
<td>(.014)</td>
<td>(.015)</td>
<td>(.015)</td>
<td>(.019)</td>
<td>(.019)</td>
<td>(.018)</td>
<td>(.018)</td>
<td></td>
</tr>
<tr>
<td>Same religion</td>
<td>.020***</td>
<td>.020**</td>
<td>.021**</td>
<td>.021**</td>
<td>.020**</td>
<td>.021**</td>
<td>.028**</td>
<td>.027***</td>
<td>.027***</td>
<td>.028***</td>
</tr>
<tr>
<td></td>
<td>(.009)</td>
<td>(.009)</td>
<td>(.009)</td>
<td>(.009)</td>
<td>(.009)</td>
<td>(.013)</td>
<td>(.012)</td>
<td>(.012)</td>
<td>(.013)</td>
<td></td>
</tr>
<tr>
<td>Subjective integration: both strong</td>
<td>.041**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.016)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subjective integration: both weak</td>
<td></td>
<td>.013</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.012)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Both think men should be the breadwinners</td>
<td></td>
<td></td>
<td>- .008</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(.009)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Both think women should clean the house</td>
<td></td>
<td></td>
<td>.008</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(.011)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Both think immigrants should keep their customs</td>
<td></td>
<td></td>
<td>.005</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(.013)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Both think natives should be open to immigrants' customs</td>
<td></td>
<td></td>
<td>.003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(.016)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Math grade: both higher than class median</td>
<td></td>
<td></td>
<td>.018</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(.016)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Math grade: both lower than class median</td>
<td></td>
<td></td>
<td>- .006</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(.016)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCL grade: both higher than class median</td>
<td></td>
<td></td>
<td>.029**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(.015)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCL grade: both lower than class median</td>
<td></td>
<td></td>
<td>0.024</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(.016)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All controls for confoundings</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>N (dyads)</td>
<td>6,858</td>
<td>6,858</td>
<td>6,861</td>
<td>6,860</td>
<td>6,758</td>
<td>6,777</td>
<td>4,072</td>
<td>4,072</td>
<td>4,143</td>
<td>4,143</td>
</tr>
</tbody>
</table>

One, two, and three asterisks indicate significance at the 10, 5, and 1 percent levels, respectively.

*: EN and SW: wave 2 only.
3.4 Concluding remarks

Among the ways to assess social cohesion in a group is to ascertain the extent of homophilic interactions among its members. In pursuing this line of inquiry, we made a case for studying homophily from a multidimensional perspective. This chapter tested two different ideas. The first is that bonds among individuals grow not exclusively out of a common ethnic background, but from a rich array of individual traits that include gender, religion, socio-economic status, and cognitive abilities. The second is that characteristics such as religious creed, normative beliefs, and other cultural attributes should not be conflated within the category of ethnicity in the context of peer selection. We also examined whether or not the relevance of each attribute varies by generational status.

We began by introducing a stylised model that views social relations as the product of group composition and homophilic preferences over multiple attributes. We then tested our hypotheses while controlling for confounding factors, unobservable characteristics at the class level, and potential endogeneity. Consistent with earlier studies (Block and Grund 2014; Rapallini and Rustichini 2016; S. Smith et al. 2014), gender emerged as the most relevant predictor of friendship, and we also found evidence of preferences for similarity with respect to religion, socio-economic conditions, and academic achievements, especially when considering the upper half of the grade distribution. The size and significance of coefficients on culture-related attributes — religion, normative beliefs, and subjective integration — indicate that they all play distinct and non-negligible roles. Controlling for generational status, we finally verified that homophily varies as a function of the time spent in a country. Recently migrated individuals bond mostly on the basis of ethnicity and religion, whereas natives and third-generation immigrants exhibit homophilic preferences over a wider set of characteristics. As we already stressed, similarity between third-generation immigrants and natives does not represent the necessary result of the time spent by non-native families in a country, and these similarities should not be used to draw inferences about integration processes that do not take into account the social context in which they take place.

Overall, the picture that emerges is one of deep segregation among groups. These results confirm the findings from the literature on the strength of cross-cultural barriers and on how these barriers change as individuals grow up. Ethnicity, religion, socio-economic status, conservative views on gender roles, and cultural closeness are all significant
predictors of friendship links. Our results add to the literature indicating that barriers tend to emerge and interactions with dissimilar individuals to decline as children move into adolescence: ‘friendships become more exclusive, the importance of similarity as a basis for friendship becomes more pronounced [...] and similar ascribed and achieved characteristics are assigned even more weight in the selection of friends’ (Hallinan and Teixeira 1987, p. 556).

