Survival of the fittest group: an empirical economic perspective on multilevel selection theory

Philippe M.L.P. Kamm
Jeroen Smits

Radboud University
September 2021

Abstract

Evolutionary biologists are divided on whether natural selection, specifically related to the evolution of altruistic behaviour, takes place only at the individual or also at the group level. This study builds on the understanding of multilevel selection theory from an empirical economic perspective. Data from 33,557 communities in 39 countries in Africa are used in OLS regression models with fixed effects dummies to examine the relationship between 20-year community development and initial community equality. It is found that increases in years of education and women's age of marriage are highest in communities where these outcomes are initially distributed equally. For education, this effect is strongest in communities with high child mortality rates. Furthermore, findings indicate that communities with high initial gender equality and low initial religious homogeneity show stronger increases in women's age of marriage and education respectively. Our evidence provides support for a positive relationship between fitness growth and altruism at the group level, as hypothesized in multilevel selection theory. Economic policies aimed at promoting cooperation at the community level in developing countries could be effective to increase levels of community development.

Table of Contents

1. Introduction

In mainstream economics, altruistic behaviour has classically been considered irrational. With the emergence of behavioural economics come explanations for often observed altruistic acts such as cooperation and reciprocity. First, bounded rationality arguments are given, stating that humans act irrationally as a result of emotions, biases, and heuristics. Second, rationality arguments are given based on investments in social capital and reputation, which can pay off in the long run (Aumann, 2019). Over the past decades, economists have not only learnt to integrate altruistic behaviour in their studies, but also more generally started to question how rational economic agents really are. But if altruistic behaviour is not rational, then how would it have ever survived in our gene pool? To answer this question, it is important to distinguish economic rationality from ecological rationality. Economic rationality revolves around a cost-benefit optimization of an individual actor's economic outcomes. Behaviour as a result of this optimization is implied to depend on these costs and benefits (Todd & Gigerenzer, 2012). Ecologically rational behaviour, on the other hand, has survived in the evolutionary selection process and often comes in the form of heuristics (Aumann, 2019; Smith, 2003; Todd & Gigerenzer, 2012). Having children, for instance, can be irrational from an economic perspective, yet it is obviously rational from an ecological perspective.

There are many similarities between economics and evolutionary biology (Schwesinger, 2013). Both fields of study aim to find logical explanations for why we observe certain behaviour and decisions. Moreover, competition is a central theme that is used in both fields to form the basis of many of their explanations. The idea of selection of firms in a competitive market and the term creative destruction by Schumpeter (1982) stem directly from the Darwinian idea of evolution. A large difference between the two domains, however, is that in economics it is not always clear if there is an economic explanation for a given observation or not. Recall the introduction on altruistic behaviour: it is still up to debate how much of our altruistic actions are driven by economic motivations. In evolutionary biology, however, this is different. Whatever behaviour we observe today is thought to be a consequence of natural selection, and therefore there exists a logical reason and environment which explains its survival (McAndrew, 2002; Smith, 2003). As such, there likely exists an ecologically rational explanation for the evolution of altruistic behaviour. Consensus on explaining altruistic behaviour in the field of economics has grown in recent years, but in the field of evolutionary biology this consensus currently sits at an historical low (Jeler, 2018).

Evolutionary biologists are divided on the levels at which natural selection take place. Since Charles Darwin's *The Origin of Species*, there is consensus that selection takes place at the level of individuals, the vehicles of genes. Supporters of multilevel selection theory, however, argue that selection takes place at more than just the individual level, most notably also at the group level. The general idea is that individual egoists beat individual altruists, but groups of altruists beat groups of egoists (Wilson & Wilson, 2007). After decades of debate, evolutionary biologists are still choosing sides and either support or reject multilevel selection theory. Some economists have picked up on this intellectual conflict, and attempted to provide additional insights using economic game theoretic models of evolution. Unfortunately, only little progress has been made so far (Safarzynska & van den Bergh, 2010). Moreover, empirical evidence on multilevel selection is largely absent. Economists have only showed meager interest in empirical studies on the topic, despite the potential policy implications such evidence could yield (Wilson & Gowdy, 2015). The debate on multilevel selection theory remains open and eagerly awaits empirical insights.

Given the ongoing debate on multilevel selection theory and the lack of empirical evidence from economists on the topic, this study aims to build on the understanding of multilevel selection theory from an empirical economic perspective. The relationship between group fitness growth and group altruism, using data from communities of households in rural Africa, will be examined to shed more light on group level selection in economics. Section 2 provides a detailed overview of the literature related to multilevel selection theory in both biology and economics, and concludes with the conceptual framework of this study. Methods, including the data, variables, and model, are discussed in Section 3. Results are presented in Section 4, and Section 5 concludes.

2. Literature review

It is important to develop an understanding of multilevel selection theory and its arguments for and against it from both a biological and economic perspective. Section 2.1 provides an historical overview of the debate in evolutionary biology and illustrates where the debate stands today. Economic studies on multilevel selection theory are discussed in Section 2.2, as well as potential implications for economic policy. Section 2.3 formulates the research problem, and the conceptual framework is presented in Section 2.4.

2.1. Multilevel selection theory in evolutionary biology

Today, there is a fierce academic debate between evolutionary biologists on the extent to which evolution, specifically the evolution of altruism and cooperation, takes place at the individual and at the group level. Charles Darwin (1859), the founding father of evolutionary biology, already recognized the difficulty of explaining altruistic behaviour in *The Origin of Species*. If the fittest genes survive, then how could eusocial species such as ants live in colonies with sterile worker ants? Ants themselves could not be the only unit of selection, as sterile worker ants are not able to reproduce and hence their genes are not able to survive in the gene pool. Darwin realized that not only individual ants, but also entire ant colonies led by their queens must be competitive. These thoughts led to the birth of multilevel selection theory: selection between individuals takes place via individually-advantageous traits, and selection between groups takes place via group-advantageous traits (Wilson & Wilson, 2007). Altruistic behaviour could be regarded as a group-advantageous trait, hence explaining its survival in the gene pool. Though multilevel selection theory existed in the early twentieth century, many evolutionary biologists were not convinced by it and stuck to Darwin's more elaborate theory on individual level selection.

The 1960s marked the important introduction of the concept of inclusive fitness, also known as kin selection, by Hamilton (1964). The basic idea is that for individuals it is not only beneficial to improve their own fitness, but also to improve the fitness of others which share their genes. For instance, an individual might share scarce food with their child or sibling to save their life. This is a sacrifice to their own individual fitness, but results in greater genetic fitness (McAndrew, 2002). This concept can be extended to small groups such as tribes with relatively high gene correlation, and thereby forms the basis of an explanation for altruistic behaviour in groups based on individual level selection (Birch, 2019). Hamilton's introduction of kin selection can be regarded as the missing puzzle piece to Darwin's individual level selection theory, as it could be used to explain behaviour of eusocial species such as ants. Kin selection theory still forms the basis of explaining altruistic behaviour for evolutionary biologists who are on the individual level selection side today, and with that the basis to reject the theory of multilevel selection (Leigh, 2010).

But not every scientist is on the individual level selections side, because despite the introduction of kin selection, multilevel selection theory is still alive. Most notably evolutionary biologist David Sloan Wilson persists to argue that only a limited range of altruistic behaviour can be explained by kin selection. He argues that the well-developed level of social, moral, and coordinative behaviour of species such as human beings is the result of altruistic groups outcompeting selfish groups. At the heat of the debate in 2010, Edward Osborne Wilson, one of the world's most influential biologists, worked together with two prominent mathematicians to reject the foundations of kin selection to explain eusocial behaviour (Nowak, Tarnita, & Wilson, 2010). They conclude that multilevel selection theory is needed to explain eusocial behaviour of species, including humans. This publication remains controversial to say the least (Abbot et al., 2011). Today, evolutionary biologists either support or completely reject multilevel selection theory.

Strong support for multilevel selection theory by biologists can be found in Wilson and Kniffin (1998), Wilson and Wilson (2007), Wilson (2012), Goodnight (2015), and Wilson (2018). Generally, their reasoning revolves around the idea that selfish individuals beat altruistic individuals within groups, but groups of altruistic individuals beat groups of selfish individuals. Perfect conditions for multilevel selection are low within-group variation and high between-group variation. Beyond multilevel selection theory at the genetic level, the theory also finds support at the cultural level. Akçay and van Cleve (2011), Tomasello (2014), Richerson *et al*. (2016), and Wilson (2018) all provide arguments that not only genes, but also group specific social structures such as culture are levels of selection. A third branch of literature rests on evolutionary game theoretic models, providing a more mathematical view on evolution. Within this branch, evidence for multilevel selection theory is presented in Nowak (2006), Traulsen and Nowak (2006), Nowak (2012) and Rand and Nowak (2013). Finally, Balliet, Wu, and De Dreu (2014) conduct a meta-analysis on ingroup favouritism and find support for a willingness of people to sacrifice personal benefits for the good of their own group. They conclude that this can have effects on intergroup competition. All in all, there are multiple roads leading to support for at least some form of multilevel selection theory.

