Diversity, semi-communication and cross-country trust: A quantitative analysis

Michael Kumove [a]

October 2019

Abstract

Existing trust research has often failed to account for the possibility that communication impairments brought on by language barriers could explain low levels of trust both within and between countries. To test whether this is the case, I construct an 'index of communication potential' for a sample of 359 cross-country dyads composed of 21 European countries. Although similar indexes have been used previously, this is the first one to include instances of 'semi-communication' between related languages when calculating communication potential. Multiple regression analysis indicated not only that greater communication potential was associated with greater cross-country trust, but that this relationship was monotonic: semi-communication was also associated with greater trust, but the link was weaker than for actually sharing a language.

[a] School of Politics and International Relations, Australian National University.
Haydon-Allen Building, 22 University Avenue, Canberra 2601, ACT, Australia.
Email: michael.kumove@anu.edu.au

Published article DOI: https://doi.org/10.1016/j.ssresearch.2019.102392

Keywords: Diversity, trust, social capital, communication, mutual intelligibility, ethnic relations

1. Introduction

Over the last two decades, a growing body of research has documented a negative relationship between ethno-cultural diversity and social trust. Not only is within-country ethnic diversity associated with lower within-country social trust (Alesina and La Ferrara, 2002; Delhey and Newton, 2006; Putnam, 2007), but cultural or ethnic differences between countries have also been identified as a source of cross-country mistrust and animosity (Delhey, 2007)

It is less clear *why* this relationship exists. Scholars have already suggested a number of reasons why diversity leads to lower trust, including perceptions of group threat, social identity theory, differences in culture, values or preferences, and in-group sanctioning effects. But one explanation remains surprisingly under-researched: cross-group language barriers. It seems likely that communication barriers both within countries and across countries could inhibit the formation of trust because, as Putnam (1993) points out, trust develops from repeated interactions which decrease participant risk and reduce incentives to defect. Given that communication barriers might prevent these from occurring, it stands to reason that this could inhibit the growth of trust.

Unfortunately, the commonly-used measure of 'linguistic fractionalisation' is a poor indicator of communication potential, and only one study (Buzasi, 2015) has attempted to test if social trust is associated with communication potential specifically. Taking my cue from her work, I constructed an 'index of communication potential' between 359 directed cross-country dyads composed of 21 European countries. This index accounts for the ability of second languages and lingua francas to reduce communication barriers. I also created a second index which counts instances where certain closely-related languages – such as Spanish and Italian, or Danish and Swedish – are similar enough to enable partial communication between their speakers. Although the first index is similar to Buzasi's, the second index is a new addition to the study of diversity and trust. No study thus far has attempted to include 'semi-communication' (Haugen, 1966) into an index of communication potential, as this one does. Both indexes were strongly and significantly associated with trust, even when included side-by-side in the same regression model. This indicates not only that communication is linked to higher trust, but also that the effect appears to depend on the degree of mutual intelligibility between the languages in question – a novel result never before demonstrated in quantitative trust research. Although these results concern cross-country trust, the findings may also be applicable to the study of within-country trust, a possibility which was previously identified by Gerritsen and Lubbers (2010).

The paper is structured as follows. First, a literature review discusses previous work on the relationship between within- and cross-country ethnic diversity, communication and trust, which is followed by a theory of how ethnic diversity can translate into reduced trust via impaired communication. Next, I describe the data and methods used to test this theory, before reporting the results of my statistical tests. The paper finishes with a discussion of the results, which considers their limitations, scholarly relevance and possible lessons for policymakers.

2. Literature review

Trust is an important ingredient for any successful society or organisation. It is one of the core components of social capital, which refers to the 'features of social organization such as networks, norms and social trust that facilitate coordination and cooperation for mutual benefit' (Putnam, 1995). Trust therefore facilitates collective action and improves the functioning of political, social or economic systems (Hooghe, 2007). But trust comes in different forms. The type which is often cited as being particularly important for maintaining within-country social cohesion is 'generalised trust', which refers to the degree of trust in unspecified strangers. In contrast, 'particularised trust' is the extent to which people trust certain known individuals and members of their own in-group (Uslaner and Conley, 2003).[1] Cross-group trust differs from both of these by examining the extent to which respondents trust a specified out-group, and is therefore particularly salient in the context of cross-country ties or ethnic relations within a single country. Cross-country trust, which is the focus of this study, can be viewed as a special case of cross-group trust.

Trust can be a difficult concept for researchers to grasp. Some authors have pointed out that trust levels sometimes seem to shift for reasons that are unclear, and it is difficult to provide anything more than ad-hoc explanations in many cases. Writing about Japan, the sociologist Ronald Dore argued that its high level of social trust was partly a *sui generis* phenomenon arising from its Confucian heritage (Fisman and Khanna, 1999). Nevertheless, other scholars have attempted to find patterns of trust that can be generalised across countries. Starting in the early 2000s, a number of studies began to report a negative relationship between ethnic diversity and within-country generalised trust. Alesina and La Ferrara (2002) found that ethnic diversity was associated with lower generalised trust in the United States, a finding which was replicated by Costa and Kahn (2003) and Stolle, Soroka and Johnston (2008). Delhey and Newton (2005) found similar evidence in a cross-country sample, and Gundelach and Tranmüller (2013) reported the same effect in Germany. Putnam (2007) made a splash by finding that ethnic diversity was related not only to lower generalised trust, but also to lower trust in one's own in-group, a finding which Putnam labelled 'constrict theory'.

Studies of cross-country trust that rely on large datasets seem to be relatively rare, but some existing research again links ethno-cultural differences to lower trust. Gerritsen and Lubbers (2010) found that cultural distance was linked to lower cross-country trust in Europe, confirming earlier results by Delhey (2007). This idea – that cultural similarities produce greater affinity and trust across groups – is known as the similarity-attraction hypothesis. As the psychologist Milton Rokeach pointed out, there is a 'natural tendency for people to associate with, socialize with and be more comfortable with others having similar belief systems', (Rokeach, 1960: 161). The similarity-attraction hypothesis was critiqued by Van Oudenhoven et al. (2002), who argued that it does not explain the existence of asymmetric cross-country trust (i.e. when people in country A trust those in country B, but those in B do not trust A). They instead find that social identity theory – favouring one's in-group and denigrating out-groups in order to elevate one's self esteem – better explains patterns of cross-

[1] Generalised and particularised trust are also known as 'thin' and 'thick' trust, respectively.

country trust. On the other hand, Kim (2007) found that cross-country trust does tend to exhibit symmetry, and if there is high trust from A to B, B is also likely to report high trust in A. He also showed that country A is more likely to trust B when: A is generally trusting; if others generally trust B; or if B is trusted by A's friends.

Despite the evidence linking diversity within and between countries to lower trust, less attention has been paid to why this relationship exists. Focusing mainly on within-country trust, Schaeffer (2014) provides a typology of some possible explanations. First, conflict theory (Blumer, 1958) posits that groups within a diverse state develop animosity towards each other as they jockey for status and resources, while the aforementioned social identity theory (Tajfel and Turner, 1986) suggests that in-group favouritism arises from the need for self-esteem. A third family of explanations points out that because people in diverse societies tend to cluster socially along ethnic lines, they are more likely to behave positively towards in-group members, who they will likely encounter again, than out-group members, who they will likely not. Habyarimana et al. (2007) found compelling evidence for this effect. Fourthly, because ethnic and cultural diversity often go hand-in-hand, a diverse state will tend to have a range of preferences and values. This could erode cross-group trust since people are aware that others may have cultural ideas that they consider immoral or offensive. This is particularly likely to occur when highly visible differences, such as dress, mark people out as members of a certain group (Schaeffer, 2014: 44). Van Oudenhoven et al. (2002) had earlier proposed a similar typology for cross-country attitudes. In their reckoning, three theories might explain cross-country attitudes, including trust: the similarity-attraction hypothesis, social identity theory (both already mentioned) and the contact hypothesis, in which cross-group trust is a function of cross-group interpersonal contact.

All of these explanations are possible and mutually compatible, and all have varying degrees of evidence behind them. But a fifth possible explanation – that communication barriers between ethnic groups or nations who speak different languages impedes the development of trust – seems to have been surprisingly neglected in the literature. Schaeffer (2014) relegates it to a short section at the end of his discussion, while Van Oudenhoven et al. (2002) mention language only as a marker of similarity that can trigger cross-national comparisons, and therefore intensify the out-group derogation predicted by social identity theory. They do not consider the idea that linguistic similarity could affect trust by increasing communication potential. That question is the focus of this study.

Although a number of studies have found a negative relationship between linguistic fractionalisation and within-country trust (Leigh, 2006; Anderson and Paskeviciute, 2006; Wang and Steiner, 2015), these results are insufficient to pinpoint communication barriers as the mechanism responsible. This is because linguistic fractionalisation, which refers to the probability that any two randomly-selected people will share a mother tongue, does not account for second languages or lingua francas, nor the possibility that two separate languages might be similar enough that their speakers could still understand each other. Spain is a good example of this shortcoming. It has a medium score (0.4132) for linguistic fractionalisation (Alesina et al, 2003) but this ignores the fact that almost everyone in the country can speak Castilian Spanish, even if their mother tongue is Basque, Catalan or something else (Fernandéz and Roth,

2006). The communication barriers between Spaniards are therefore almost zero. The same is true across countries: using only mother tongues to estimate the communication potential between the nations of the European Union would lead to inaccurate results, since this ignores Europe's many lingua francas such as English, German or Russian. By measuring only mother tongue, linguistic fractionalisation seems to be measuring something much closer to ethnic groupings instead of true communication potential.