Understanding which characteristics matter the most in establishing social relationships and whether the effects of these attributes vary with age and with the time spent in the host country is key to informed and well-designed immigration policies. Our results first suggest that the emphasis placed on ethnicity, which is typical of the multicultural approach, may be misguided, since other individual traits seem to matter as much as ethnicity in establishing friendship relations. Second, our analysis confirms the need for increased efforts in improving inclusiveness and developing integration programs. Such programs would likely benefit from being tailored to the needs of target audiences. Among first-generation immigrants, for example, promotion of religious diversity and removal of language barriers that hamper communication have obvious importance. Anti-segregation programs aimed at natives and immigrants of second or third generation pose different challenges, i.e., tackling divides related to socio-economic conditions and other facets of cultural diversity.

Our work does not come without limitations. First, the availability of only two waves of data, which were collected very close together in time, did not allow us to investigate the co-evolution of similarity and fellowship. Second, despite the reassuring outcomes of tests checking the validity of instruments, the instrumentation method that we proposed has no general validity. As we have maintained, the average influence exerted by Ego on her indirect friends through Alter is likely to have a limited extent. A second effect, however, might also take place, consisting in Alter influencing both Ego and Ego’s indirect friends in a similar way. The fact that instruments are constructed by averaging over the characteristics of multiple indirect friends is likely to smooth this effect out, and our results seem to confirm this. The possibility of instruments capturing the same influence that affects endogenous regressors, however, cannot be ruled out a priori.
A. **Appendices**

A.1 *Appendix to Chapter 2*

The following code is also available for download at:

[https://github.com/ncampigotto/pairwise_imitation](https://github.com/ncampigotto/pairwise_imitation).

A.1.1 MATLAB code

```matlab
N=20; % Population size
p_revision=1/2; % Probability of action revision
p_error=0.05 % Probability of mutation
N_rounds=50000; % Number of rounds
k=N-1; % Required for looping over N

% Players are repeatedly and randomly matched in pairs
% to play a Stag Hunt stage game
% Stag=1, Hare=0

% ACTIONS PAYOFFS
% P1   P2   P1   P2
% ------------------
% 1    1    10   10
% 1    0    0    3
% 0    1    3    0
% 0    0    3    3

% The evolution of play under Revision protocol 1 (PRI)
% is described by the array Action1
Action1 = [zeros(1,N/2) ones(1,N/2)]; % Initial state
Action1 = Action1(randperm(numel(Action1)));
% Choose one of the following:
% - Random initial condition: Action1=round(rand(1,N));
```
% - Everyone plays Hare at round 1: Action1 = zeros(1,N);
% - Everyone plays Stag at round 1: Action1 = ones(1,N);
% - 1/2 of the population plays Stag and
% 1/2 of the population plays Hare at round 1:
% Action1 = [zeros(1,N/2) ones(1,N/2)];
% Action1 = Action1(randperm(numel(Action1)));

round=1; % Set first round of play

% The agents’ payoffs under Revision protocol 1 (PRI)
% are described by the array Payoff1
% Remark: agent i (odd) plays against agent i+1 (even)
for i=1:2:N-1
  if(Action1(round,i)==1 && Action1(round,i+1)==1)
    Payoff1(round,i)=10;
    Payoff1(round,i+1)=10;
  end
  if(Action1(round,i)==0 && Action1(round,i+1)==0)
    Payoff1(round,i)=3;
    Payoff1(round,i+1)=3;
  end
  if(Action1(round,i)==0 && Action1(round,i+1)==1)
    Payoff1(round,i)=3;
    Payoff1(round,i+1)=0;
  end
  if(Action1(round,i)==1 && Action1(round,i+1)==0)
    Payoff1(round,i)=0;
    Payoff1(round,i+1)=3;
  end
end

% The evolution of play and the agents’ payoffs under
% Revision protocol 2 (PII) are described by the arrays
% Action2 and Payoff2, respectively
Action2=Action1;
Payoff2=Payoff1;

%%%%%%%%%%
%% LOOP %
%%%%%%%%%%
for round=2:N_rounds
  % Initialise play:
  Action1(round,:)=Action1(round-1,:);
  Action2(round,:)=Action2(round-1,:);
  Payoff1(round,:)=Payoff1(round-1,:);
  Payoff2(round,:)=Payoff2(round-1,:);
end
% Every player has a probability $p_{\text{revision}}$ of being
% selected for action revision.
revision1 = rand(1,N) < $p_{\text{revision}}$;
revision2 = rand(1,N) < $p_{\text{revision}}$;

% In every round, Player $i$ plays Action(round,$i$) and
% receives a payoff of Payoff(round,$i$).