Despite the extensive range of support, however, arguments to completely reject multilevel selection theory are easily found. Abbot *et al*. (2011), a group of over 100 scientists, come with a direct response to the aforementioned critical paper on kin selection by Nowak, Tarnita, and Wilson (2010). In this response, they state that the arguments used by Nowak *et al*. (2010) "are based upon a misunderstanding of evolutionary theory and a misrepresentation of the empirical literature". Similar highly critical responses come from Boomsma *et al*. (2011), Strassmann et al. (2011), Ferriere and Michod (2011) and Herre and Wcislo (2011). Moreover, Dawkins (2012), a leading authority in the field, refers to selection at the group level as a "poorly defined and incoherent view" of evolution. More recently, Gardner (2015) and Jeler (2018) have also expressed fierce criticism. Though a wide range of arguments is given, generally it is argued that any evolutionary outcome that can be reached through multilevel selection can also be reached through kin selection alone. Additionally, many researchers are not convinced that groups can evolve, as they cannot clearly reproduce or die like genes can. Even with the intellectual warfare going on, Van Veelen (2009), Kramer and Meunier (2016), Birch (2019), and Apicella and Silk (2019) are convinced that both kin and multilevel selection theories have their rights and wrongs. They conclude that a combination of both perspectives could increase understanding of group evolution.

2.2. Multilevel selection theory in economics

If anything, it should be clear by now that evolutionary biologists have not been able to provide a convincing answer on how altruism within groups came to be. Some economists have picked up on this, and, albeit from a slightly different angle, attempted to study the strength of evolutionary theories themselves. Mostly, economic game theoretic models are used to study multilevel selection mechanisms in economics. Seldom however, do we find empirical analyses on the topic. The following paragraphs provide a brief overview of the most relevant literature to date and its potential implications for economic policy.

By far the most significant contributions to the debate on multilevel selection theory from economists come from game theoretic models. It is important to note that economic game theoretic models are not identical to evolutionary/biological game theoretic models. Zinovyeva (2010) explains that economic agents are commonly able to learn and adapt behaviour of others swiftly, and can also develop institutions. Therefore, the role of the group is more important in economics and hence multilevel selection is more plausible. Economic agents, however, can also make decisions based on their own expectations of future behaviour of others, adding an extra layer of complexity to economic game theoretic models. Within the realm of economic game theoretic models on multilevel selection theory, public goods games and prisoner's dilemma games are most relevant. Eaton, Eswaran, and Oxoby (2011), Waring, Goff, and Smaldino (2017), and DeMartini and Marriott (2018) all find support for group level selection in public goods games. Generally speaking, in such public goods games agents live in a 2-dimensional plane and harvest resources which are needed for survival and reproduction. They interact with neighbouring agents and have genes or traits which influence their decision to cooperate/share or not. Results show that through multilevel selection, resource sharing traits evolve and cooperative groups survive whereas non-cooperative groups die off. Next, García and van den Bergh (2011) find that multilevel selection can explain parochial altruism behaviour, meaning altruism towards one's own social group, in a prisoner's dilemma game. Though support for multilevel selection theory is found in economic game theoretic models, Safarzynska and van den Bergh (2010) stress that there are currently many different methods and models to study the same phenomenon, and all are in their early stages. Richerson *et al*. (2016) highlight that it is difficult to develop an empirical game theory which captures multilevel selection due to the complexity of human interaction. Both conclude that much important work remains to be done.

On the empirical side, however, work has hardly been done at all. Schwesigner (2013) argues that senses of emotions and fairness are hardwired in our brains as a result of selection at the gene level, which has allowed for cultural and economic group selection over the past 10,000 years. The latter part of this finding is confirmed by Tomaszewski (2021), who conducted a review of literature on human cooperation and economic development from antiquity until today. Tomaszewski (2021) concludes that both cooperation and egoism have always affected economic development throughout history. These two studies provide some empirical support for a relevant role of group selection since the Neolithic Revolution.[1] Waring and Acheson (2018) find more specific empirical evidence for multilevel selection theory. They conduct a study on territoriality and conservation rules in Maine's lobstering industry (1700s–present), and find empirical evidence for selection at both the level of harbor gangs (composed of residents of a single harbor town) and at the level of individual lobstermen. The studies by Tomaszewski (2021) and Waring and Acheson (2018) are both long-term analyses of intergroup competition and are of qualitative nature. Apart from these two studies, not much empirical evidence exists on multilevel selection in economics. It is a gap in literature waiting to be filled, if the right data can be found.

If an empirical relationship between fitness growth and altruism at the group level is found, however, it could still be that the genes in all groups survive. Especially in today's modern age, some groups could theoretically live on in poverty and others in comfort, but both would be able to reproduce and their genes would thus survive. For selection of genes to take place, the weakest groups would actually have to die off (or at least reproduce at a lower rate), and the strongest

[1] Marks the transition of hunters and gatherers to agriculture and settlement lifestyles.

would have to flourish.[2] As such, differences between the fitness levels of groups would need to grow large such that only the fittest groups survive. This is theorized to take place under specific contexts, meaning that the strength of the relationship between group fitness growth and group altruism might be moderated by the contexts which are theorized to be optimal for group level selection. If the relationship between group fitness growth and group altruism is stronger when the group level selection context is better, there is more reason to believe that actual group selection would take place in an extreme (optimal) context.

Two moderating factors which might affect the strength of the relationship between group fitness growth and group altruism are heterogeneity between groups and high stakes of success. Waring *et al*. (2017) argue that the optimal context for group level selection is probably warfare, where the stakes of success are survival and the stakes of failure are death. Notice how this describes a context of high intergroup competition where the consequences of (a lack of) group altruism are enormous for group fitness growth. Heterogeneity between groups is related to intergroup competition, and thereby to the consequences of group altruism. The idea here is that if there is high heterogeneity between groups, these groups are more likely to engage in intergroup conflict and potentially fight each other to death. As such, the effect of group altruism on group fitness growth is expected to increase as heterogeneity between groups grows. Waring and Acheson (2018) substantiate the importance of heterogeneity between groups in their empirical study on the lobstering industry, and add a second possible context of high stakes of success. The context of high stakes of success also connects to numerous studies that regard resource management as an important driving factor of intergroup competition (Hodler, 2006; Hodler & Knight, 2012). As the stakes of success grow, group altruism becomes increasingly important for group fitness growth. Again however, empirical evidence on the possible interaction between optimal contexts for group selection and the relationship between group fitness growth and group altruism is currently limited.

Next to heterogeneity between groups, researchers have identified that group selection can occur when there is homogeneity within groups (DeMartini & Marriott, 2018). Eaton *et al*. (2011)

---

[2] On the other hand, Richerson *et al.* (2016) and Waring and Acheson (2018) highlight that cultural group selection could still take place when the genes in all groups survive, by imitation of social norms and behaviours that prove to be successful for the survival of a group. Essentially, the fittest culture survives and is adopted by all groups, whereas the weak cultures die off. Nonetheless, the group selection optimal contexts discussed in this section are relevant in cultural group selection too.

confirm that ethnically fractionalized groups are less likely to progress at the group level. It is important to understand, however, that homogeneity within groups is not a moderating factor in the relationship between group fitness growth and group altruism. Instead, homogeneity within groups tells us something about the likelihood that individual group members will give up direct personal benefits in favour of the group, and is thereby directly related to group altruism. Still, the role that homogeneity within groups plays in group altruism and thereby indirectly in group fitness growth is an important part of multilevel selection theory. It is therefore relevant to gather empirical economic evidence on the role that homogeneity within groups plays in group fitness growth.

Developing our understanding of multilevel selection in economics could have significant implications for economic policy. Economic policy has enjoyed large influence from Darwinian selection ideas, but Field (2008) points out that notable economists such as Paul Samuelson, Gary Becker, Jack Hirshleifer, and Friedrich Hayek all emphasized the importance of multilevel selection. Gowdy and Seidl (2004) and Wilson and Gowdy (2015) find that individual behaviour which is best for the group seldom maximizes that individual's relative fitness within the group. In other words: doing what is best for your group is rarely also best for yourself as an individual. Therefore, economic policies based on Adam Smith's invisible hand or New Welfare Economics, which rely on the maximization of individual level 'economic fitness', do not necessarily result in socially optimal outcomes. Richerson *et al*. (2016) recognize this problem, and argue that institutions in modern economies often promote economic group selection by benefitting ordinary citizens, thereby meeting common social goals over individual goals. Gowdy and Seidl (2004) and Wilson and Gowdy (2015) conclude that economic policy should extend its focus towards intermediate level small groups, in between the government and individual agents/firms. This bears substantial similarities with Elinor Ostrom's concept of polycentric governance (Ostrom, 2010).

2.3. Research problem

All in all, it can be concluded that the heated debate on multilevel selection theory is far from settled. Evolutionary biologists are still choosing sides, and the limited contributions from economists have not significantly changed the attractiveness of either side yet. Despite the potential implications for economic policy, empirical evidence regarding multilevel selection theory is largely absent. Richerson *et al.* (2016) discuss how modern-day group selection does not result in an immediate change in our gene pool, but has been “important in the ongoing evolution of complex societies”. Thus, data on modern group selection processes can yield insights in current evolution of societies, such as cultural evolution. Moreover, it can broaden our general understanding of multilevel selection, both from the genetic and cultural perspectives (Field, 2008). Therefore, the goal of this study is to build on the understanding of multilevel selection theory from an empirical economic perspective.

2.4. Conceptual framework

As we know, group selection occurs through groups with altruists outcompeting groups with egoists due to their better abilities to cooperate and resulting higher levels of group fitness. In accordance with multilevel selection theory, in this study it is hypothesized that the relationship between group fitness growth and group altruism is positive. By using data on development and inequality from communities of households in rural Africa, the relationship between fitness growth and altruism can be examined at the group level. The way these communities change over time in education, women’s height, and women’s age at first marriage are used as three unique measures of group fitness growth. Likewise, community (in)equality in education, women’s height, and women’s age at first marriage are used as three unique measures of group altruism.