Recognising this shortcoming, some scholars have focused on how communication potential specifically, rather than mere 'fractionalisation', is linked to trust. Buzasi (2015) constructed an index of communication potential which accounted for the total set of languages spoken by an individual, and found that this was associated with greater generalised trust within sub-national regions in a sample of countries in sub-Saharan Africa. That appears to be the only study so far which has used quantitative methods to test how communication barriers affect trust, although there is some work in the fields of management and organisational psychology which has shown how language barriers can inhibit trust and worsen cross-group attitudes, such as Thuesen (2017) and Tenzer et al. (2014). Moreover, while Buzasi's study represented an encouraging turn towards focusing on communication potential instead of using the rather crude measure of linguistic fractionalisation, it still did not account for the possibility that the degree of mutual intelligibility affects the relationship between communication and trust. Perhaps when groups speak closely-related languages which allow them to partially understand each other, like Spanish and Italian, the negative effect on trust will be reduced. This paper is the first study which integrates this idea of semi-communication into a quantitative study of communication and trust. It is also the first study which applies Buzasi's idea of an 'index of communication potential' to a cross-country research context.

3. The 'segregation' theory of communication and trust

Scholars have long sought to determine the factors which generate trust. Some researchers have pointed to historical or institutional factors, or the role of intergroup relations: the famous 'contact hypothesis' emphasises that trust and other positive attitudes arise because social contact with other groups allows people to discover common ground and overcome stereotypes (Allport, 1954). In this view, trust results from gaining more accurate information about other groups. But another popular strand of research views trust through the lens of rational choice theory, and often uses game-theoretic models to illustrate how it emerges. For instance, Putnam noted that trust is a product of the degree of social ties between the actors concerned: by playing 'repeated games' with one another, actors 'reduce incentives to defect, reduce uncertainty, and provide models for future cooperation. Trust itself is [therefore] an emergent property of the social system' (Putnam, 1993: 177). Repeated games enable actors to trust and cooperate with each other because they offer the possibility of punishing non-cooperators in future encounters, something not possible in single-shot or finitely-repeated games (Ostrom, 2003: 23-24). This is much the same phenomenon as was found by Habyarimana et al. (2007), where the potential to be reprimanded in future encounters led people to behave more positively towards in-group members.

It therefore makes sense that when there are a large number of ongoing social ties between two groups, this should lead to a higher level of cross-group trust. These ties could be between ethnic groups within a single state, or cross-country ties between residents of different states. The next question is whether (and how) communication barriers inhibit the formation of those cross-group ties. Buzasi (2015) has already pointed out that communication enables the formation of trust because it permits repeated interaction. The proposed mechanism linking diversity to low trust therefore works as follows: first, assuming the two groups speak different languages, the communication barriers create a degree of social segregation between groups. Lacking a way to communicate, the two groups will each have a high density of in-group ties but few cross-group ones. Over time, the effect will be that people tend to have many more trusted in-group members than out-group members, and because people are prone to using heuristics such as 'representativeness', this lower personal trust will spill over into lower trust towards the out-group as a whole (Tversky and Kahneman, 1974; Buzasi, 2015). Depending on the salience of out-group perceptions when evaluating trust in 'people in general', this low level of cross-group trust is likely to lead also to lower generalised trust. The potential for social segregation to inhibit trust also fits well with existing research showing that physical segregation has the same effect (Uslaner, 2010; Robinson, 2017).

This paper's main theoretical contribution is to refine the above model by moving beyond a binary understanding of language differences. Previous work, including Buzasi (2015), has assumed that people either speak the same language, and can therefore communicate, or don't speak the same language, and can't. But the reality is often more complicated than this. Numerous language pairings are similar enough that monolingual speakers can still have a conversation with each other, although perhaps with some straining: Danish and Swedish, Spanish and Italian, Russian and Ukrainian. It would be incorrect to assume that communication across these divides is zero. To account for these instances of 'semi-communication' (Haugen, 1966), I instead propose a three-level categorisation for spoken mutual intelligibility:

1: Full intelligibility: e.g. American English and British English, Serbian and Croatian, Swedish and Norwegian

2: Semi-communication: e.g. Spanish and Italian, Russian and Ukrainian, German and Yiddish

3: No intelligibility: e.g. English and Chinese, Spanish and Arabic, most other language pairings

This is a simple model. It does not account for the full spectrum of mutual intelligibility, and nor does it account for the possibility that one's ability to understand a closely-related language might increase with exposure to it. Nevertheless, it represents a first step towards modelling mutual intelligibility when analysing the relationship between diversity, communication and trust. This is theoretically relevant because if communication is positively related to trust, then it follows that the relationship may be monotonic: trust increases at each step as we move from no intelligibility, to semi-communication, to full intelligibility. Introducing new theoretical expectations that are consistent with the implications of existing theories also forms part of

what Imre Lakatos described as 'progressive' science. In the Lakatosian philosophy of science, successive theories should introduce both new theoretical implications and verify those implications empirically, thus generating 'novel facts' (Musgrave and Pidgen, 2016).

This study aims to fulfil both of those criteria. It would be logically consistent with Buzasi (2015) if communication and trust were monotonically related, and testing that allows this study to potentially produce a 'novel fact' and would provide additional validation for the theory that communication contributes to trust. Another possibility is that the relationship is non-monotonic, and semi-communication actually leads to the lowest trust of all. It is impossible to tell without a way to measure mutual intelligibility.

I theorise that mutual intelligibility and trust are positively and monotonically related.[2] This is likely to be the case because increases in intelligibility make possible new and more complicated forms of social ties. Two ethnic groups which cannot communicate verbally will find that almost all types of cross-group ties are impossible, and because repeated interaction is therefore extremely limited, trust will be at a commensurately low level. Two groups which can enjoy semi-communication might be able to form some kinds of cross-group ties, although the type of ties possible will still be limited. Playing sports together or chatting in the street might be possible, but more complex forms of interaction might be off-limits. Nevertheless, the increased potential for cross-group interaction should translate into greater cross-group trust. Finally, with full intelligibility there should be no limits on the types of social ties possible. The quantity of repeated interactions should therefore be greater, and cross-group trust should reach its highest level in this situation.

If correct, this theory should produce the following observable implications:

H_1: Full intelligibility should be positively associated with cross-group trust.

This means that sharing a language should be associated with greater trust than not sharing a language. The shared language index, which will be introduced below and measures the probability that two people have at least one language in common (and therefore can enjoy full intelligibility), should show a positive and statistically significant association with trust.

H_2: Semi-communication should be positively associated with cross-group trust, but the association should be weaker than for full intelligibility.

If the relationship is really monotonic, then semi-communication should also be positively associated with trust, but its ability to generate trust will be weaker than full intelligibility. The semi-communication index (also introduced below) which measures the probability that two people are able to engage in semi-communication (but nothing else), should show a positive and statistically significant association with trust, but the magnitude of this association should be smaller than for the shared language index.

[2] By monotonic, I mean that an increase in the level of communication should always generate some increase in trust, even though the magnitude of that increase may differ across the domain (i.e., the relation may be non-linear). Formally, this means that the first-order derivative of the function describing the relationship between communication and trust should be positive for all values of communication. This is technically known as "strict" monotonicity.

H_3: Structural equation modelling should reveal a positive indirect effect of communication on trust through the channel of cross-group contact.

If the relationship between communication and trust is mediated to some extent by cross-group contact, then a structural equation model should show a positive indirect relationship between communication and trust via that channel. However, I do not necessarily expect that mediation will be total: there may also be a direct effect of communication on trust.

4. Data and methods

This paper uses multiple regression and some bivariate statistical techniques in order to test the theory described above. The analysis is based on 359 cross-country dyads composed of 21 European countries. The 21 countries are those which have both (1) data on trust from the European Election Study 2004 as well as (2) language data from Eurobarometer 64.3. Those are:

Austria	*Belgium*	*Czech Republic*	*Denmark*	*Estonia*	*Finland*	*France*
Greece	*Hungary*	*Ireland*	*Italy*	*Latvia*	*Lithuania*	*Luxembourg*
Netherlands	*Poland*	*Portugal*	*Slovakia*	*Slovenia*	*Spain*	*Sweden*

Although 21 countries should produce a maximum possible number of 420 dyads, three countries (Belgium, Lithuania and Sweden) did not have data available on their trust levels toward the other countries, and one other dyad (Hungary-Portugal) was missing this data as well. This leaves 359 'directed dyads', each of which consists of a 'truster' and a 'trustee'. The level of trust attached to that dyad refers to how much the truster trusts the trustee, not the other way around. When I write the dyads with a hyphen, the truster is always listed first, so the dyad Latvia-Finland is concerned with how much Latvia trusts Finland. That level of trust will be different to the dyad Finland-Latvia.

Independent variable: communication

As noted above, many previous studies have failed to really measure communication potential and have instead measured something much closer to ethnic groupings. To overcome this problem, we need to construct an 'index of communication potential', as recommended by Laitin (2000) based on work by Greenberg (1956) and later used by Buzasi (2015). To create this index, we divide the population up not by their mother tongues, but by their linguistic repertoires, so those who are monolingual English speakers are taken as a different group to those who speak both English and Spanish. If a country has three languages A, B and C, then there are seven possible groups:

- monolingual A speakers
- monolingual B speakers
- monolingual C speakers

- those who speak A and B
- those who speak A and C
- those who speak B and C
- those who speak A, B and C

Buzasi and Laitin are referring to calculating within-country communication potential. To generate a within-country measure of communication potential, we first determine the percentage of the population of each of those seven groups and then find the product of each of those groups with each of the other. We then sum up all the products where communication is possible - so we include the product of A+B and A in the sum, but not B+C and A, since in the latter case there is no language in common (Laitin, 2000: 149).

The challenge here is to adapt this method for cross-country use. Instead of finding the products of each group with each of the others in the same country, I instead found the products of each group with each group in the *other* country. These were then summed to generate a score for cross-country communication potential. I obtained language data from Eurobarometer 64.3, which was conducted in 2005 and asks Europeans about their native language as well as any other languages they speak well. In order to simplify the calculations, I took only the first additional language mentioned by each person.[3] I also ignored mother tongue groupings with less than 10 respondents, and instead placed these into a single 'Other' category.