%%%%%%% REVISION PROTOCOL 1:
%%%%%%% PAIRWISE RANDOM IMITATION (PRI)

for $i$=1:N

  % Check whether agent $i$ has received a revision
  % opportunity
  if revision1($i$)==true % If yes,...
    other_players=[1:$i$-1,$i$+1:N]; % ...consider $P\{i\}$...
    reference=other_players(ceil(rand()*(k))); % ...and draw
      % one reference at random from $P\{i\}$.
    % Check whether the payoff earned by $i$’s reference
    % in the previous round is higher than the payoff
    % earned by $i$. If yes, then $i$ copies their
    % reference.
    if Payoff1(round-1,reference)>Payoff1(round-1,$i$)
      Action1(round,$i$)=Action1(round-1,reference);
    end
    % If agent $i$ makes a mistake, they selects the
    % action that is NOT prescribed by the revision
    % protocol.
    if rand()<$p_{\text{error}}$
      Action1(round,$i$)=abs(Action1(round,$i$)-1);
    % Action1(round,$i$)= randi([0,1]);
    % select if mistakes consist in choosing
    % one action at random
    end
  end

% Randomly shuffle Action1/Payoff1 (new pairs are formed
% at random)
lastPlayPayoff1=[Action1(end,:);Payoff1(end,:)];
shuffled_lastPlayPayoff1 = lastPlayPayoff1(:,randperm(size(
lastPlayPayoff1,2)));
Action1(end,:)=shuffled_lastPlayPayoff1(end-1,:);
Payoff1(end,:)=shuffled_lastPlayPayoff1(end,:);
clear lastPlayPayoff1 shuffled_lastPlayPayoff1;

% Compute payoffs
for $i$=1:2:N-1
  if (Action1(round,$i$)==1 && Action1(round,$i$+1)==1)
    Payoff1(round,$i$)=10;
  end
end
Appendix A

Payoff1(round,i+1)=10;
if(Action1(round,i)==0 & Action1(round,i+1)==0)
    Payoff1(round,i)=3;
    Payoff1(round,i+1)=3;
end
if(Action1(round,i)==0 & Action1(round,i+1)==1)
    Payoff1(round,i)=3;
    Payoff1(round,i+1)=0;
end
if(Action1(round,i)==1 & Action1(round,i+1)==0)
    Payoff1(round,i)=0;
    Payoff1(round,i+1)=3;
end

%%%%%%% REVISION PROTOCOL 2:
%%%%%%% PAIRWISE INTERACTION-AND-IMITATION (PII)

for i=1:N
    % Check whether agent i has received revision opportunity
    if revision2(i)==true % If yes,...
        opponent=mod(i,2)*(2)+i-1; % ...consider agent i’s last opponent.
        if Payoff2(round-1,opponent)>Payoff2(round-1,i)
            Action2(round,i)=Action2(round-1,opponent);
        end
    end
    % When an agent makes a mistake, they selects the action that is NOT prescribed by the revision protocol
    if rand()<p_error
        Action2(round,i)=abs(Action2(round,i)-1);
        % select if mistakes consist in choosing one action at random
    end
end

% Randomly shuffle Action2/Payoff2 (new pairs are formed at random)
lastPlayPayoff2=[Action2(end,:);Payoff2(end,:)];
shuffled_lastPlayPayoff2 = lastPlayPayoff2(:,randperm(size(lastPlayPayoff2,2)));
Action2(end,:)=shuffled_lastPlayPayoff2(end-1,:);
Payoff2(end,:)=shuffled_lastPlayPayoff2(end,:);
clear lastPlayPayoff2 shuffled_lastPlayPayoff2;

% Compute payoffs
for i=1:2:N-1
    if(Action2(round,i)==1 && Action2(round,i+1)==1)
        Payoff2(round,i)=10;
        Payoff2(round,i+1)=10;
    end
    if(Action2(round,i)==0 && Action2(round,i+1)==0)
        Payoff2(round,i)=3;
        Payoff2(round,i+1)=3;
    end
    if(Action2(round,i)==0 && Action2(round,i+1)==1)
        Payoff2(round,i)=3;
        Payoff2(round,i+1)=0;
    end
    if(Action2(round,i)==1 && Action2(round,i+1)==0)
        Payoff2(round,i)=0;
        Payoff2(round,i+1)=3;
    end
end
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% EVOLUTION OF COOPERATION %%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% If Player i played Stag at time t under protocolX, we have:
% ActionX(i,t)=1
% So the proportions of players who played Stag at round t are:
%pl_Stag(t)=sum(Action1(t,:))/N;
%p2_Stag(t)=sum(Action2(t,:))/N;
for t=1:N_rounds
    pl_Stag(t)=sum(Action1(t,:))/N;
    p2_Stag(t)=sum(Action2(t,:))/N;
end

A.2 Appendix to Chapter 3

A.2.1 Questionnaire items

The CILS4EU questionnaire is structured in several sections. Our variables were constructed from information assessed in the sections named ‘Youth Main’ (YM), ‘Youth Classmates’ (YC), and ‘Youth Parents’ (YP).