First, education plays a key role in development, being one of the three dimensions in the UNDP Human Development Index. Second, height is related to child stunting and thereby provides information on development. Stunting is the result of poor nutrition at a young age, and shows its effects in underdeveloped height for an individual’s lifetime. Child stunting is high in Africa, with an estimated 37.9% and 29.1% of African children affected in 2000 and 2019 respectively (World Health Organization, 2020). Third, women’s age at first marriage is positively connected to levels of income and education, which again are both strongly related to human development (Garenne, 2004).

Next, inequality in education, height, and women’s age at first marriage all indicate how these outcomes are distributed between individuals in the same community. Indirectly, they provide information on how (public) resources such as schooling and child nutrition are shared in the community. Additionally, community age difference between spouses is used as a measure of group gender altruism. Age difference between spouses can be seen as a “proxy for conjugal distance and gender inequalities”, allowing for the examination of the role of gender inequality in group selection (Barbieri, Hertrich, & Grieve, 2005, p. 654). Lastly, the role of homogeneity within groups can be examined. Both ethnic and religious fractionalization can be used to measure (the lack of) homogeneity within groups. As mentioned earlier, Eaton *et al*. (2011) stress the importance of ethnically homogeneous groups in group selection. DeMartini and Marriott (2018) discuss the role of shared culture and norms within groups, which is exactly what religious fractionalization can capture. To test these claims, both ethnic and religious fractionalization measures are used to examine the relationship between homogeneity within communities and community development.

Furthermore, it is hypothesized that the strength of the relationship between group fitness growth and group altruism is moderated by contexts which are theorized to be optimal for group selection. These two contexts are heterogeneity between groups and high stakes of success. Unfortunately, measuring heterogeneity between groups in this study is not feasible due to data constraints on the geographical proximity of communities. High stakes of success, however, can be measured using child mortality rates. The context of high stakes of success mentioned by Waring and Acheson (2018) and Waring *et al*. (2017) is best characterized by a life-or-death context such as warfare, where the consequences of (a lack of) group altruism on group fitness growth are enormous.[3] Child mortality rates are used to mimic this life-or-death context. High child mortality rates in a community indicate that survival is not straightforward, and that group altruism could make the difference between life or death. If the empirical relationship between group fitness growth and group altruism grows stronger as child mortality rates increase, we could have more robust empirical support for multilevel selection theory.

Finally, a number of control factors are expected to affect community development. First and foremost, the initial level of community development is controlled for. If the initial level of community development is relatively high, then there is less room left for development. For instance, if all individuals in a community initially enjoy 15 years of education, the room for improvement before reaching a development ceiling is smaller than when those individuals would initially only enjoy 1 year of education. Moreover, catching up in development is easier than leading in development. In an area with low development, for instance, it is easier for a community to find elementary level teachers than it is to find university level teachers.

[3] We will stick to the term ‘high stakes of success’, as this is how it is referred to in existing literature. However, in the context of warfare or child mortality, ‘high stakes of failure’ would be more accurate.

Next, the community's number of children per woman during the development period is expected to have a negative relationship with community development, and is therefore controlled for. Essentially, this can be interpreted as a measure of the youth dependency ratio in the community. A large number of children per woman during the development period results in relatively high competition for personal development among children during this period (Lam & Marteleto, 2008). Hadley *et al.* (2011), for instance, report that youth in rural Ethiopia from households with high dependency ratios have a relatively poor nutritional status. Likewise, Lam and Marteleto (2005) find that a decrease in the size of the school-age population results in large increases in school enrollment rates in Brazil. Moreover, when a community's number of children per woman is relatively high, a relatively large share of the community's resources will have to be used to support the expanding population, by for instance expanding housing and sanitation. Such support does not change the mean level of community development, and leaves relatively little resources left to be invested in community development. As such, the growth in community development is expected to be hampered by a relatively high number of children per woman during the development period.

Finally, the year of data collection and the subnational region in which the community is located could influence community development. Therefore, dummy variables for the year of data collection and the subnational region are used to control for these two possible disturbances. Controlling for subnational regions is beneficial for obvious reasons, such as controlling for differences in initial levels of development between subnational regions. However, it also helps to control for regional effects we might be unaware for. Migration, for instance, could have an impact on community development. As data on community migration is not available, it is not possible to control for this directly. However, as Nshimbi and Moyo (2017) point out, migration in Africa is usually a result of either climate pressure, conflict, or poverty. It is therefore likely that any significant form of (mass) migration is specific to certain regions. With the subnational region dummies, these effects could also be captured.

3. Methods

Ordinary Least Squares (OLS) regression models with fixed effects dummies for the years of data collection and the subnational regions in which the communities are located (at the level of provinces) are used to examine the relationship between community development over a 20-year period and initial community equality. Control variables and an interaction term are added to the models. Section 3.1 covers the data used in this study. Sections 3.2 and 3.3 present the main variables and control variables respectively. Missing observations are discussed in Section 3.4, and finally an overview of the models is provided in Section 3.5.

3.1. Data

Individual and household level data from the Demographic and Health Surveys Program (DHS) on 39 developing countries in Africa over the period 1992-2019 are used. The data are sampled in geographical clusters of households, where each cluster represents a village or an urban area/neighbourhood. The data from rural clusters, or villages, are excellent representations of communities which can compete in development. Data from 33,836 of such rural clusters in Africa are available, of which 33,557 are used in the final analyses in this study. 279 clusters are dropped from analyses due to missing data. On average each cluster contains data from 26 households which contain an average total of 135 individuals. Table 1 below provides an overview of the countries, subnational regions, and clusters used for analyses. An overview of the years of data collection per country can be found in Table A1 in the Appendix section.

Table 1. Countries used for analysis by country name, number of subnational regions, and number of clusters.

Country	Subnational regions	Clusters	Country	Subnational regions	Clusters
Angola	18	422	Liberia	15	203
Benin	6	1420	Madagascar	22	445
Burkina Faso	13	960	Malawi	13	2427
Burundi	5	749	Mali	7	1250
Cameroon	10	800	Mauritania	11	106
Central African Republic CAR	5	123	Mozambique	10	994
Chad	7	678	Namibia	13	723
Comoros	3	214	Niger	7	747
Congo Brazzaville	10	313	Nigeria	37	2361
Congo Democratic Republic	10	550	Rwanda	5	1640
Cote d'Ivoire	10	464	Sao Tome & Principe	4	52
Eritrea	6	350	Senegal	10	1436
Eswatini	4	164	Sierra Leone	13	843
Ethiopia	11	1641	South Africa	9	540
Gabon	9	232	Tanzania	25	1771
Gambia	6	134	Togo	5	356
Ghana	10	939	Uganda	8	1127
Guinea	8	825	Zambia	9	1344
Kenya	7	2308	Zimbabwe	9	1031

Lesotho	10	875	Totals: 39	400	33557

Community data are only collected at one moment in time, and hence measuring changes in fitness over time is not straightforward. Likewise, it is challenging to compute control variables from the beginning of the development period. However, communities contain data from individuals of varying ages. Using this information, individuals can be put in different age groups. All individuals aged 40 to 49 represent the community at the beginning of the 20-year development period, and all individuals aged 20 to 29 represent the community at the end of the 20-year development period. The use of age groups introduces a constraint on the usability of variables though. Specifically, the only variables which can be used are individual level variables of which the observations are stable over time for adults, as they need to essentially take us back to the situation in the community about 20 years earlier. Using income, for instance, would not suffice as the income of individuals aged 40-49 today need not be the same as their income 20 years ago. Contrarily, an example of a variable which is relatively stable over time is years of education. Due to the low average years of education in rural Africa, an individual's lifetime years of education do not change significantly after the first 20 years of their life (Lewin, 2009; Mugisha, 2006). In other words: the data on years of education of individuals aged 40-49 today tell us something about the level of their education 20 years ago. This is crucial, as it adds a time dimension to the data, even though the data are only collected at one moment in time. If, for example, the data from a given community was collected in the year 2000, we can thus examine this community's change in development over the period ranging from 1980 to 2000.

Such estimations over time come with a margin of error though, due to migration and death of individuals who actually populated the community 20 years ago. These inaccuracies are inevitable given the nature of the data. As mentioned in Section 2.4, the use of fixed effects dummies for the subnational region in which a community is located do help to restrict the erroneous effects caused by for instance migration. It is, however, not possible to control for these errors completely, and thus some degree of inaccuracy remains in the model.

3.2. Main variables

As discussed in Section 2.4, community development in education, women's height, and women's age at first marriage are used as three unique measures of group fitness growth. Specifically, community development over the 20-year period is computed as the difference in development levels between individuals currently aged 20-29 and individuals currently aged 40-49. At the beginning of the 20-year development period, a community's educational development level is thus measured as the mean years of education of all individuals currently aged 40-49 in that community. Likewise, a community's educational development level at the end of the 20-year development period is measured as the mean years of education of all individuals currently aged 20-29 in that community. The same method is used to compute community development in women's height and women's age at first marriage, using women's mean height and mean women's age at first marriage respectively. Notice how years of education, women's height, and women's age at first marriage are all variables which are stable over time for adults. Individuals in rural Africa are rarely enrolled in education after they reach 20 years of age (Lewin, 2009; Mugisha, 2006). Their total years of education and women's height are unlikely to change significantly in the 30 years thereafter. Women's age at first marriage is by definition only available for ever married women and then stable for the rest of their lifetime. For the remainder of this study, the three dependent variables are referred to as *Change in education*, *Change in women's height*, and *Change in women's age of marriage*.