This allowed me to generate language repertoire groups for each country including residents' native tongues as well as their first additional language. These are just like the A, B and C groups in Laitin's example. For instance, in Austria 959 people gave German as their mother tongue, and 43 spoke 'Other' mother tongues. Of the 959 whose mother tongue was German, 539 also spoke English, another 14 spoke French and a further 11 spoke Italian, while of the 43 who spoke 'Other', 35 spoke German. This means there are six groups of people in Austria:

- Those who speak only German (39.4%)
- Those who speak German and English (53.8%)
- Those who speak German and French (1.4%)
- Those who speak German and Italian (1.1%)
- Those who speak 'Other' (0.8%)
- Those who speak 'Other' and German (3.5%)

By generating these lists of language repertoires for each of the 21 countries and then comparing them against each other, I was able to determine the probability that a randomly selected person in the first country will share at least one language with a randomly selected person in the second country. These scores are referred to as the **shared language index**. Further detail on how these scores were calculated is contained in Appendix A.

[3] The first language mentioned is also likely to be the one in which the respondent is most proficient, so this has the additional advantage of avoiding counting languages in which the respondent may have only a middling level of proficiency.

But the process does not end there. Since I am interested in the possibility that the degree of mutual intelligibility between languages may affect trust, I need to find a way to account for this in the analysis. In particular, I want to know whether the 'semi-communication' between languages like Spanish and Italian might be linked to higher trust than if there was no communication at all. To do this, I took all of the languages in my sample and identified the pairs where there is sufficient mutual intelligibility for semi-communication to occur. Information on how this was done is contained in Appendix B, but it relied heavily on Gooskens et al.'s (2018) study of mutual intelligibility. I also relied on testimonies from native speakers and language learners accessed online. Through this method I derived the following list of language pairs in the 21 countries which allow for semi-communication:

Catalan-Spanish	*Catalan-Italian*	*Galician-Portuguese*	*Galician-Spanish*	*Italian-Spanish*
Portuguese-Spanish	*Croatian – Slovenian*	*Czech-Slovak*	*Polish-Slovak*	*Danish-Swedish*

For dyads where any of these pairings occur, I calculated the sum of all the encounters where semi-communication occurs. This is a similar process to that used for the shared language index: the only difference is that in this case, I am summing up the semi-communication encounters rather than those where there is a shared language. These scores, which refer to the probability that two randomly-selected people from across the dyad will be able to engage in semi-communication, are referred to as the **semi-communication index.** Most dyads (297 of the 359) did not allow for any semi-communication. However, in the dyads where it was possible, the probability of semi-communication was sometimes quite large and represented a vast change over and above the shared language index. The 'shared language index' score for Italy-Spain was 0.064, but the probability of semi-communication was 0.846. To reiterate, the inclusion of semi-communication means that we can now construct two separate indexes of communication potential:

- The **shared language index**, which measures the probability that any two randomly-selected people from each country have at least one language in common. This is similar to the index used by Buzasi (2015), although this index is between-country rather than within-country.
- The **semi-communication index**, which measures the probability that any two randomly-selected people from each country are able to engage in semi-communication (e.g. one of them speaks Spanish and the other Italian).

The development of the semi-communication index represents an innovation in the quantitative study of communication potential. Buzasi's study included only a shared language index. By accounting for semi-communication, I hope to be able to move towards a much more accurate understanding of how mutual intelligibility affects trust. To see how this all works, let's take another look at the Italy-Spain dyad referred to above. Its score on the shared language index is 0.064, indicating there is only a 6.4% chance that a randomly-selected Italian and a randomly-selected Spaniard will have at least one language in common. However, there is a 0.846 probability that they will be able to engage in some form of semi-communication, for

instance if one speaks Spanish and the other Italian, and this is its score on the semi-communication index. The descriptive statistics for the two indexes are presented in Table 1 below. A final point to note is that the scores on both communication indexes are non-directed. That is, the communication scores for the dyad Finland-Latvia will be exactly the same as for Latvia-Finland. This is in contrast to the trust scores: as mentioned above, the trust level for Latvia-Finland will be different to Finland-Latvia.

Dependent variable: cross-country trust

Cross-country trust data was obtained from the 2004 European Election Study. While newer data would have been preferable, this appears to be the most recent dataset available on trust between European countries and is the same source used by Gerritsen and Lubbers (2010) in their analysis of European cross-country trust. The survey asked respondents in various countries how much they trust specified other European nationalities ("How much trust do you have in Danes? How much trust do you have in Poles?" and so on). Three answers were possible: a lot of trust, not much trust, or don't know/not applicable. For each of the dyads, I calculated the level of trust held by the truster towards the trustee. This was obtained by first removing all the "don't know/not applicable" answers from the respondents in the truster country. Then, the total number of people responding that they had "a lot of trust" in the trustee country was divided by that figure:

$$Trust\ score = \frac{Number\ responding\ "A\ lot\ of\ trust"}{(Total\ responses - number\ of\ "DK/NA"\ responses)}$$

The resulting trust scores can range from 0 to 1. This is cross-group trust, rather than generalised trust, although as discussed above the former is likely to affect the latter. The descriptive statistics are presented in Table 1 below.

Control variables

Two major competing explanations for low trust also need to be controlled for. First among these is cultural distance between groups. This was measured using three different specifications:

- First, I used ***ethnic similarity*** as a proxy for cultural similarity by developing an 'ethnic similarity index' for each dyad. I obtained data on ethnic groupings from Alesina et al (2003). Then, using a similar procedure as with the communication indexes, I identified whether there were any shared ethnic groups across the countries and summed up the probabilities of those encounters occurring. 117 of the dyads had some degree of ethnic similarity, while the other 242 had scores of zero.
- The second measure of cultural distance was whether the two countries in the dyad have a ***shared religious background***. This is taken to be the religion dominant in the country (Catholicism, Protestantism or Orthodox) immediately prior to World War Two. Gerritsen and Lubbers (2010) also used this technique to measure cultural

distance. This is a dummy variable, coded as 1 if the two countries share their religious background and 0 otherwise.

- The third measure was whether the two countries were from the ***same "cultural region"*** of Europe. The European Union encyclopedia, Eurovoc, divides Europe into four cultural regions: Western Europe, Northern Europe, Southern Europe and Central/Eastern Europe. This dummy variable is coded as 1 if the two countries are from the same region, and 0 if not.

The second major control is for the level of contact between groups. This was measured using five different specifications:

- I used Eurostat to obtain figures for the ***amount of tourism*** from the truster to the trustee country. I used 2004 figures wherever possible, but where these were unavailable I used the figure from the closest available year. I divided the total number of tourist trips by the population of the truster country, to provide an estimate for the proportion of the truster country which visited the trustee in that year.
- ***Distance between capitals*** was obtained from http://tjpeiffer.com/crowflies.html. This provides the distances in miles which I then converted into kilometres for use in my dataset. All distances are as the crow flies.
- I coded a dyad as ***"crossing the Iron Curtain"*** if one (but not both) of the countries involved was formerly communist. Among the 21 countries the following were formerly communist: the Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Slovakia and Slovenia. This is taken to be an indicator of contact as the level of contact and travel between communist and non-communist European states was very limited prior to 1989.
- The ***value of exports from the truster to the trustee*** was obtained from Eurostat DS-018995 and measures the value, in billions of Euros, of the exports from the truster country to the trustee country during 2004.
- Finally, I obtained scores for the ***familiarity with the trustee country*** by taking the number of "don't know/not applicable" answers given by respondents in the truster country, and dividing that by the total number of respondents. This is the same technique used by Gerritsen and Lubbers (2010) to control for cross-country contact and the resulting scores can range from 0 to 1. As the proportion of "don't know/not applicable" answers actually measures unfamiliarity, I converted this to familiarity by calculating *familiarity = 1 – (level of unfamiliarity)*. The lowest level was from Italy to Lithuania (0.398) and the highest was France to Belgium (0.985).

Country fixed effects dummies were used for both the truster and trustee countries. These capture country-specific variables which may be linked to both communication and trust, such as economic development, corruption, inequality, and a country's underlying propensity to trust.

Descriptive statistics

The descriptive statistics for all variables are presented in Table 1 below. Note that all have been rounded to three decimal places.[4]

Table 1: *Descriptive statistics for all variables*

	Mean	Median	Standard Deviation	Minimum	Maximum
Shared language index	0.154	0.108	0.141	0.012	0.743
Semi-communication index	0.023	0.000	0.124	0.000	0.879
Cross-country trust	0.636	0.651	0.172	0.216	0.976
Shared ethnicity index	0.004	0.000	0.016	0.000	0.124
Shared religious background	0.501	1.000	0.501	0.000	1.000
Shared cultural region	0.214	0.000	0.411	0.000	1.000
Amount of tourism	0.024	0.007	0.059	0.000	0.703
Distance between capitals (1000 kms)	1.330	1.251	0.714	0.055	3.363
Dyad crosses Iron Curtain	0.493	0.000	0.501	0.000	1.000
Value of exports (billion €)	1.858	0.308	4.864	0.000	35.230
Familiarity with trustee	0.738	0.744	0.135	0.398	0.985

The semi-communication index has a median of zero, because the majority of dyads do not actually have any of the eligible language pairs which allow for semi-communication. This also explains why its mean is far lower than for the shared language index. As for the dependent variable, the lowest level of trust was Portugal's trust towards Lithuania (0.216) and the highest was Denmark's trust in Sweden (0.976). However, most dyads had a level of trust in the mid-range, between about 0.4 and 0.8.

Histograms for both communication indexes and the trust scores are displayed below. Because of the skewed nature of the data, I also include histograms of the natural log of each communication index. The semi-communication index takes a value of zero for most dyads, and the log of zero cannot be computed. A token small value (0.000000001) was added to the dyads with semi-communication scores to allow the logarithm of these to be calculated.