YM-1 Are you a boy or a girl?
Appendix A

YM-2  When were you born?
YM-3  In which country were you born?
YM-19 Which grades did you get in the last school year in the following subjects?

(a) Math
(b) <Survey country language>

YM-27 Did your mother complete university? [Yes; No; Don’t know]
YM-33 Did your father complete university? [Yes; No; Don’t know]
YM-63 In a family, who should do the following?

(a) Earn money [Mostly the man; Mostly the woman; Both about the same]
(b) Clean the house [Mostly the man; Mostly the woman; Both about the same]

YM-66 How strongly do you feel <survey country member>? [Very strongly; Fairly strongly; Not very strongly; Not at all strongly]
YM-71 What is your religion? [No religion; Buddhism; Christianity; Christianity: Catholic; Christianity: Protestant; Hinduism; Islam; Judaism; Sikh; Other: specify]
YM-75 How much do you agree or disagree with each of these statements?

(a) The <survey country> people should be open to the customs and traditions of immigrants [Strongly agree; Agree; Neither agree nor disagree; Disagree; Strongly disagree]
(b) Immigrants should do all they can to keep their customs and traditions [Strongly agree; Agree; Neither agree nor disagree; Disagree; Strongly disagree]

YM-95 How many rooms are there in your home (not counting kitchen and bathroom)?
YM-96 About how many books are there in your home? [0-25; 26-100; 101-200; 201-500; More than 500]
YC-1 Who are your best friends in class? Here you may write down no more than five numbers.
YC-9  Which classmates live within a 5 minute walk from your home?
YP-9  How strongly do you feel <survey country member>? [Very strongly; Fairly strongly; Not very strongly; Not at all strongly]
A.2.2 Variables description

Dependent and controls for confoundings

- Binary: 1 if a friendship nomination from Ego to Alter is reciprocated, as assessed by question YC-1.
- Share of Ego’s (Alter’s) classmates similar to her in each characteristic considered.
- Binary: 1 if Ego and Alter live within a five-minute walking distance, as assessed by question YC-9.
- Binary: 1 if Ego and Alter are both members of a club, as assessed by question YM-76.
- Number of best friends that Ego and Alter share, as assessed by question YC-1.

Demographic characteristics

- Information on generational status was retrieved from previous work by Dollmann and Konstanze (2016).
- Other variables are constructed straightforwardly from questions YM-1, YM-2, YM-3, YM-71.

Socio-economic status

- Binary: 1 if at least one among both Ego’s and Alter’s parents completed tertiary education, as assessed by questions YM-27 and YM-33.
- Difference (absolute value) between the answers given by Ego and Alter to question YM-95 (size of the house in number of rooms).
- Difference (absolute value) between the answers given by Ego and Alter to question YM-96 (number of books in the house).

Normative beliefs

- Binary: 1 if both Ego and Alter answered ‘mostly the man’ to question YM-63a.
- Binary: 1 if both Ego and Alter answered ‘mostly the woman’ to question YM-63b.
- Binary: 1 if both Ego and Alter answered ‘strongly agree’ to question YM-75a.
- Binary: 1 if both Ego and Alter answered ‘strongly agree’ to question YM-75b.
Normative beliefs (instruments)

- Binary: 1 if both Ego (Alter) and the majority of Alter’s friends who are not friends of Ego (Ego’s friends who are not friends of Alter) answered ‘mostly the man’ to question YM-63a.
- Binary: 1 if both Ego (Alter) and the majority of Alter’s friends who are not friends of Ego (Ego’s friends who are not friends of Alter) answered ‘mostly the woman’ to question YM-63b.
- Binary: 1 if both Ego (Alter) and the majority of Alter’s friends who are not friends of Ego (Ego’s friends who are not friends of Alter) answered ‘strongly agree’ to question YM-75a.
- Binary: 1 if both Ego (Alter) and the majority of Alter’s friends who are not friends of Ego (Ego’s friends who are not friends of Alter) answered ‘strongly agree’ to question YM-75b.