Next, community (in)equality in education, women's height, and women's age at first marriage are used as three unique measures of group altruism. Community inequality is measured at the beginning of the 20-year development period, by computing the Gini coefficient for years of education, women's height, and women's age at first marriage for individuals currently aged 40-49 in the community. The resulting three independent variables are called: *Initial inequality in education*, *Initial inequality in women's height*, and *Initial inequality in women's age of marriage*, where 'initial' refers to the beginning of the 20-year development period. The Gini coefficient is chosen as a measure of inequality because it is a widely used mean and population size independent measure of inequality. Moreover, as the coefficient ranges from 0 (complete equality) to 1 (complete chaos), the resulting coefficients from a regression are relatively easy to interpret. The Gini coefficient is computed in SPSS for all communities, and the computation follows the method used in the INEQDECO Stata module by Jenkins (2021) which can be used to calculate numerous inequality indices in Stata.

Additionally, community age difference between spouses is used as a measure of gender altruism in the group. The data on age difference between spouses is defined as the age of the husband minus the age of the wife, and is only available for married/cohabitating females. Again, the age difference between spouses is relatively stable over time for adults, making this a suitable variable for our analyses.[4] As the age difference between spouses is already a measure of gender inequality, the community mean is used instead of the Gini coefficient. Notice that the mean age difference between spouses can be both positive and negative, depending on whether the community is male or female dominated. One could therefore argue that the absolute value (modulus) of age difference between spouses would be a better measure of gender inequality. However, as gender inequality in rural Africa is usually the result of male dominance, a negative mean age difference is interpreted as gender equality in this study. This also benefits the interpretation of regression coefficients. As such, community gender inequality at the beginning of the 20-year development period is computed as the mean age difference between spouses for female individuals currently aged 40-49 in the community. The resulting independent variable is called: *Initial age difference between spouses*.

On top of that, both ethnic and religious fractionalization can be used to examine the relationship between homogeneity within communities and community development. Data on the ethnic and religious group to which individuals belong is available, and can be used to compute ethnic and religious fractionalization indices. Fractionalization is measured at the beginning of the 20-year development period, using data from individuals currently aged 40-49 in the community. Since ethnicity and religion remain mostly identical over an adult's lifetime, the fractionalization indices can provide information about the situation 20 years ago. Fractionalization is computed as described in Fearon (2003). The measure describes the probability that two randomly selected individuals from a community belong to different ethnic/religious groups: $Fractionalization \equiv 1 - \sum_{i=1}^{n} p_i^2$, where the population shares of ethnic/religious groups in the community are denoted $p_1, p_2, \ldots, p_n$. The resulting independent variables are called: *Initial ethnic fractionalization* and *Initial religious fractionalization*.

Finally, as discussed in Section 2.4, community child mortality rates are used to measure a stakes of success (life-or-death) context in the community, which is theorized to play a moderating role in group selection. The community child mortality rate is computed as the number of children

[4] Divorce followed by a new marriage/cohabitation could potentially result in inaccuracies if the age difference in the new marriage/cohabitation is different from the initial age difference. The probability that such a divorce leads to a significant change in a community's mean age difference between spouses is limited, though not impossible. Given the large number of communities in the dataset, however, these potential inaccuracies are not expected to significantly impact the overall analyses.

that died divided by the number of children that were born from all mothers aged 40-49 in the community. The variable gives a good indication of how harsh living conditions have been in the community since the beginning of the 20-year development period. It is called: *Initial child mortality rate*. Interaction variables are then computed as the product of the relevant main independent variable and the *Initial child mortality rate*. Centered versions of these variables are used to compute the interaction variables.

3.3. Control variables

The most important control factor is the initial level of community development, at the beginning of the 20-year development period. In correspondence with our three dependent variables, the three respective control variables for the initial level of community development are computed as the mean years of education for individuals aged 40-49 in the community, the women's mean height for individuals aged 40-49 in the community, and the women's mean age of marriage for individuals aged 40-49 in the community. Only one of these three control variables is included in each model, matching the dependent variable used in the respective the model. The three control variables are called: *Initial education, Initial women's height,* and *Initial women's age of marriage* respectively. Moreover, the community's number of children per woman during the development period is controlled for. This control variable is computed as the mean number of currently living children per woman aged 40-49 in the community, and reflects the community youth dependency ratio during the 20-year development period. The variable name is: *Initial birth rate*. Table 2 below provides some additional information about the data in general and an overview of all the variables.

Table 2. Descriptive statistics of aggregated community level data in rural Africa, DHS 1992-2019.

Variable name	N	Min	Max	Mean	SD
Number of individual level observations in the community	33557	19	594	135.8013	50.1943
Number of households sampled in the community	33557	10	134	25.7702	7.0923
Mean household wealth in the community (IWI)	33557	0	96.83	21.4479	13.7781
Change in education	29110	-6.54	10.77	1.3163	1.8378
Change in women's height	4498	-152.83	179.4	-6.3324	34.3009
Change in women's age of marriage	11707	-19.55	7.8	-0.7575	2.3583
Initial inequality in education	25136	0	0.97	0.5333	0.251
Initial inequality in women's height	5525	0	0.1	0.018	0.0074
Initial inequality in women's age of marriage	14113	0	0.32	0.0986	0.0454
Initial age difference between spouses	9328	-12.86	35.8	10.0959	4.6597
Initial education	29410	0	15.86	3.2672	2.8826
Initial women's height	5525	1316.4	1810.8	1584.7314	37.798
Initial women's age of marriage	14113	10.67	36.4	18.3125	2.7056
Initial ethnic fractionalization	23565	0	0.84	0.1585	0.2211

Initial religious fractionalization	23556	0	0.83	0.2885	0.2503
Initial birth rate	14580	0.6	10.4	5.2418	1.241
Initial child mortality rate	14580	0	0.7	0.19	0.1118

3.4. Missing observations

As can be observed in Table 2, not all community variables have observations for each community. One reason for this is the difference in data collection between different regions and years, resulting in some variables not being collected in all regions for all years. The other reason is that in this study, community level observations are deemed useful for analyses if they are based on data from at least five unique individuals, and if the community contains any data from at least ten unique households. The initial community level of development and initial community inequality could, theoretically, be computed using data from only two individuals. The fewer individuals used to compute a community level variable, however, the lower the accuracy of the community level variable will be. To guarantee a minimal degree of accuracy of all community level variables, the minimal of five individual observations per community variable has been set. Moreover, only communities with data from at least ten unique households are kept for the analyses. This requirement is set to minimize the probability that even when the minimal of five unique individuals is satisfied in a community, these are not five individuals representing only one or two unique households in the community. The main reason for the large share of missing observation for specifically the variables related to women's age of marriage, age difference between spouses, and child birth and mortality, is that data for these variables are only collected for women. This makes it harder to reach the minimal of five individual observations per community. Data on women's height has not been collected consistently in some year and country combinations, resulting in even fewer communities with five valid individual observations. Also, there are relatively many young individuals in rural Africa which is why there are more observations for individuals aged 20-29 than for individuals aged 40-49.

Table 3 below provides an overview of how both requirements affect the number of valid observations for key community level components over different age groups, which are used to compute the final variables. Note in Table 2 that the *Initial inequality in education* variable only has 25,136 valid observations, whereas the *Initial education* variable has 29,410 valid observations. Both are computed using the same data, yet there is a difference of 4,274 observations between the two variables. The reason for this difference is that the mean years of education can equal zero in a community, but the Gini coefficient cannot be calculated for any population with zero mean years of education. One might argue that a community with zero mean years of education is a perfectly equal community. In this study, however, the Gini coefficient is used to measure community altruism to explain community development. The lack of an accessible school in a community for instance, resulting in zero mean years of education, does not reflect any information about altruism in the community. Moreover, roughly 25% of communities with zero mean years of education at the start of the 20-year development period also have zero mean years of education at the end of this period, most likely due to the lack of an accessible school. This information should not be used to draw the conclusion that perfect community equality in years of education results in extremely low community development. The difference of 4,274 valid observations between the *Initial inequality in education* and *Initial education* variables is thus explained by the 4,274 communities where the mean years of education is zero for individuals aged 40-49.

Table 3. Overview of how the number of valid community observations drops for key components due to two requirements.

Component name	N, no restrictions	N, min. of 5 individuals	N, min. of 10 households
Years of education for ages 20-29	33816	33241 (-575)	33081 (-160)
Years of education for ages 40-49	33730	29465 (-4265)	29410 (-55)
Women's height for ages 20-29	25139	15449 (-9690)	15443 (-6)
Women's height for ages 40-49	23163	5525 (-17638)	5525 (0)
Women's age of marriage for ages 20-29	33207	25633 (-7574)	25597 (-36)
Women's age of marriage for ages 40-49	33195	14119 (-19076)	14113 (-6)
Age difference between spouses for ages 40-49	29860	9331 (-20529)	9328 (-3)
Child births for ages 40-49	32721	14586 (-18135)	14580 (-6)
Child mortality for ages 40-49	32721	14586 (-18135)	14580 (-6)
Ethnicity for ages 40-49	28540	23614 (-4926)	23565 (-49)
Religion for ages 40-49	27929	23600 (-4329)	23556 (-44)

3.5. Models

The relationship between community development and community inequality is analyzed using OLS regression models with fixed effects dummies for the years of data collection and the subnational regions in which the communities are located (at the level of provinces). Dummies are computed for all 28 years of data collection, ranging from 1992 to 2019, and for all 400 subnational regions.