[4] I thank an anonymous reviewer for their suggestions regarding the formatting and presentation of the variables and descriptive statistics.

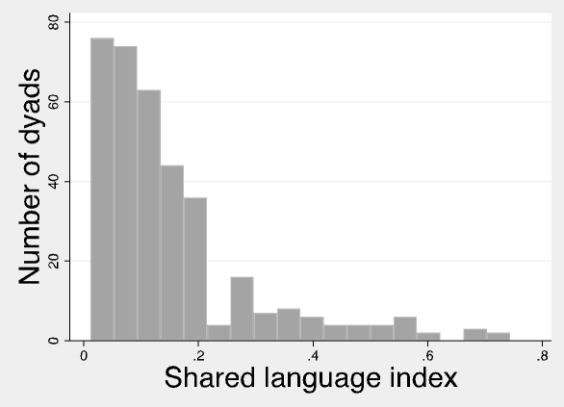

Figure 1: *Histogram of SLI scores*

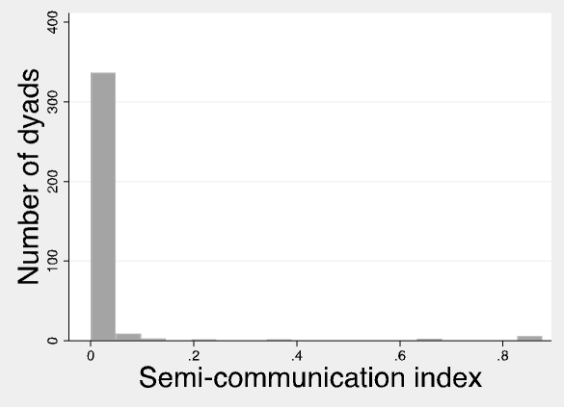

Figure 2: *Histogram of SCI score*

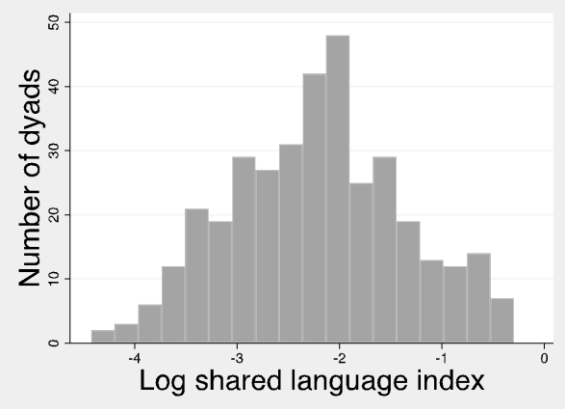

Figure 3: *Histogram of logged SLI scores*

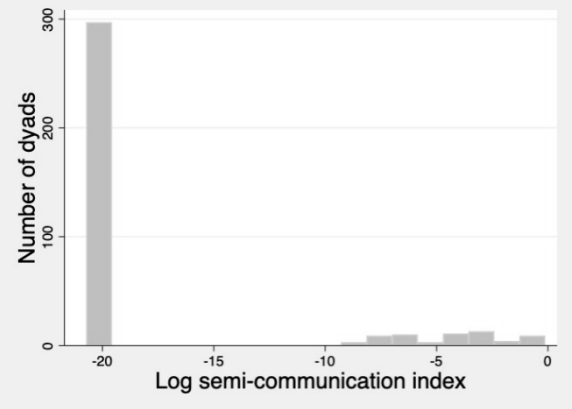

Figure 4: *Histogram of logged SCI scores*

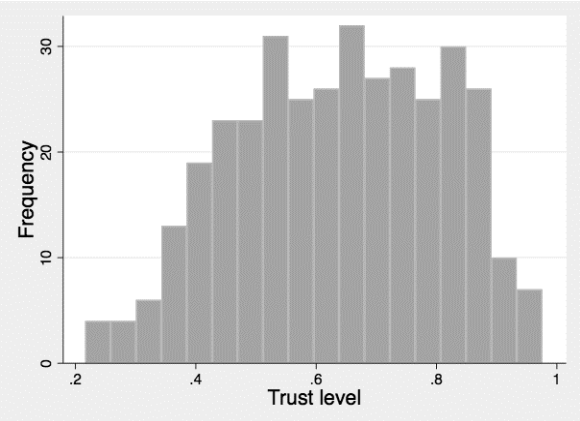

Figure 5: *Histogram of dyad trust scores*

5. Results

The theory outlined above predicts that there should be a positive and monotonic relationship between communication potential and trust. This means that both the shared language index and the semi-communication index should be positively associated with trust, but the strength of that association should be weaker for the semi-communication index. An initial set of bivariate analyses seem to broadly align with those expectations. The shared language index showed a pairwise correlation of 0.339 with cross-country trust, while the semi-communication

index showed 0.138. Scatterplots illustrating these relationships are displayed below. Scatterplots of each index's log transformation are also included.

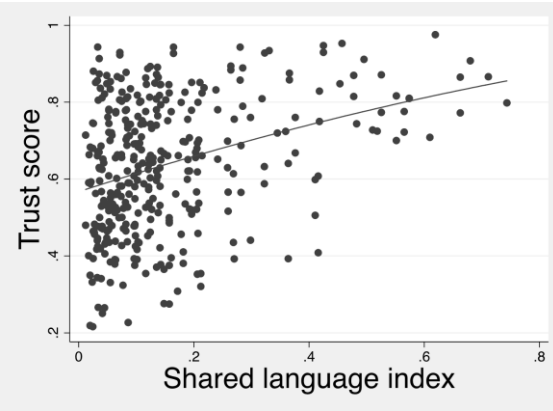

Figure 6: *SLI and trust*

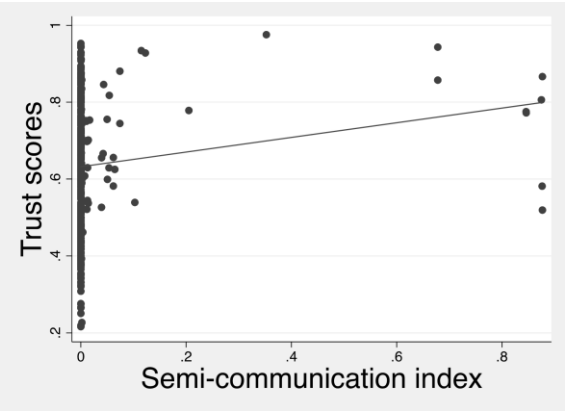

Figure 7: *SCI and trust*

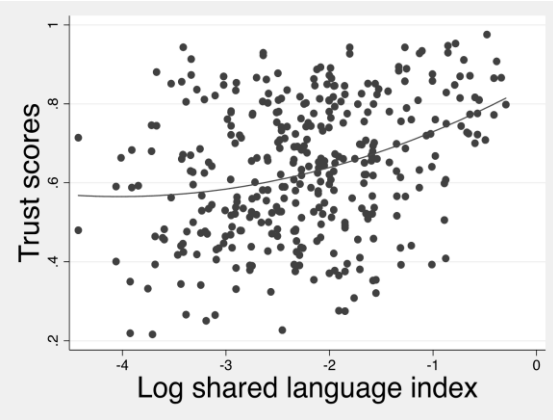

Figure 8: *Logged SLI scores and trust*

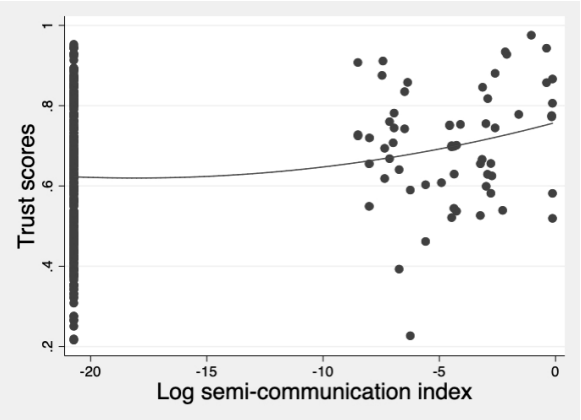

Figure 9: *Logged SCI scores and trust*

The relationship is not completely linear. While having high communication potential appears to equate to high trust, low communication potential does not necessarily equate to low trust. It is also important to be cautious because the number of high-communication, high-trust dyads is relatively small, raising the possibility that a small number of dyads are exerting excessive leverage on the trend line through random chance.

I turn next to multivariable analysis. Multiple regression was used to measure the strength and significance of the relationship between communication and trust while controlling for a range of other variables. The results are contained in Table 2. In model (1), I ran a simple ordinary least squares regression using shared language index scores as the independent variable and controlling only for country fixed effects. The index was significant at the 0.001 level with a β-coefficient of 0.312. Then I began adding controls. In model (2), I added the controls for cultural distance: shared cultural region, shared religious background, and the shared ethnicity index. This is a major competing explanation for why diversity may reduce trust and overcoming it would provide a boost for my theory. It appears to have passed that hurdle here: the shared language index remained significant at the 0.001 level even though one cultural distance measure, shared cultural region, also reached statistical significance. In model (3), I introduced the controls for cross-group contact. As expected, the index showed a sizeable drop in its coefficient as well as a lower level of statistical significance. Since the theory specifies

that cross-group contact is the mediator, it makes sense that after controlling for contact, the effect of communication would be reduced.

Table 2: *Dependent variable is cross-country trust*

Independent variable	(1)	(2)	(3)	(4)	(5)	(6)
Shared language index	0.312*** (0.044)	0.208*** (0.049)	0.167** (0.051)	0.278*** (0.042)	0.196*** (0.048)	0.167** (0.051)
Semi-communication index				0.195*** (0.038)	0.145*** (0.039)	0.090* (0.040)
Truster and trustee fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Cultural distance controls		Yes	Yes		Yes	Yes
Cross-group contact controls			Yes			Yes
R^2	0.785	0.805	0.834	0.801	0.813	0.834

Notes: standard errors in parentheses. * indicates $p < 0.05$, ** is $p < 0.01$, *** is $p < 0.001$. N=359 except for models (3) and (6) where N=315, because tourism data does not include data for 44 dyads.