Bond with one’s country

- Binary: 1 if both Ego and Alter answered ‘very strongly’ to question YM-66.
- Binary: 1 if both Ego and Alter answered ‘not very strongly’ or ‘not at all strongly’ to question YM-66.

Bond with one’s country (instruments)

- Binary: 1 if both Ego (Alter) and the majority of Alter’s friends who are not friends of Ego (Ego’s friends who are not friends of Alter) answered ‘very strongly’ to question YM-66.
- Binary: 1 if both Ego (Alter) and the majority of Alter’s friends who are not friends of Ego (Ego’s friends who are not friends of Alter) answered ‘not very strongly’ or ‘not at all strongly’ to question YM-66.
- Binary: 1 if both Ego (Alter) and Alter’s parent (Ego’s parent) answered ‘very strongly’ to questions YM-66 and YP-9.
- Binary: 1 if both Ego (Alter) and Alter’s parent (Ego’s parent) answered ‘not very strongly’ or ‘not at all strongly’ to questions YM-66 and YP-9.

School grades

- Binary: 1 if both Ego’s and Alter’s grades in math (as assessed by question YM-19a) are above the class median.
- Binary: 1 if both Ego’s and Alter’s grades in math (as assessed by question YM-19a) are below the class median.

- Binary: 1 if both Ego’s and Alter’s grades in <survey country language> (as assessed by question YM-19b) are above the class median.

- Binary: 1 if both Ego’s and Alter’s grades in <survey country language> (as assessed by question YM-19b) are below the class median.


Brandenburger, Adam (2014), The Language of Game Theory: Putting Epistemics into the Mathematics of Games, World Scientific, Hackensack (NJ) and London.


HENRICH, JOSEPH, ROBERT BOYD, SAMUEL BOWLES, COLIN F. CAMERER, ERNST FEHR, and HERBERT GINTIS (eds.) (2004), Foundations of


Kalter, Frank, Anthony F. Heath, Miles Hewstone, Jan O. Jonsson, Matthijs Kalmijn, Irena Kogan, and Frank van Tu-bergen (2016), *Children of Immigrants Longitudinal Survey in Four European Countries (CILS4EU) – Reduced version*, Reduced data file for download and off-site use. GESIS Data Archive, Cologne, ZA5656 Data file Version 2.3.0, doi:10.4232/cils4eu.5656. 2.3.0.


MATZ, DAVID (1964), Delinquency and Drift, John Wiley and Sons, New York.
Perry, Matthew C. (1856), Narrative of the Expedition to the China Seas and Japan, 1852-1854, ed. by Francis L. Hawks, Beverley Tucker, Senate Printer, Washington.
PROTO, EUGENIO, ALDO RUSTICINI, and ANDIS SOFIANOS (2017),
“Intelligence, Personality and Gains from Cooperation in Repeated
Interactions”, working paper.
nity in the Twenty-first Century. The 2006 Johan Skytte Prize Lecture”,
RAKOVE, JACK N. (1996), Original Meanings: Politics and Ideas in the
RAKOVE, JACK N., ANDREW R. RUTTEN, and BARRY R. WEINGASt
(2005), “Ideas, Interests, and Credible Commitments in the American
Revolution”, working paper.
RANKIN, FREDERICK W., JOHN B. VAN HUYCK, and RAYMOND C.
BATTALIO (2000), “Strategic Similarity and Emergent Conventions: Evi-
dence from Similar Stag Hunt Games”, Games and Economic Behavior,
vol. 32, no. 2, pp. 315-337.
RAPALLINI, CHIARA and ALDO RUSTICINI (2016), “Elective affinities
matter as much as ethnicity in multi-ethnic schools”, Journal of Eco-
REID, JOHN P. (1995), Constitutional History of the American Revolution,
University of Wisconsin Press, Madison (WI).
RENDELL, LUKE, ROBERT BOYD, DANIEL COWNDEN, MAGNUS EN-
QUIST, KIMMO ERIKSSON, MARCUS W. FELDMAN, LAUREL FOGARTY,
STEfANO GHIRLANDA, DAVID LLICCRAP, and KEVIN N. LALAND
Equilibrium Selection in Evolutionary Games with Random Matching”,
Contract and The First and Second Discourses, ed. by Susan Dunn,
SANDHOLm, WILLIAM H. (2010), “Pairwise Comparison Dynamics and
Evolutionary Foundations for Nash Equilibrium”, Games, vol. 1, no. 1,
pp. 3-17.
SANTos, FRANCISCO C., JORGE M. PACHECO, and TOM LENAERTS
(2006), “Cooperation Prevails When Individuals Adjust Their Social