In total, nine unique combinations of dependent and independent variables are analyzed. First, the three dependent variables which measure community development (*Change in education*, *Change in women's height,* and *Change in women's age of marriage*) are matched with their respective community inequality independent variables (*Initial inequality in education*, *Initial inequality in women's height,* and *Initial inequality in women's age of marriage*). Next, we aim to examine the role of *Initial age difference between spouses*, *Initial ethnic fractionalization,* and *Initial religious fractionalization* in community development. To do so, these three independent variables are matched with both *Change in education* and *Change in women's age of marriage*, as these two dependent variables have the most valid observations. This results in $3 + 6 = 9$ unique combinations of dependent and independent variables which are used for analyses. An overview of the resulting nine OSL regression models with fixed effects dummies is presented in Table 4 below.

Table 4. Overview of the nine unique combinations of dependent and independent variables used for analysis.

Model	Dependent variable	Main independent variable
1	Change in education	Initial inequality in education
2	Change in women's height	Initial inequality in women's height
3	Change in women's age of marriage	Initial inequality in women's age of marriage
4	Change in education	Initial age difference between spouses
5	Change in women's age of marriage	Initial age difference between spouses
6	Change in education	Initial ethnic fractionalization
7	Change in women's age of marriage	Initial ethnic fractionalization
8	Change in education	Initial religious fractionalization
9	Change in women's age of marriage	Initial religious fractionalization

Next to the dependent and relevant independent and control variables, all nine models also include an interaction with the *Initial child mortality rate* variable. Adding the interaction term to the models has very little impact on the coefficients and significance of the other variables, and hence the versions of the models excluding the interaction term are only presented in the Appendix section.

To add additional robustness to the results we control for missing observations of variables. In this study, observations of the *Initial birth rate* and *Initial child mortality rate* variables are not valid for all observations of the other variables in the models. To prevent a loss in the models' degrees of freedom, the dummy variable adjustment procedure for missing observations has been followed (Allison, 2001). All missing observations of the original two variables are replaced by the variables' means. Moreover, a dummy variable is created which equals one for every missing observation in the original two variables. Since the two variables have identical missing observations, only one dummy variable is needed. Thereby, all models can still utilize all valid observations of their dependent variables, and the effects of missing data on childbirths (and consequently child mortality) are controlled for. The updated *Initial child mortality rate* variable is also used to create the interaction variables, with the same dummy variable representing the missing values of the original interaction variable.

Finally, the Variance Inflation Factor (VIF) has been computed for all independent variables in all models to diagnose potential multicollinearity issues. The VIF is below the critical value of 10 for all independent variables in all models. We therefore conclude that there are no substantial multicollinearity issues in any of the models.

4. Results

The results of the nine OLS regression models with fixed effects dummies for the years of data collection and the subnational regions in which the communities are located (at the level of provinces) are presented in three sections. Section 4.1 shows the three models related to years of education, women's height, and women's age of marriage to examine the relationship between community development and community inequality. Section 4.2 presents two models using data on the mean age difference between spouses in communities to explore the role of gender inequality in community development. In Section 4.3, models are presented in which both ethnic and religious fractionalization in the community are used as independent variables to explore the role of community homogeneity in community development.

The number of observations and number of subnational regions included in the model, as well as the adjusted R-squared, are presented at the bottom of all results tables. Note, however, that all models include over 300 subnational region dummy variables and around 20 year of data collection dummy variables. As a result, the adjusted R-squared should be interpreted with caution in this study, as the explanatory power of the models could be inflated by the large number of dummy variables.

4.1. Inequality

The results of the first of nine models are presented in Table 5 below. Almost all results are in line with our hypotheses. The relationship between the *Change in education* and the *Initial inequality* *in education*, after controlling for other factors, is negative and highly significant. In other words, communities in which years of education were distributed relatively equally at the beginning of the 20-year development period showed a significantly larger increase in mean years of education during these 20 years. Next, *Initial education* at the start of the development period is also significantly negatively related to the *Change in education*. Surprisingly, no significant effect is found for the *Initial birth rate*. The *Initial child mortality rate*, however, is significantly and negatively related to the *Change in education*. Finally, the interaction term in the model provides information on whether the effect of the *Initial inequality in education* is sensitive to changes in the *Initial child mortality rate* in the community. The significant and negative coefficient indicates that the effect of the *Initial inequality in education* on the *Change in education* is more negative in communities where the *Initial child mortality rate* is high.

Table 5. OLS regression model on the relationship between community change in education and community initial inequality in education in rural Africa.

Independent variable	Beta coefficient	Standard error
Intercept	3.718***	0.132
Initial inequality in education	-0.672***	0.099
Initial education	-0.543***	0.009
Initial birth rate	-0.009	0.012
Initial child mortality rate	-1.207***	0.142
Dummy for missing data on childbirths	-0.329***	0.020
Interaction between Initial inequality in education and Initial child mortality rate	-2.142***	0.521
Number of observations	24952	
Number of subnational regions	399	
Adjusted R-squared	0.427	

****p<0.001, **p<0.01, *p<0.05.*

Table 6 below shows the results of the model using the women's height variables. First of all, notice how the number of observations drop from 24,952 in Table 5 to 4,429 in Table 6 due to the large difference in valid observations between the years of education and women's height variables. Next, the relationship between the *Change in women's height* and the *Initial inequality in women's height* is not significant. Apart from the intercept, the only significant variables in this model are the *Initial women's height* and *Initial child mortality rate* variables. Both variables are negatively related to the *Change in women's height*, which is according to expectations.

Table 6. OLS regression model on the relationship between community change in women's height and community initial inequality in women's height in rural Africa.

Independent variable	Beta coefficient	Standard error
Intercept	1265.434***	21.328
Initial inequality in women's height	75.82	54.406
Initial women's height	-0.802***	0.013
Initial birth rate	-0.056	0.368
Initial child mortality rate	-21.78***	4.507
Dummy for missing data on childbirths	2.39	2.294
Interaction between Initial inequality in women's height and Initial child mortality rate	-558.364	447.236
Number of observations	4429	
Number of subnational regions	308	
Adjusted R-squared	0.509	

****p<0.001, **p<0.01, *p<0.05.*

Results of the model using the women's age of marriage variables are presented in Table 7 below. As with the years of education model, most results are in line with our hypotheses. The relationship between the *Change in women's age of marriage* and the *Initial inequality in women's age of marriage* is negative and highly significant. The variables *Initial women's age of marriage*, *Initial birth rate*, and *Initial child mortality rate* are also all significant and are negatively related to the *Change in women's age of marriage*. Last, the interaction term in the model is not significant, indicating that the effect of the *Initial inequality in women's age of marriage* is not sensitive to changes in the *Initial child mortality rate* in the community.

Table 7. OLS regression model on the relationship between community change in women's age of marriage and community initial inequality in women's age of marriage in rural Africa.

Independent variable	Beta coefficient	Standard error
Intercept	15.794***	0.180
Initial inequality in women's age of marriage	-1.529***	0.305
Initial women's age of marriage	-0.896***	0.006
Initial birth rate	-0.03**	0.011
Initial child mortality rate	-1.019***	0.135
Dummy for missing data on childbirths	0.852**	0.303
Interaction between Initial inequality in women's age of marriage and Initial child mortality rate	1.97	2.281
Number of observations	11711	
Number of subnational regions	378	
Adjusted R-squared	0.744	

****p<0.001, **p<0.01, *p<0.05.*

4.2. Gender inequality

Table 8 below displays the results of the models using the *Initial age difference between spouses* main independent variable. The *Change in education* dependent variable is used for the results on the left, whereas the *Change in women's age of marriage* dependent variable is used for the results on the right. First, it can be observed that the *Initial age difference between spouses* variable, which reflects community gender inequality, is only significantly and negatively related to the *Change in women's age of marriage*. When a community's mean age difference between spouses grows by one year, which implies that the average husband's age grows by one year relative to the average wife's age, the community's average women's mean age of marriage is predicted to decrease by 0.009 years. When examining the control variables, both the coefficients of *Initial education* and *Initial women's age of marriage* are significant and negative, as expected. The coefficients of the *Initial child mortality rate* variable are also significant and negative in both models. The *Initial birth rate* variable is only significant in the education model, and the coefficient is positive. This is not in line with expectations, as it implies that more children per woman result in an increase in the community's mean years of education. The dummy variable for missing data on childbirths is excluded as there are no such missing observations in these two models. The interaction term is not significant in both models.

Table 8. OLS regression models on the relationships between community age difference between spouses and both community change in education and community change in women's age of marriage in rural Africa.

	Change in education		Change in women's age of marriage	
Independent variable	Beta coefficient	S.E.	Beta coefficient	S.E.
Intercept	1.796***	0.181	15.399***	0.228
Initial age difference between spouses	-0.005	0.004	-0.009*	0.004
Initial education	-0.329***	0.011	—	—
Initial women's age of marriage	—	—	-0.893***	0.007
Initial birth rate	0.031*	0.015	-0.009	0.014
Initial child mortality rate	-1.022***	0.181	-0.827***	0.170
Interaction between Initial age difference between spouses and Initial child mortality rate	-0.048	0.031	0.025	0.028
Number of observations	9085		7881	
Number of subnational regions	387		371	
Adjusted R-squared	0.372		0.734	

****p<0.001, **p<0.01, *p<0.05.*

4.3. Homogeneity

In this section, both *Initial ethnic fractionalization* and *Initial religious fractionalization* are used as independent variables to explore the role of homogeneity within communities in community development. The results of the models using the *Initial ethnic fractionalization* independent variable are presented in Table 9 below. *Initial ethnic fractionalization* does not seem to affect neither the *Change in education* nor the *Change in women's age of marriage*. Effects of the control variable are mostly consistent with the other models. The results of the *Initial birth rate* variable are slightly inconsistent though, as no significant effect is found in the education model, but a negative and significant effect is found in the age of marriage model.