Models (1), (2) and (3) are similar to Buzasi (2015), but are novel in that cross-country communication and trust are used instead of within-country measures, and the region being analysed is Europe rather than sub-Saharan Africa. The results support H_1, which supposes that sharing at least one language will be associated with higher trust, as Buzasi had previously found. But this paper's real theoretical contribution is the idea that semi-communication may also be associated with trust, albeit to a weaker degree than full communication (i.e. sharing a language). This would make the relationship between communication potential and trust monotonic, and in Lakatosian parlance would constitute a 'novel fact'. This implication is contained in H_2.

Models (4), (5) and (6) are used to test H_2. This hypothesis suggests that semi-communication has its own impact on trust above and beyond that of full communication, but also that is the size of that relationship should be smaller than for full communication. This appears to be confirmed by the results. When the semi-communication index was placed alongside the shared language index in model (4), both indexes reached statistical significance, and remained significant after adding the controls for cultural distance and cross-country contact in models (5) and (6). This indicates that semi-communication does indeed have a positive relationship with trust independent of that which already exists from sharing a language. Critically, the semi-communication index had smaller coefficients than the shared language index, indicating that semi-communication is effective at generating trust, but less so than full intelligibility. This suggests a monotonic relationship between trust and the level of mutual intelligibility, and is exactly what my theory and hypotheses predicted.

H_3 supposes that the relationship between communication and trust is mediated by cross-group contact. To test this, I generated a structural equation model (SEM) in order to examine the direct and indirect effects of communication on trust. However, including all five cross-group contact variables would make for an overly complicated model. To solve this problem I used principal component analysis to extract the first principal component from those five variables, and then used the first principal component in the SEM (where it is labelled as 'cross-group contact'). The SEM results are displayed in Figure 10 below. They provide clear evidence that the relationship is partially mediated by cross-group contact: both the shared language index and semi-communication index have positive and highly significant indirect effects on trust through the channel of cross-group contact. The direct effects of communication are smaller, and only the shared language index has a significant direct effect on trust. Furthermore, the coefficient for the shared language index on cross-group contact remains higher than that for the semi-communication index, which is consistent with the idea that full communication is more effective than semi-communication at facilitating repeated interactions.

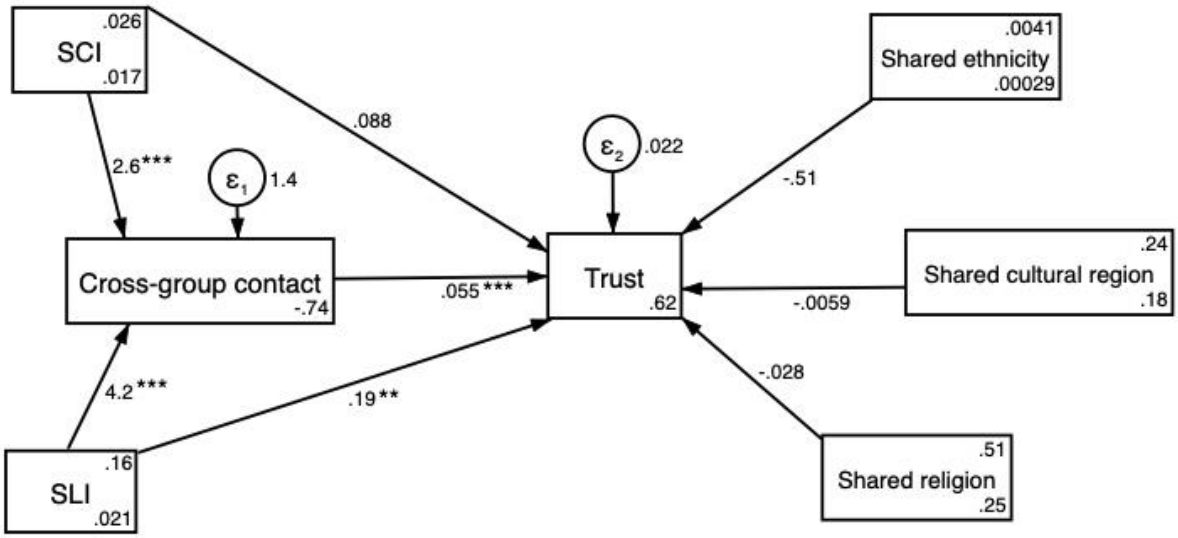

Figure 10: Structural equation model. SCI refers to semi-communication index, SLI is shared language index. * indicates $p < 0.05$, ** is $p < 0.01$, *** is $p < 0.001$.

6. Robustness checks

A number of robustness checks were used to confirm that these results are accurate. First, I ran the same regressions contained in Table 2, but used the natural log of the two indexes in place of the originals. A token small value (0.000000001) was added to the dyads with semi-communication scores of zero in order to enable the logarithm of these to be calculated. The results were very similar, with the indexes still reaching statistical significance even after controlling for cultural distance, and once again they both showed a drop in the level of significance and in the size of their coefficients after controlling for the level of cross-country contact. These results are contained in Table 4, and appear to confirm that the link between communication and trust is genuine, rather than simply an artifact arising from the skewed nature of the independent variable.

Secondly, all regressions were run again using an alternative measure of trust. In this measure, the trust level was calculated by dividing the number of people expressing 'a lot of trust' by the total number of respondents, without subtracting the 'don't know/not applicable' answers. Since some countries might be more likely to shy away from saying they distrust a group, and instead might simply respond 'I don't know', using a second measure which includes this in the denominator provides a way of avoiding this bias. The results were very similar, with the two indexes reaching statistical significance in all the same models. The only difference was that the size of the coefficients tended to be larger when the alternative trust measure was used. I also tested for sensitivity to outliers by re-running the regressions after removing all observations with studentised residuals above 2.0 or below -2.0. This did not appear to change the results in any significant way. Finally, variance inflation factor (VIF) analysis was used to confirm that multicollinearity was not present among the independent variables.

Table 4: *Dependent variable is cross-country trust*

Independent variable	(1)	(2)	(3)	(4)	(5)	(6)
Log shared language index	0.052*** (0.008)	0.032** (0.010)	0.011 (0.012)	0.043*** (0.008)	0.027** (0.010)	0.011 (0.012)
Log semi-communication index				0.003*** (0.001)	0.002* (0.001)	0.001 (0.001)
Truster and trustee fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Cultural distance controls		Yes	Yes		Yes	Yes
Cross-group contact controls			Yes			Yes
R^2	0.778	0.801	0.827	0.790	0.804	0.829

Notes: standard errors in parentheses. * indicates $p < 0.05$, ** is $p < 0.01$, *** is $p < 0.001$. N=359 except for models (3) and (6) where N=315, because tourism data does not include data for 44 dyads.

7. Discussion

These results show strong support for all three of the hypotheses:

H_1: Full communication potential, arising from sharing at least one language, is an important contributor to trust (the Buzasi hypothesis)

H_2: Semi-communication is also positively associated with trust, although less strongly than when there is a shared language (the monotonicity hypothesis)

H_3: The structural equation model should show a positive indirect effect of communication on trust via cross-group contact (the contact-as-mediator hypothesis)

While Buzasi (2015) had previously demonstrated a link between communication and trust, this study is novel for several reasons. First, Buzasi's study tested only the first of the three bullet points above. It did not attempt to measure the effect of semi-communication or include any explicit test of cross-group contact as a mediator. Secondly, Buzasi's study measured within-country generalised trust rather than cross-country trust, and it did so only among

countries in sub-Saharan Africa. This study, which focuses on Europe, is the first attempt to test quantitatively for an association between communication and trust between high-income countries. The fact that the shared language index reached statistical significance is strongly supportive of H_1: communication potential does have a positive relationship with trust, even after controlling for a range of alternative explanations such as cultural distance and country fixed effects such as economic development. This aligns well with Buzasi's earlier findings.

But this study's most important theoretical contribution is to introduce the idea that semi-communication might have its own impact on trust separate to that which arises from actually sharing a language. I hypothesised that the relationship between communication potential and trust is monotonic: all else equal, the level of trust will increase as groups move from no communication, to semi-communication, to full communication. The evidence here seems to support that hypothesis. The semi-communication index showed a strong and significant positive relationship with trust, even when included alongside the shared language index. Crucially, it also showed smaller coefficients than the shared language index. This is consistent with the notion that semi-communication generates trust but does so less effectively than full intelligibility, and implies a monotonic relationship of the type just described. This offers strong support for H_2. It also constitutes a 'novel fact' consistent with Buzasi's (2015) existing theory and findings, and allows for much stronger inference about the role of communication because a monotonic relationship is less likely to be confounded by other factors.

H_3 tests whether the communication-trust relationship is mediated by cross-group contact. The results also offer support for this hypothesis. The size of the coefficients and the levels of statistical significance in Table 2 declined after controlling for cross-group contact, while standard errors tended to increase. If communication does affect trust through the channel of cross-group social ties, it makes sense that after controlling for cross-group contact, the independent effect of communication would be reduced. More importantly, the structural equation model (SEM) provided clear evidence that the relationship between communication and trust is mediated by cross-group contact. The SEM results showed that both communication indexes had a positive and significant association with cross-group contact, which in turn had a positive and significant association with trust. Just as we would expect, the shared language index had a stronger association with contact than the semi-communication index. This suggests that communication has an indirect effect on trust through the channel of cross-group contact, and is consistent with the existence of partial mediation.