Table 9. OLS regression models on the relationships between community ethnic fractionalization and both community change in education and community change in women's age of marriage in rural Africa.

	Change in education		Change in women's age of marriage	
Independent variable	Beta coefficient	S.E.	Beta coefficient	S.E.
Intercept	2.202***	0.111	15.907***	0.201
Initial ethnic fractionalization	0.083	0.049	-0.106	0.068
Initial education	-0.478***	0.006	—	—
Initial women's age of marriage	—	—	-0.905***	0.006
Initial birth rate	-0.014	0.011	-0.032**	0.013
Initial child mortality rate	-1.267***	0.129	-1.119***	0.150
Dummy for missing data on childbirths	-0.297***	0.019	0.87**	0.306
Interaction between Initial ethnic fractionalization and Initial child mortality rate	0.1	0.519	0.366	0.516
Number of observations	23464		9767	
Number of subnational regions	373		350	
Adjusted R-squared	0.425		0.736	

****p<0.001, **p<0.01, *p<0.05.*

Table 10 below shows the results of the models using *Initial religious fractionalization* as the main independent variable. Surprisingly, the relationship between the *Change in education* and the *Initial religious fractionalization* is significant and positive. This implies that community homogeneity negatively affects community development, if religious fractionalization is used to measure homogeneity. No significant relationship, however, is found with the *Change in women's age of marriage*. The rest of the results are consistent with expectations, except for the interaction term which is not significant.

Table 10. OLS regression models on the relationships between community religious fractionalization and both community change in education and community change in women's age of marriage in rural Africa.

Independent variable	Change in education		Change in women's age of marriage	
	Beta coefficient	S.E.	Beta coefficient	S.E.
Intercept	2.105***	0.112	15.739***	0.197
Initial religious fractionalization	0.335***	0.050	-0.042	0.072
Initial education	-0.47***	0.006	—	—
Initial women's age of marriage	—	—	-0.901***	0.006
Initial birth rate	-0.025*	0.011	-0.03*	0.012
Initial child mortality rate	-1.313***	0.131	-1.048***	0.147
Dummy for missing data on childbirths	-0.286***	0.019	0.536	0.311
Interaction between Initial religious fractionalization and Initial child mortality rate	-0.732	0.466	0.316	0.471
Number of observations	23450		9985	
Number of subnational regions	379		366	
Adjusted R-squared	0.425		0.742	

******p<0.001, **p<0.01, *p<0.05.*

5. Conclusion and Discussion

The aim of this study was to build on the understanding of multilevel selection theory from an empirical economic perspective. We used data from 33,557 communities of households in rural Africa, spread over 39 countries and 400 subnational regions. OLS regression models with fixed effects dummies for the years of data collection and the subnational regions in which the communities are located (at the level of provinces) were used to examine the relationship between community development over a 20-year period and initial community equality. Additionally, the effects of gender equality and homogeneity within communities on community development were studied. In short, our results reveal that, over 20 years, community increases in years of education and women's age of marriage are highest in communities where these outcomes are initially distributed equally. We also find that communities with high gender equality show a stronger increase in women's age of marriage, whereas educational development seems unaffected. No evidence for a positive effect of homogeneity within communities on community development is found. Instead, we find that communities with high religious homogeneity show less improvement in education. Last, our analyses reveal that the positive relationship between community development and equality in education is strongest in communities with high child mortality rates. This suggests a moderating role of a high stakes of success (life-or-death) context, though the finding was not replicated in other models in this study.

Our empirical evidence provides support for a positive relationship between fitness growth and altruism at the group level, as hypothesized in multilevel selection theory. This marks the first empirical support for group selection based on a large-scale statistical analysis. Studies by Tomaszewski (2021) and Waring and Acheson (2018) also provide empirical support for group selection, but are both based on long-term qualitative analyses of intergroup competition. The fierce debate on multilevel selection theory in evolutionary biology benefits from the availability of our new empirical evidence. Some degree of uncertainty surrounding multilevel selection theory can be taken away, as there is now finally support based on a large amount of data from human communities. New lessons can be learned about the evolution of human beings, but also about how our evolutionary process differs from that of other social species.

Additionally, our evidence provides limited support for a positive effect of gender equality on group fitness growth. The role of gender equality in multilevel selection theory had not been studied before, and therefore this study opens up a new area of interest for both evolutionary biologists and economists. The level of gender equality in a community tells us something about altruism specifically related to gender. Possibly, the gender roles associated with high gender inequality are not effective in today's community development. This could explain why communities with high gender equality show high development, and would imply that (cultural) selection of gender roles could happen. That is, gender roles which are beneficial for community development survive and are imitated, whereas others eventually disappear (Waring & Acheson, 2018). This explanation, however, remains speculative and is sensitive to the large variety of cultures and corresponding gender roles in rural Africa. As such, the role of gender equality in group selection needs further investigation from both a theoretical and empirical perspective.

Next, the empirical evidence in this study calls for further examination of the role of homogeneity within communities in multilevel selection theory. Our evidence provides limited support for a negative effect of religious homogeneity within groups on group fitness growth. This contradicts existing theoretical and empirical evidence on multilevel selection theory. The discrepancy could possibly be related to the measurement of homogeneity used in this study. A fractionalization index was computed as the probability that two randomly selected individuals from a community belong to different religious groups. Therefore, communities with for instance 50% Catholics and 50% Muslims have a lower fractionalization value than communities with 33% Catholics, 33% Protestants, and 33% Orthodox. The latter example, however, is 100% Christian and therefore perhaps more homogeneous as the measure suggests. Moreover, in this study we only examined the relationship between community homogeneity and development, but not the relationship between homogeneity and altruism directly. All in all, questions are raised which could be answered with future empirical evidence on the role of homogeneity within groups in multilevel selection theory.

Finally, our empirical evidence provides limited support for a moderating role of a high stakes of success (life-or-death) context. This finding adds robustness to our thesis, as it indicates that the effect of group altruism on group fitness growth is largest in an optimal group selection context, in our case when community child mortality rates are high. Data on intergroup conflict, natural resource proximity to groups, and additional measures of poverty and other harsh living conditions could all potentially improve our understanding of the moderating role of a high stake of success context in the future. The use of additional interaction terms with different contextual variables could teach us more about the context under which group selection may take place. Additionally, heterogeneity between communities is theorized to play an important moderating role in multilevel selection theory, but no data was available on the proximity of communities. Future studies could investigate this context if data becomes available.

The evidence presented in this study comes with a few limitations. First, there could be inaccuracies in the data from individuals currently aged 40-49, who reflect the community at the beginning of the 20-year development period, due to possible deaths and migration. This could give a false idea about the initial levels of inequality and development in the community. Fixed effects dummies for years of data collection and subnational regions can control for this to some extent, for instance by controlling for mass migration events specific to certain years and subnational regions. However, inaccuracies remain in the estimation of the situation in the communities at the beginning of the 20-year development period, and therefore results should be interpreted with caution. A breakthrough improvement for future studies would be to use community data collected in two different time periods to overcome the inaccuracies of using age groups as a measure of time.

Next, (in)equality is used as a measure of group altruism, and changes in years of education, in women's height, and in women's age of marriage are used to reflect group fitness growth. These measures are the best possible given the data available, but there is room for potential improvements. For instance, community redistribution of income and wealth/assets could be an important form of group altruism, but cannot be captured given the nature of our data. Moreover, group fitness growth could ideally be measured as a broad index such as the Human Development Index, instead of only considering changes in years of education, women's height, or women's age of marriage. Part of these hurdles could again be overcome by using community data collected in two different time periods, as this reliefs the constraint of using only variables which are stable over an adult's lifetime. Generally, future studies could improve on our analyses when more and better data are available to measure group altruism and group fitness growth.

Additionally, the relationship between fitness growth and altruism at the group level is only examined over a relatively short time period in this study. The longest possible time period for community development to be examined in this study was 20 years, given the data constraints. Hence, the observed effects of inequality on development seem rather small in this study. When looking at years of education, for example, an increase in the Gini coefficient of 0.1 results in 0.672 less years of educational development in the community over a 20-year period. In contrast, Wilson and Kniffin (1998) look at 24 generations in their theoretical model, which would correspond to over 480 years in the context of this study if one generation is assumed to be 20 years. Waring and Acheson (2018) investigate group selection over the past century, and Tomaszewski (2021) studies human cooperation over the past millennia. Unfortunately, community level data over such long time periods are not available. If the relationship found in this study holds over a longer time period, however, the effect of sustained inequality would grow larger over time. This would lead to substantial differences in group fitness, potentially resulting in group selection.

Last, no significant relationship was found between community initial equality in women's height and change in women's height. Compared to the other models in this study, there is a substantial difference in the number of observations and subnational regions. Still, it would be interesting to see if the results would hold for height too if more data are available. Alternatively, it could be that community initial equality in women's height and change in women's height, through for instance child stunting, are not empirically related. This would be a novel finding, though more evidence is needed for such conclusions.