Questions might be asked about the external validity of these results. This study measures the impact of cross-country communication on cross-country trust. It is less clear whether these results tell us anything interesting about the determinants of within-country trust, with which most of the trust literature has tended to concern itself. However, I think there are good reasons to expect that these results may apply to the study of within-country trust as well. The countries which make up my sample are all members of the EU, a highly-integrated confederation of polities with a shared currency, freedom of movement and a shared justice system which exhibits a kind of 'pooled sovereignty' (Keohane, 2002). With its many different ethnicities, integrated economy and free movement, we might view the EU as analogous to a large state which contains multiple ethno-linguistic groups, such as India, Indonesia, the Philippines, and

so on, and there is not necessarily any reason to expect that conclusions about trust between EU nationalities would not also apply to the study of trust between ethnic groups in a single, diverse state. The fact that the EU allows for free movement within its borders would seem to be particularly important for this kind of external validity, as this allows EU citizens to move between countries and therefore experience cross-group contact across linguistic divides, much as they can within a single country. Some authors, like Gerritsen and Lubbers (2010), have already used cross-country evidence to contribute to debates in the within-country trust literature.

The data and methods used in this analysis also have some limitations. The first of these is data quality. As I touched on above, some studies of linguistic diversity fail to capture the true dynamics of communication potential because they focus only on mother tongue groupings. While this study overcomes that problem by accounting for second languages and lingua francas, the communication indexes I developed here still may not always truly reflect communication potential. Because I only counted one second language from each Eurobarometer survey respondent, it is likely that my index scores still underestimate true communication potential, since in this system trilingual or quadrilingual people could only have a maximum of two languages counted. If some countries have more trilingual or quadrilingual people than others (the Netherlands is a possible example), this may lead to systematic bias where such countries suffer greater under-counting of their communication potential compared to countries where few people are trilingual or quadrilingual. Future researchers may wish to include respondents' third or fourth languages if time and resources permit.

A second data quality issue is the problem of how to distinguish two different languages from one another. The question of what constitutes a separate 'language' has long proven difficult to answer: for instance, the mutually-intelligible Swedish and Norwegian are considered different languages but the mutually-unintelligible Mandarin and Cantonese are considered by their speakers to simply be 'dialects' of Chinese. Many linguists instead point defeatedly to the aphorism that 'a language is a dialect with an army and a navy' (Rajagopalan, 2001).

My adjustments for linguistic distance help to ameliorate this problem by moving the analysis away from a binary understanding of language differences. However, the fact that the Eurobarometer survey asks respondents about what language they consider themselves to speak opens the door to subjective definitions of language boundaries which may not correspond to communication divides. Italy is a good example here. Not one single Italian respondent reported that they spoke an Italian regional language like Neapolitan, Ligurian or Sicilian. Instead almost all of them reported that their mother tongue was 'Italian', which is in line with the common Italian practice of treating those regional languages as mere dialects. This means that my data is under-counting linguistic repertoires in countries like Italy. This could be problematic if I was conducting a within-country analysis, but since this chapter focuses on cross-country communication, and Italian regional languages are not spoken in other European countries anyway, it is doubtful that their absence from the data has affected the communication index scores used here.

Data quality aside, the communication indexes also exclude a number of real-world elements. First, no consideration is made of who is speaking 'whose' language. As Nelson Mandela said, 'when you talk to a man in a language he understands, that goes to his head. When you talk to him in his language, that goes to his heart'. It is plausible that one might feel greater affinity and trust towards an out-group if that group has learned your language, rather than if you have learned theirs. This was not accounted for in the analysis. Second, the way semi-communication was modelled here does not account for asymmetries: if two languages are coded as partly intelligible, the level of intelligibility is assumed to be the same in both 'directions'. In reality this is not always the case: Portuguese speakers can reportedly understand Spanish with relative ease, but Spanish speakers have to strain to understand Portuguese because of its more complex phonology (Jensen, 1989). Thirdly, the analysis of semi-communication does not allow for the possibility that intelligibility can increase over time with exposure to the other language. This seems to be well-established in the mutual intelligibility literature, but the indexes here assume that intelligibility takes only a single fixed value. Third, my theoretical framework assumes that 'repeated games' necessarily lead to greater trust, but it is possible that in certain situations increased contact could actually worsen cross-group attitudes. This may occur when (as Schaeffer suggested) the groups have widely differing cultures and values, so greater exposure would lead only to greater animosity. These limitations leave ample room for refinement in future research.

The final major limitation of this analysis is that the regression techniques used here cannot confirm the causal nature of the relationships. Communication potential may be associated with cross-country trust, but this does not mean that communication potential causes cross-country trust. Perhaps the causal direction could be the reverse – countries who trust each other more are more likely to want to communicate, and will learn each other's languages in order to enable this to happen. Or the relationship may not be causal at all. It could be a spurious association driven by some other factor not included in my model. While these results are consistent with the theory that communication causes trust, it is impossible to know this with certainty from mere observational data. However, there is at least one reason to suspect that reverse causation may be less of a problem with this data in particular. Among the 21 countries, much of the communication potential across the dyads arises from people in both countries who have learned lingua francas such as English, German and sometimes Russian. These are usually not the main language of either country in the dyad. It therefore seems less likely that the regression results could have arisen from reverse causation because the communication generally did not actually arise from people learning the language of the other country in the dyad, although this does not rule out the possibility that the relationship is spurious.

8. Conclusion

This study found clear support for the theory that communication potential is linked to trust, and that this relationship is both positive and monotonic. This may go some way to explaining the numerous findings which have shown that both within-country and between-country diversity is linked to lower trust, and is consistent with Buzasi (2015). But this study is novel in some important ways. It is the first quantitative study of communication and trust which accounts for variations in mutual intelligibility: by including a measure of semi-communication, we can see whether the relationship between communication and trust is monotonic.

The results showed strong support for all three hypotheses. Both communication indexes reached statistical significance, even after adding controls, which strongly suggests that communication does play a role in generating or inhibiting trust. The semi-communication index was significant even when added alongside the shared language index, indicating that semi-communication has an independent, positive relationship with trust that is somewhat weaker than for full intelligibility. This indicates that the relationship between communication and trust is monotonic. Finally, the changes in size of the regression coefficients and the results of the structural equation model indicate that contact partially mediates the relationship between communication and trust. By providing evidence that communication facilitates trust through 'repeated games', these results are consistent with a rational choice approach to trust. My hope is that by combining this with historical, institutional and intergroup relations perspectives, scholars will be able to gain greater insight into this complex but important concept.

Acknowledgements:

This research was supported by funding from the Australian Government. I would like to thank Ben Goldsmith, Paul Kenny and Boris Bizumic for reading and commenting on the draft manuscript. Their feedback greatly improved the final product. I am also grateful to Keith Dowding for his helpful suggestions on the selection and framing of the dependent variable, and to Katrin Travouillon for her feedback on the nature and measurement of trust.

References

Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Warcziarg. 'Fractionalization'. *Journal of Economic Growth* 8 (2003): 155–94.

Alesina, Alberto, and Eliana La Ferrara. 'Who Trusts Others?' *Journal of Public Economics* 85, no. 2 (2002): 207–34.

Allport, Gordon. *The Nature of Prejudice*. Perseus Books, 1954.

Anderson, Christopher J., and Aida Paskeviciute. 'How Ethnic and Linguistic Heterogeneity Influence the Prospects for Civil Society: A Comparative Study of Citizenship Behavior'. *The Journal of Politics* 68, no. 4 (2006): 783–802.

Blumer, Herbert. 'Race Prejudice as a Sense of Group Position'. *The Pacific Sociological Review* 1, no. 1 (1958): 3–7. https://doi.org/10.2307/1388607.

Buzasi, Katalin. 'Languages, Communication Potential and Generalized Trust in Sub-Saharan Africa: Evidence Based on the Afrobarometer Survey'. *Social Science Research* 49 (2015): 141–55. https://doi.org/10.1016/j.ssresearch.2014.08.005.

Costa, Dora L., and Matthew E. Kahn. 'Civic Engagement and Community Heterogeneity: An Economist's Perspective'. *Perspectives on Politics* 1, no. 1 (2003): 103–11. https://doi.org/10.1017/S1537592703000082.

Delhey, Jan. 'Do Enlargements Make the European Union Less Cohesive? An Analysis of Trust between EU Nationalities*'. *JCMS: Journal of Common Market Studies* 45, no. 2 (2007): 253–79. https://doi.org/10.1111/j.1468-5965.2007.00711.x.

Delhey, Jan, and Kenneth Newton. 'Predicting Cross-National Levels of Social Trust: Global Pattern or Nordic Exceptionalism?' *European Sociological Review* 21, no. 4 (1 September 2005): 311–27. https://doi.org/10.1093/esr/jci022.

Fernandéz, Francisco Moreno, and Jaime Otero Roth. 'Demografía de La Lengua Española'. Instituto Complutense de Estudios Internacionales, 2006. http://www.escrituradigital.net/wiki/images/Demografia_de_la_lengua_española.pdf.

Fieldhouse, Edward, and David Cutts. 'Does Diversity Damage Social Capital? A Comparative Study of Neighbourhood Diversity and Social Capital in the US and Britain'. *Canadian Journal of Political Science/Revue Canadienne de Science Politique* 43, no. 2 (2010): 289–318. https://doi.org/10.1017/S0008423910000065.

Fisman, Raymond, and Tarun Khanna. 'Is trust a historical residue? Information flows and trust levels'. *Journal of Economic Behavior and Organization* 38 (1999): 79-92.

Gerritsen, Debby, and Marcel Lubbers. 'Unknown Is Unloved? Diversity and Inter-Population Trust in Europe'. *European Union Politics* 11, no. 2 (2010): 267–87.

Gooskens, Charlotte, Vincent J. van Heuven, Jelena Golubović, Anja Schüppert, Femke Swarte, and Stefanie Voigt. 'Mutual Intelligibility between Closely Related Languages in Europe'. *International Journal of Multilingualism* 15, no. 2 (3 April 2018): 169–93. https://doi.org/10.1080/14790718.2017.1350185.

Greenberg, Joseph H. 'The Measurement of Linguistic Diversity'. *Language* 32, no. 1 (1956): 109–15. https://doi.org/10.2307/410659.

Gundelach, Birte, and Richard Traunmüller. 'Beyond Generalised Trust: Norms of Reciprocity as an Alternative Form of Social Capital in an Assimilationist Integration Regime'. *Political Studies* 62, no. 3 (2013): 596–617. https://doi.org/10.1111/1467-9248.12064.

Habyarimana, James, Macartan Humphreys, Daniel N. Posner, and Jeremy M. Weinstein. 'Why Does Ethnic Diversity Undermine Public Goods Provision?' *American Political Science Review* 101, no. 4 (2007): 709–25.

Haugen, Einar. 'Semicommunication: The Language Gap in Scandinavia'. *Sociological Inquiry* 36, no. 2 (1966): 280–97. https://doi.org/10.1111/j.1475-682X.1966.tb00630.x.

Hooghe, Marc. 'Social Capital and Diversity: Generalized Trust, Social Cohesion and Regimes of Diversity'. *Canadian Journal of Political Science* 40, no. 3 (2007): 709–32.

Jensen, John B. 'On the Mutual Intelligibility of Spanish and Portuguese'. *Hispania* 72, no. 4 (1989): 848–52.

Kawachi, I, B P Kennedy, K Lochner, and D Prothrow-Stith. 'Social Capital, Income Inequality, and Mortality.' *American Journal of Public Health* 87, no. 9 (September 1997): 1491–98.

Kim, Chae-Han. 'Explaining Interstate Trust/Distrust in Triadic Relations'. *International Interactions* 33, no. 4 (19 November 2007): 423–39. https://doi.org/10.1080/03050620701681932.

Laitin, David. 'What Is a Language Community?' *American Journal of Political Science* 44, no. 1 (2000): 142–55.

Leigh, Andrew. 'Trust, Inequality and Ethnic Heterogeneity'. *Economic Record* 82, no. 258 (2006): 268–80.

Musgrave, Alan, and Charles Pigden. 'Imre Lakatos'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Winter 2016. Metaphysics Research Lab, Stanford University, 2016. https://plato.stanford.edu/archives/win2016/entries/lakatos/.

Ostrom, Elinor. 'Toward a Behavioural Theory Linking Trust, Reciprocity, and Reputation'. In *Trust and Reciprocity: Interdisciplinary Lessons From Experimental Research*. New York: Russell Sage Foundation, 2003.

Papacostas, Antonis. 'Eurobarometer 64.3: Foreign Languages, Biotechnology, Organized Crime, and Health Items, November-December 2005: Version 3'. Inter-university Consortium for Political and Social Research, 20 December 2006. https://doi.org/10.3886/ICPSR04590.v3.

Putnam, Robert. 'E Pluribus Unum: Diversity and Community in the Twenty-First Century'. *Scandinavian Political Studies* 30, no. 2 (2007): 137–74.

———. *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton: Princeton University Press, 1993.

Rajagopalan, Kanavillil. 'The Politics of Language and the Concept of Linguistic Identity'. *CAUCA, Revista de Filología y Su Didáctica* 24 (2001): 17–28.

Robinson, Amanda Lea. 'Ethnic Diversity, Segregation and Ethnocentric Trust in Africa'. *British Journal of Political Science*, 2017, 1–23.

Rokeach, M. *The Open and Closed Mind*. New York: Basic Books, 1960.

Schaeffer, Merlin. *Ethnic Diversity and Social Cohesion: Immigration, Ethnic Fractionalization and Potentials for Civic Action*. Farnham: Ashgate, 2014.

Schmitt, Hermann, Stefano Bartolini, Wouter van der Brug, Cees van der Eijk, Mark Franklin, Dieter Fuchs, Gabor Toka, Michael Marsh, and Jacques Thomassen. 'European Election Study, 2004 (2nd Edition)'. Cologne: GESIS Data Archive, 2009.

http://europeanelectionstudies.net/european-election-studies/ees-2004-study/voter-study.

Stolle, Dietlind, Stuart Soroka, and Richard Johnston. ‘When Does Diversity Erode Trust? Neighbourhood Diversity, Interpersonal Trust and the Mediating Effect of Social Interactions’. *Political Studies* 56, no. 1 (2008).

Tajfel, Henri, and John C. Turner. ‘The Social Identity Theory of Integroup Behaviour’. In *Psychology of Intergroup Relations*, edited by S. Worchel and W. G. Austin, 7–24. Chicago: Nelson-Hall, 1986.

Tenzer, Helene, Markus Pudelko, and Anne-Wil Harzing. ‘The Impact of Language Barriers on Trust Formation in Multinational Teams’. *Journal of International Business Studies* 45, no. 5 (2014): 508–35.

Thuesen, Frederik. ‘Linguistic Barriers and Bridges: Constructing Social Capital in Ethnically Diverse Low-Skill Workplaces’. *Work, Employment and Society* 31, no. 6 (2017): 937–53.

Tversky, Amos, and Daniel Kahneman. ‘Judgment under Uncertainty: Heuristics and Biases’. *Science* 185, no. 4157 (1974): 1124–31.

Uslaner, Eric M. ‘Segregation, Mistrust and Minorities’. *Ethnicities* 10, no. 4 (1 December 2010): 415–34. https://doi.org/10.1177/1468796810378325.

Uslaner, Eric M., and Richard S. Conley. ‘Civic Engagement and Particularized Trust: The Ties That Bind People to Their Ethnic Communities’. *American Politics Research* 31, no. 4 (July 2003): 331–60. https://doi.org/10.1177/1532673X03031004001.

Van Oudenhoven, Jan Pieter Van, Françoise Askevis-Leherpeux, Bettina Hannover, Renske Jaarsma, and Benoît Dardenne. ‘Asymmetrical International Attitudes’. *European Journal of Social Psychology* 32, no. 2 (2002): 275–89. https://doi.org/10.1002/ejsp.89.

Wang, Cong, and Bodo Steiner. ‘Can Ethno-Linguistic Diversity Explain Cross-Country Differences in Social Capital? A Global Perspective’. *Economic Record* 91, no. 294 (2015): 338–66.

Appendix A: Calculation of shared language index and semi-communication index scores

The aim of the shared language index was to determine the probability that two randomly selected people, one from each country in the dyad, will have at least on language in common. To figure this out, I created language repertoire lists for each of the 21 countries. Then, I compared the repertoires against one another for each of the 359 dyads. For instance, the language repertoire groups for the dyad Austria-Portugal are:

Figure A1: Language repertoires in Austria and Portugal

Country:	Austria		Country:	Portugal	
German	395	0.394211577	Portuguese	664	0.66599799
German and English	539	0.537924152	Portuguese and English	213	0.21364092
German and French	14	0.013972056	Portuguese and French	77	0.0772317
German and Italian	11	0.010978044	Portuguese and Spanish	30	0.03009027
Other	8	0.007984032	Portuguese and German	13	0.01303912
Other and German	35	0.03493014			

Now, I can calculate the product of any group encountering any of the groups in the other country, which reflects the probability of that encounter occurring. Each of the cells in the table below is obtained by finding the product of the proportions of each group from the two countries. For instance, since the proportion of monolingual German speakers in Austria is 0.394, and the proportion of monolingual Portuguese speakers in Portugal is 0.666, the product is 0.263. This is entered in the relevant cell in figure A2:

Figure A2: Probabilities of language encounters between Austria and Portugal

	Portuguese	Portuguese and English	Portuguese and French	Portuguese and Spanish	Portuguese and German
German	0.263	0.084	0.030	0.012	0.005
German and English	0.358	0.115	0.042	0.016	0.007
German and French	0.009	0.003	0.001	0.000	0.000
German and Italian	0.007	0.002	0.001	0.000	0.000
Other	0.005	0.002	0.001	0.000	0.000
Other and German	0.023	0.007	0.003	0.001	0.000

The same process was repeated for each of the other products in the table. Each product is taken to represent the probability that an encounter between one resident of each country will involve those two specific groups – so if a randomly-selected Austrian and a randomly-selected Portuguese were to meet, the probability that it will be a German monolingual and a Portuguese monolingual is 0.263. Next, I follow the procedure used by Greenberg and Laitin and sum up the encounters where the participants actually have a language in common. Those are shaded in figure A3 below:

Figure A3: Austria-Portugal language encounters where full communication is possible

	Portuguese	Portuguese and English	Portuguese and French	Portuguese and Spanish	Portuguese and German
German	0.263	0.084	0.030	0.012	0.005
German and English	0.358	0.115	0.042	0.016	0.007
German and French	0.009	0.003	0.001	0.000	0.000
German and Italian	0.007	0.002	0.001	0.000	0.000
Other	0.005	0.002	0.001	0.000	0.000
Other and German	0.023	0.007	0.003	0.001	0.000

The Portuguese who can speak English can speak to the Austrians who speak English (an encounter with a 0.115 probability of occurring). Similarly, Portuguese who speak French can talk to the Austrians who speak French, and the Portuguese who speak German can talk to the Austrians who speak German (which is almost all of them). I then sum up all the products for encounters where communication is possible. The result is 0.129, which I take as the 'shared language index' score between Austria and Portugal.