All in all, this study contributes to existing literature by providing the first empirical support for group selection based on a large-scale statistical analysis. Our findings suggests that group selection could be expected to take place under the assumptions that (1) the relationships found in this study accurately reflect the relationships between group altruism and group fitness growth, and that (2) these relationships are also consistently true over longer-term time periods. Consequently, fruitful lessons can be derived for policymakers, specifically in developing countries. Given that there is a relationship between community development and community equality in rural Africa, economic policies aimed at promoting cooperation at the community level could result in higher levels of community development. Inspiration could be drawn from Elinor Ostrom's concept of polycentric governance, with multiple local governing authorities.

6. References

Abbot, P., Abe, J., Alcock, J., Alizon, S., Alpedrinha, J. A., Andersson, M., ... & Zink, A. (2011). Inclusive fitness theory and eusociality. *Nature*, *471*(7339), E1-E4.

Akçay, E., & Cleve, J. V. (2012). Behavioral responses in structured populations pave the way to group optimality. *The American Naturalist*, *179*(2), 257-269.

Allison, P. D. (2001). *Missing data*. Sage publications.

Apicella, C. L., & Silk, J. B. (2019). The evolution of human cooperation. *Current Biology*, *29*(11), R447-R450.

Atkisson, C., & Mulder, M. B. (2020). The structure of multiplex networks predicts play in economic games and real-world cooperation. *arXiv preprint arXiv:2012.07669*.

Aumann, R. J. (2019). A synthesis of behavioural and mainstream economics. *Nature human behaviour*, *3*(7), 666-670.

Balliet, D., Wu, J., & De Dreu, C. K. (2014). Ingroup favoritism in cooperation: a meta-analysis. *Psychological bulletin*, *140*(6), 1556.

Barbieri, M., Hertrich, V., & Grieve, M. (2005). Age difference between spouses and contraceptive practice in sub-Saharan Africa. *Population*, *60*(5), 617-654.

Birch, J. (2019). Are kin and group selection rivals or friends?. *Current Biology*, *29*(11), R433-R438.

Boomsma, J. J., Beekman, M., Cornwallis, C. K., Griffin, A. S., Holman, L., Hughes, W. O., ... & Ratnieks, F. L. (2011). Only full-sibling families evolved eusociality. *Nature*, *471*(7339), E4-E5.

Darwin, C. (1859). On the Origin of Species.

Dawkins, R. (2012). The descent of Edward Wilson. *Prospect Magazine*, *195*, 1-59.

DeMartini, S., & Marriott, C. (2018, July). Multilevel Selection Can Lead to Cooperation in a Public Goods Game. In *Artificial Life Conference Proceedings* (pp. 485-492). One Rogers Street, Cambridge, MA 02142-1209 USA journals-info@ mit. edu: MIT Press.

Eaton, B. C., Eswaran, M., & Oxoby, R. J. (2011). 'Us' and 'Them': the origin of identity, and its economic implications. *Canadian Journal of Economics/Revue canadienne d'économique*, *44*(3), 719-748.

Fearon, J. D. (2003). Ethnic and cultural diversity by country. *Journal of economic growth*, *8*(2), 195-222.

Ferriere, R., & Michod, R. E. (2011). Inclusive fitness in evolution. *Nature*, *471*(7339), E6-E8.

Field, A. J. (2008). Why multilevel selection matters. *Journal of Bioeconomics*, *10*(3), 203-238.

García, J., & van den Bergh, J. C. (2011). Evolution of parochial altruism by multilevel selection. *Evolution and Human Behavior*, *32*(4), 277-287.

Gardner, A. (2015). More on the genetical theory of multilevel selection. *Journal of evolutionary biology*, *28*(9), 1747-1751.

Garenne, M. (2004). Age at marriage and modernisation in sub-Saharan Africa. *Southern African Journal of Demography*, 59-79.

Goodnight, C. J. (2015). Multilevel selection theory and evidence: a critique of Gardner, 2015. *Journal of evolutionary biology*, *28*(9), 1734-1746.

Gowdy, J., & Seidl, I. (2004). Economic man and selfish genes: the implications of group selection for economic valuation and policy. *The Journal of Socio-Economics*, *33*(3), 343-358.

Hadley, C., Belachew, T., Lindstrom, D., & Tessema, F. (2011). The shape of things to come? Household dependency ratio and adolescent nutritional status in rural and urban Ethiopia. *American journal of physical anthropology*, *144*(4), 643-652.

Hamilton, W. D. (1964). The genetic theory of social behavior. I and II. *Journal of theoretical biology*, *7*, 1-52.

Herre, E. A., & Wcislo, W. T. (2011). In defence of inclusive fitness theory. *Nature*, *471*(7339), E8-E9.

Hodler, R. (2006). The curse of natural resources in fractionalized countries. *European Economic Review, 50*(6), 1367-1386.

Hodler, R., & Knight, D. S. (2012). Ethnic fractionalisation and aid effectiveness. *Journal of African Economies*, *21*(1), 65-93.

Jeler, C. (Ed.). (2018). *Multilevel Selection and the Theory of Evolution: Historical and Conceptual Issues*. Springer.

Jenkins, S. (2021). INEQDECO: Stata module to calculate inequality indices with decomposition by subgroup.

Kramer, J., & Meunier, J. (2016). Kin and multilevel selection in social evolution: a never-ending controversy?. *F1000Research*, *5*.

Kramer, K. L. (2010). Cooperative breeding and its significance to the demographic success of humans. *Annual Review of Anthropology*, *39*, 417-436.

Lam, D., & Marteleto, L. (2005). Small families and large cohorts: The impact of the demographic transition on schooling in Brazil. *The changing transitions to adulthood in developing countries: selected studies*, 56-83.

Lam, D., & Marteleto, L. (2008). Stages of the demographic transition from a child's perspective: Family size, cohort size, and children's resources. *Population and development review*, *34*(2), 225-252.

Leigh Jr, E. G. (2010). The group selection controversy. *Journal of Evolutionary Biology*.

Lewin, K. M. (2009). Access to education in sub-Saharan Africa: patterns, problems and possibilities. *Comparative Education*, *45*(2), 151-174.

Makowsky, M. D., & Smaldino, P. E. (2016). The evolution of power and the divergence of cooperative norms. *Journal of Economic Behavior & Organization*, *126*, 75-88.

McAndrew, F. T. (2002). New evolutionary perspectives on altruism: Multilevel-selection and costly-signaling theories. *Current Directions in Psychological Science*, *11*(2), 79-82.

Mugisha, F. (2006). School enrollment among urban non-slum, slum and rural children in Kenya: Is the urban advantage eroding?. *International Journal of Educational Development*, *26*(5), 471-482.

Nowak, M. A. (2006). Five rules for the evolution of cooperation. *science*, *314*(5805), 1560-1563.

Nowak, M. A. (2012). Evolving cooperation. *Journal of theoretical biology*, *299*(0), 1-8.

Nowak, M. A., Tarnita, C. E., & Wilson, E. O. (2010). The evolution of eusociality. *Nature*, *466*(7310), 1057-1062.

Nshimbi, C., & Moyo, I. (2017). *Migration, cross-border trade and development in Africa*. Palgrave Studies of Sustainable Business in Africa). Cham, Switzerland: Palgrave Macmillan.

Ostrom, E. (2010). Beyond markets and states: polycentric governance of complex economic systems. *American economic review*, *100*(3), 641-72.

Rand, D. G., & Nowak, M. A. (2013). Human cooperation. *Trends in cognitive sciences*, *17*(8), 413-425.

Richerson, P., Baldini, R., Bell, A. V., Demps, K., Frost, K., Hillis, V., ... & Zefferman, M. (2016). Cultural group selection plays an essential role in explaining human cooperation: A sketch of the evidence. *Behavioral and Brain Sciences*, *39*.

Safarzyńska, K., & van den Bergh, J. C. (2010). Evolutionary models in economics: a survey of methods and building blocks. *Journal of Evolutionary Economics*, *20*(3), 329-373.

Schumpeter, J. A. (1982). The theory of economic development: An inquiry into profits, capital, credit, interest, and the business cycle (1912/1934). *Transaction Publishers.–1982.–January*, *1*, 244.

Schwesinger, G. (2013). *Natural and economic selection: Lessons from the evo-devo and multilevel selection debate* (No. 2013-014). Jena Economic Research Papers.

Smith, V. L. (2003). Constructivist and ecological rationality in economics. *American economic review*, *93*(3), 465-508.

Strassmann, J. E., Page, R. E., Robinson, G. E., & Seeley, T. D. (2011). Kin selection and eusociality. *Nature*, *471*(7339), E5-E6.

Todd, P. M., & Gigerenzer, G. E. (2012). *Ecological rationality: Intelligence in the world*. Oxford University Press.

Tomasello, M. (2014). The ultra-social animal. *European journal of social psychology*, *44*(3), 187-194.

Tomaszewski, M. (2021). Egoism and cooperation in economic development-a historical approach. *Economic Research-Ekonomska Istraživanja*, 1-16.

Traulsen, A., & Nowak, M. A. (2006). Evolution of cooperation by multilevel selection. *Proceedings of the National Academy of Sciences*, *103*(29), 10952-10955.

Van Veelen, M. (2009). Group selection, kin selection, altruism and cooperation: when inclusive fitness is right and when it can be wrong. *Journal of Theoretical Biology*, *259*(3), 589-600.