What about the semi-communication scores? To calculate these, I took all of the languages in my sample and identified the pairs where there is sufficient mutual intelligibility for semi-communication to occur. Information on how this was done is contained in Appendix B, but it relied heavily on Gooskens et al.'s (2018) study of mutual intelligibility. I also relied on testimonies from native speakers and language learners accessed online. Through this method I derived the following list of language pairs in the 21 countries which allow for semi-communication:

Catalan-Spanish	*Catalan-Italian*	*Galician-Portuguese*	*Galician-Spanish*	*Italian-Spanish*
Portuguese-Spanish	*Croatian – Slovenian*	*Czech-Slovak*	*Polish-Slovak*	*Danish-Swedish*

For dyads where any of these pairings occur, I calculated the sum of all the encounters where semi-communication occurs. This is essentially the same process as before, except that instead of counting encounters where there is a language in common, I counted those which involve one of the semi-communication pairings just listed. For instance in the Austria-Portugal dyad, there is one encounter where semi-communication is possible. This is highlighted with dark shading in figure A4:

Figure A4: Austria-Portugal encounters with semi-communication added

	Portuguese	Portuguese and English	Portuguese and French	Portuguese and Spanish	Portuguese and German
German	0.263	0.084	0.030	0.012	0.005
German and English	0.358	0.115	0.042	0.016	0.007
German and French	0.009	0.003	0.001	0.000	0.000
German and Italian	0.007	0.002	0.001	0.000	0.000
Other	0.005	0.002	0.001	0.000	0.000
Other and German	0.023	0.007	0.003	0.001	0.000

The small number of Austrians who speak Italian can partially converse with the small number of Portuguese who speak Spanish. In this case, the value of the dark-shaded square is quite tiny (0.00033 with rounding removed). This would be the score for the semi-communication index for the dyad Austria-Portugal. However in other dyads where semi-communication encounters are far more likely, the likelihood of semi-communication was quite large. The Italy-Spain semi-communication index score was 0.846, compared to s shared language index score of just 0.064.

Appendix B: Which language pairings should be coded as allowing for semi-communication?

In the 21 countries in my sample, I coded a total of 28 different languages (not including "other"). Only languages which had a mother tongue population of at least ten were counted, otherwise they were included in the "other" category. The 28 languages are:

Arabic	*Basque*	*Catalan*	*Croatian*
Czech	*Danish*	*Dutch*	*English*
Estonian	*Finnish*	*French*	*Galician*
German	*Greek*	*Hungarian*	*Irish Gaelic*
Italian	*Latvian*	*Lithuanian*	*Luxembourgish*
Polish	*Portuguese*	*Romanian*	*Russian*
Slovak	*Slovenian*	*Spanish*	*Swedish*

I assumed that only languages within the same subfamily even have the possibility of being intelligible with each other. This is in line with Gooskens et al's (2018) approach, which examined mutual intelligibility only within, not across, European language subfamilies. The subfamilies that these languages fall into are:

- **Romance**: Catalan, French, Galician, Italian, Portuguese, Romanian, Spanish
- **Slavic**: Croatian, Czech, Polish, Russian, Slovak, Slovenian
- **Germanic**: Danish, Dutch, English, German, Luxembourgish, Swedish
- **Baltic**: Latvian, Lithuanian
- **Celtic**: Irish Gaelic
- **Hellenic**: Greek

There are also five languages which fall outside of the Indo-European language family entirely. Their families are:

- **Finno-Ugric**: Estonian, Finnish, Hungarian
- **Basque**: Basque
- **Afro-Asiatic**: Arabic

I am concerned only with spoken intelligibility for the purposes of this chapter, so any references to "intelligibility" henceforth refer to spoken intelligibility. To see whether any languages within the Germanic, Slavic or Germanic subfamilies are intelligible with each other, I relied first on Gooskens et al's (2018) study. This examined mutual intelligibility between several languages across those families, although it did not include Catalan, Galician, Russian or Luxembourgish. I counted pairs of languages as partly intelligible if they had an average two-way listening score of above 40% for minimal-exposure listeners. For instance, Swedes can understand 56.0% of Danish, while Danes can understand 43.8% of Swedish. The average is 49.9%, which is above 40% and therefore qualifies this language pairing as partially intelligible. Using this procedure, I coded the following languages as partially intelligible:

Italian-Spanish	*Portuguese-Spanish*	*Croatian-Slovenian*
Czech-Slovak	*Polish-Slovak*	*Danish-Swedish*

This still leaves the question of Galician, Catalan, Luxembourgish and Russian. For these I relied on other online sources to indicate whether they are intelligible with other languages in their subfamilies. These were primarily online forums, where native speakers of the language in question report on whether they are able to partially understand other languages as well. These sources indicate that:

- **Catalan is partly intelligible with Spanish and Italian:**
    - "I'm Spanish. For us Spaniards, Catalan is understandable. The languages are different but pretty close. With a little imagination you can guess the meaning of most words. If you happen to know some French, that will definitely help too" [5]
    - Even non-native speakers of Spanish report being able to understand some spoken Catalan: "I found Catalan highly decipherable at times both written (very easy) and spoken (trickier but manageable)" [6]
    - Italian is reportedly even closer to Catalan than Spanish is: "For Catalan speakers, Italian is definitely very intelligible. Most of the time, they don't even subtitle Italian on Catalan TV...I have extensive experience talking with Italians, using western Catalan while they were using Italian, with little problem for understanding". Another poster writes that "When the lamguages are spoken , if catalan is spoken slowly and with valencian accent it's almost all intelligible, no each word but each sentence yes." [7]
    - Native speaker reports indicate that probably around 40% of Catalan is understandable to Italians who have never studied it, or a closely related language like Spanish. Italians who have studied some Spanish can understand Catalan even better, reporting 70% - 80% intelligibility.[8]
- **Galician is partly intelligible with both Portuguese and Spanish**
    - "If you speak Portuguese, you'll find that you understand Galician better than Spanish speakers". Posters tend to report that Galician is very intelligible with Portuguese, and somewhat less so with Spanish.[9]
    - "I am catalan myself … Galician is rather similar to Spanish, they may be easy to understand".[10]
- **Russian is not really intelligible with any of the other Slavic languages in the sample.**
    - Russian is reported to be highly intelligible with Belorussian and partly intelligible with Ukrainian. But other Slavic languages such as Croatian, Polish, Czech and so on are a little bit too different to be understandable. A Russian native speaker reports: "It is mutually intelligible with Belarusian and borderline mutually intelligible with Ukrainian. A Russian speaker can pick out

[5] https://spanish.stackexchange.com/questions/3825/difference-between-spanish-and-catalan/3828
[6] https://www.reddit.com/r/AskEurope/comments/6vlysq/spaniards_are_the_different_languages_in_spain/
[7] https://www.quora.com/Are-Catalan-and-Italian-mutually-intelligible
[8] Anna Reboldi, informal email conversation, 22 August 2018.
[9] https://www.quora.com/How-mutually-intelligible-is-Galician-to-Spanish-and-Portuguese-speakers
[10] https://forum.wordreference.com/threads/catalan-spanish-galician-mutual-intelligibility.162650/

isolated words and phrases in spoken Polish or Bulgarian. Croatian creates a sense of cognitive dissonance, because it sounds a lot like Russian except you can't recognize any words." [11]

- **Luxembourgish is not really intelligible with any of the other Germanic languages in the sample**.
    - Luxembourgish is reportedly closest to Dutch and German, but even these languages (in their standard forms) are too distinct for my standard of partial intelligibility to be met. A Dutch speaker reports that "I can understand quite some things of what they say, but understanding 50%? No, it's very hard to understand it without having learned it". [12]
    - However, dialects close to the Luxembourg border may be partially intelligible with Luxembourgish. "If you learn Standard German... no [you cannot understand it]. If you learn the western dialects, you could probably understand not only Luxembourgish, but Dutch too." [13]

This still leaves the question of the Baltic languages (Latvian and Lithuanian) and the Finno-Ugric languages (Finnish, Estonian and Hungarian). Sources indicate that:

- **Latvian and Lithuanian are not intelligible with each other to any real extent.**
    - Although from the same subfamily, they are too distinct for communication to be possible. "It's not the case of Latvian and Lithuanian [that they can communicate]. As far as I know, Bulgarian and Serbian are mutually inteligible, while Latvians and Lithuanians, in order to understand eachother, have to communicate in Russian or any other language they both speak." [14]
    - Another native speaker reports that "we are not able speak with each other. We can understand some single words only. In order to communicate we have to use English or Russian." [15]
- **Finnish, Estonian and Hungarian are not mutually or partly intelligible with each other to any real extent.**
    - Finnish and Estonian are the most closely-related of the three, but are still too different to allow mutual conversations, even simple ones. "Finnish and Estonian are not mutually intelligible. They belong to different sub-branches of Finnic languages. Finns and Estonians can understand [only] some words of each other's language". Speakers report that the languages are far less mutually intelligible than Spanish and Portuguese.[16]
    - Since Hungarian is considerably more distant from Finnish and Estonian than they are from each other, I have assumed that it therefore is also unintelligible with them.

[11] https://www.quora.com/Is-Russian-mutually-intelligible-with-any-other-Slavic-language
[12] https://forum.wordreference.com/threads/luxembourgish-mutual-intelligibility-with-other-germanic-languages-and-french.2173106/
[13] https://answers.yahoo.com/question/index?qid=20130702100425AAQE2vC
[14] https://forum.wordreference.com/threads/lithuanian-latvian-how-similar-different-are-they.2768830/
[15] https://hinative.com/en-US/questions/2234223
[16] https://www.quora.com/Are-Estonian-and-Finnish-as-mutually-intelligible-as-Spanish-and-Portuguese

Finally, since the Celtic, Basque, Hellenic and Afro-Asiatic families consist of only one language each in this sample, there are no possible languages with which they may be mutually intelligible. The process above, where Gooskens et al's (2018) research is supplemented by testimony from online sources, gives me the following list:

Catalan-Spanish	*Catalan-Italian*	*Galician-Portuguese*	*Galician-Spanish*	*Italian-Spanish*
Portuguese-Spanish	*Croatian – Slovenian*	*Czech-Slovak*	*Polish-Slovak*	*Danish-Swedish*

These ten pairings were classed as partly intelligible and allowing for semi-communication.