Waring, T. M., Goff, S. H., & Smaldino, P. E. (2017). The coevolution of economic institutions and sustainable consumption via cultural group selection. *Ecological economics*, *131*, 524-532.

Waring, T., & Acheson, J. (2018). Evidence of cultural group selection in territorial lobstering in Maine. *Sustainability science*, *13*(1), 21-34.

Wilson, D. S. (2018). Multilevel selection and Tomasello's A Natural History of Human Morality: A translation manual. *Philosophical Psychology*, *31*(5), 669-679.

Wilson, D. S., & Gowdy, J. M. (2015). Human ultrasociality and the invisible hand: Foundational developments in evolutionary science alter a foundational concept in economics. *Journal of Bioeconomics*, *17*(1), 37-52.

Wilson, D. S., & Kniffin, K. M. (1999). Multilevel selection and the social transmission of behavior. *Human Nature*, *10*(3), 291-310.

Wilson, D. S., & Wilson, E. O. (2007). Rethinking the theoretical foundation of sociobiology. *The Quarterly review of biology*, *82*(4), 327-348.

Wilson, E. O. (2012). *The social conquest of earth*. WW Norton & Company.

World Health Organization. (2020). UNICEF/WHO/The World Bank Group joint child malnutrition estimates: levels and trends in child malnutrition: key findings of the 2020 edition.

Zinovyeva, N. (2010). Multilevel selection processes in economics: theory and methods. *Evolutionary and Institutional Economics Review*, *6*(2), 277-298.

7. Appendix

Table A 1. An overview of the years of data collection per country.

Country	Year of data collection (number of clusters)					
Angola	2011 (142)	2016 (280)				Total (422)
Benin	1996 (108)	2001 (128)	2006 (440)	2011 (440)	2018 (304)	Total (1420)
Burkina Faso	1993 (115)	1998 (138)	2003 (310)	2010 (397)		Total (960)
Burundi	2010 (301)	2017 (448)				Total (749)
Cameroon	1998 (107)	2004 (218)	2011 (283)	2018 (192)		Total (800)
Central African Republic CAR	1994 (123)					Total (123)
Chad	1997 (127)	2004 (90)	2015 (461)			Total (678)
Comoros	1996 (70)	2012 (144)				Total (214)
Congo Brazzaville	2005 (62)	2011 (251)				Total (313)
Congo Democratic Republic	2007 (175)	2013 (375)				Total (550)
Cote d'Ivoire	1994 (100)	1999 (33)	2005 (140)	2011 (191)		Total (464)
Eritrea	1995 (101)	2002 (249)				Total (350)
Eswatini	2006 (164)					Total (164)
Ethiopia	2000 (400)	2005 (391)	2011 (410)	2016 (440)		Total (1641)
Gabon	2000 (81)	2012 (151)				Total (232)
Gambia	2013 (134)					Total (134)
Ghana	1998 (261)	2003 (238)	2008 (229)	2014 (211)		Total (939)
Guinea	1999 (175)	2005 (195)	2012 (193)	2018 (262)		Total (825)
Kenya	1993 (364)	1998 (437)	2003 (269)	2008 (266)	2014 (972)	Total (2308)
Lesotho	2004 (288)	2010 (306)	2014 (281)			Total (875)
Liberia	2013 (203)					Total (203)

Madagascar	2009 (445)									Total (445)
Malawi	1992 (167)	2000 (444)	2004 (448)	2010 (691)	2016 (677)					Total (2427)
Mali	1995 (179)	2001 (273)	2006 (260)	2013 (297)	2018 (241)					Total (1250)
Mauritania	2001 (106)									Total (106)
Mozambique	1997 (265)	2003 (375)	2011 (354)							Total (994)
Namibia	2000 (154)	2006 (286)	2013 (283)							Total (723)
Niger	1998 (174)	2006 (227)	2012 (346)							Total (747)
Nigeria	1999 (221)	2003 (195)	2008 (607)	2013 (525)	2018 (813)					Total (2361)
Rwanda	1992 (146)	2000 (351)	2005 (351)	2010 (413)	2015 (379)					Total (1640)
Sao Tome & Principe	2009 (52)									Total (52)
Senegal	2005 (214)	2011 (244)	2012 (121)	2014 (121)	2015 (130)	2016 (130)	2017 (214)	2018 (132)	2019 (130)	Total (1436)
Sierra Leone	2008 (208)	2013 (277)	2019 (358)							Total (843)
South Africa	1998 (280)	2016 (260)								Total (540)
Tanzania	1992 (255)	1996 (251)	1999 (116)	2004 (362)	2010 (359)	2015 (428)				Total (1771)
Togo	1998 (154)	2014 (202)								Total (356)
Uganda	2006 (308)	2011 (285)	2016 (534)							Total (1127)
Zambia	1996 (170)	2002 (217)	2007 (195)	2014 (415)	2018 (347)					Total (1344)
Zimbabwe	1994 (145)	1999 (144)	2006 (271)	2011 (237)	2015 (234)					Total (1031)
										Total (33557)

Tables A2 to A7 below show the regression models with fixed effects dummies for the years of data collection and the subnational regions in which the communities are located (at the level of provinces), excluding the interaction term with the *Initial child mortality rate* variable.

Table A 2. OLS regression model on the relationship between community change in education and community initial inequality in education in rural Africa.

Independent variable	Beta coefficient	Standard error
Intercept	3.696***	0.132

Initial inequality in education	-0.666***	0.099
Initial education	-0.545***	0.009
Initial birth rate	-0.004	0.011
Initial child mortality rate	-1.348***	0.138
Dummy for missing data on childbirths	-0.306***	0.019
Number of observations	24952	
Number of subnational regions	399	
Adjusted R-squared	0.426	

****p<0.001, **p<0.01, *p<0.05.*

Table A 3. OLS regression model on the relationship between community change in women's height and community initial inequality in women's height in rural Africa.

Independent variable	Beta coefficient	Standard error
Intercept	1266.313***	21.317
Initial inequality in women's height	67.844	54.033
Initial women's height	-0.802***	0.013
Initial birth rate	-0.065	0.368
Initial child mortality rate	-21.829***	4.507
Dummy for missing data on childbirths	2.367	2.294
Number of observations	4429	
Number of subnational regions	308	
Adjusted R-squared	0.509	

****p<0.001, **p<0.01, *p<0.05.*

Table A 4. OLS regression model on the relationship between community change in women's age of marriage and community initial inequality in women's age of marriage in rural Africa.

Independent variable	Beta coefficient	Standard error
Intercept	15.785***	0.180
Initial inequality in women's age of marriage	-1.494***	0.302
Initial women's age of marriage	-0.896***	0.006
Initial birth rate	-0.029**	0.011
Initial child mortality rate	-1.021***	0.135
Dummy for missing data on childbirths	0.852**	0.303
Number of observations	11711	
Number of subnational regions	378	
Adjusted R-squared	0.744	

****p<0.001, **p<0.01, *p<0.05.*

Table A 5. OLS regression models on the relationships between community age difference between spouses and both community change in education and community change in women's age of marriage in rural Africa.

	Change in education		Change in women's age of marriage	
Independent variable	Beta coefficient	S.E.	Beta coefficient	S.E.
Intercept	1.804***	0.181	15.396***	0.228

Initial age difference between spouses	-0.005	0.004	-0.008*	0.004
Initial education	-0.33***	0.011	—	—
Initial women's age of marriage	—	—	-0.893***	0.007
Initial birth rate	0.032*	0.015	-0.009	0.014
Initial child mortality rate	-1.037***	0.181	-0.815***	0.169
Number of observations	9085		7881	
Number of subnational regions	387		371	
Adjusted R-squared	0.372		0.734	

****p<0.001, **p<0.01, *p<0.05.*

Table A 6. OLS regression models on the relationships between community ethnic fractionalization and both community change in education and community change in women's age of marriage in rural Africa.

	Change in education		Change in women's age of marriage	
Independent variable	Beta coefficient	S.E.	Beta coefficient	S.E.
Intercept	2.201***	0.111	15.9***	0.201
Initial ethnic fractionalization	0.084	0.049	-0.097	0.067
Initial education	-0.478***	0.006	—	—
Initial women's age of marriage	—	—	-0.905***	0.006
Initial birth rate	-0.014	0.011	-0.032*	0.012
Initial child mortality rate	-1.267***	0.129	-1.115***	0.149
Dummy for missing data on childbirths	-0.297***	0.019	0.874**	0.306
Number of observations	23464		9767	
Number of subnational regions	373		350	
Adjusted R-squared	0.426		0.736	

****p<0.001, **p<0.01, *p<0.05.*

Table A 7. OLS regression models on the relationships between community religious fractionalization and both community change in education and community change in women's age of marriage in rural Africa.

	Change in education		Change in women's age of marriage	
Independent variable	Beta coefficient	S.E.	Beta coefficient	S.E.
Intercept	2.103***	0.112	15.738***	0.197
Initial religious fractionalization	0.333***	0.050	-0.038	0.071
Initial education	-0.47***	0.006	—	—
Initial women's age of marriage	—	—	-0.901***	0.006
Initial birth rate	-0.025*	0.011	-0.03*	0.012
Initial child mortality rate	-1.292***	0.130	-1.055***	0.146
Dummy for missing data on childbirths	-0.288***	0.019	0.539	0.311
Number of observations	23450		9985	
Number of subnational regions	379		366	
Adjusted R-squared	0.425		0.742	

****p<0.001, **p<0.01, *p<0.05.*