# Online appendices for "Language, ethnicity and separatism"

**Kyle L. Marquardt**

# A Survey methodology

The Moldovan survey firm IMAS-Inc implemented the survey under my supervision. Respondents were sampled according to a strategy that stratified first by urban and rural localities, then by ethnic composition; the Gagauz sample includes 23 localities (three urban and 20 rural) and the Pridnestrovian sample 27 localities (eight urban and 19 rural). As no publicly available listing of households exists in Pridnestrovie or Gagauzia, wholly random sampling was impossible. Clusters for sampling were created by randomly selecting a geographic starting point in a selected locality from a grid, then mapping households from this starting point following a systematic protocol, strictly avoiding overlap. Each locality included in the survey had a minimum of three thus-designed clusters, assigned proportionally to its population size. Twenty households from each cluster were sampled (i.e. each locality had a minimum of 60 houses included in the sample); respondents were randomly-selected from each household using Kish tables. Due to small population sizes that are prohibitive to the use of such a method, nine Pridnestrovian and eight Gagauz localities were fully mapped and 60 households were randomly selected from the universe of households in each each of these localities for inclusion in the survey.

Enumerators made at least two attempts at different times of the day to contact each sampled household, and returned at a later time if the selected individual respondent was unavailable. Survey enumerators were fluent in at least two of the three relevant languages in each region; if a respondent was unable to communicate in one of these two languages, she was re-contacted by another enumerator fluent in her language.

# B Perceptions of language

Here I examine perceptions of the relative utility of relevant languages in Pridnestrovie/Gagauzia vis-á-vis Moldova proper. These analyses provide *prima facie* evidence that language is likely to be salient in determining support for separatism for two reasons. First, survey respondents believe their regional governments are more supportive of non-Moldovan languages than the government of Moldova proper. Second, they also perceive the use of these languages to be more widespread in Pridnestrovie and Gagauzia than in Moldova proper. As a result, regional sovereignty has clear bearing on the status of non-Moldovan languages. I discuss the results from Pridnestrovie and Gagauzia in turn.

## B.1 Pridnestrovie

Table B.1 reports statistics about Pridnestrovian respondents' perceptions of 1) government support for relevant languages, and 2) the use of these languages in both Pridnestrovie and Moldova proper. The first two columns of Table B.1 represent the proportion of respondents who believe the relevant government "supports" or "completely supports" the use of a given language in the territory which it controls; these are the top two categories in a five-point Likert scale question. The second two columns represent the proportion of respondents who perceive the population of a given territory to use the relevant language "frequently" or "almost all the time," the top two items on a five-point Likert scale question regarding

Table B.1: Perceptions of languages in Pridnestrovie and Moldova proper

| | Government supports language | | Language frequently used | |
|---|---|---|---|---|
| | Pridnestrovie | Moldova | Pridnestrovie | Moldova |
| Russian | 0.97 ($n = 538$) | 0.52 ($n = 333$) | 0.99 ($n = 542$) | 0.59 ($n = 542$) |
| Ukrainian | 0.86 ($n = 530$) | 0.34 ($n = 307$) | 0.50 ($n = 538$) | 0.17 ($n = 335$) |
| Moldovan | 0.84 ($n = 532$) | 0.91 ($n = 372$) | 0.53 ($n = 537$) | 0.97 ($n = 412$) |

the frequency of language use in the two territories. Both of these concepts have strong bearing on the utility of a language. The degree to which a government supports a language proxies its usefulness for accessing state resources, as well as the long-term trajectory of the language; while the frequency with which a a region's population uses a language proxies its present usefulness for day-to-day life.

The results indicate that respondents perceive the Pridnestrovian government to be supportive of the Russian and Ukrainian languages in absolute terms, as well as more supportive of these languages than the Moldovan government (over 60 percent of respondents report that the Pridnestrovian government is more supportive of these languages than the Moldovan government). Results regarding the Moldovan language are more ambiguous, with large majorities believing that both the Moldovan government and the Pridnestrovian government support the language, though a relatively large proportion of respondents perceive the Pridnestrovian government as being less supportive of the language than the Moldovan government (32 percent of respondents, compared to 11 percent who believe the Pridnestrovian government is more supportive).

The data also confirm that Russian use is widespread in Pridnestrovie, with a large majority of respondents perceiving the language to be frequently spoken in Pridnestrovie in absolute terms; a further 64 percent of respondents believe Russian is more frequently used in Pridnestrovie than in Moldovan proper. The data also reinforce the link between the population of Moldova and the Moldovan language: 61 percent of respondents believe Moldovan is more frequently used in Moldova proper than in Pridnestrovie.

## B.2 Gagauzia

Table B.2 reports results regarding perceptions of language use and government support in Gagauzia, presented using similar data and aggregation as in Pridnestrovie. The data show that an overwhelming majority of respondents believe that both Russian and Gagauz enjoy the support of the Gagauz government, and this level of support is greater than that of the Moldovan government (88 percent for Russian and 90 percent for Gagauz). As in Pridnestrovie, respondents perceive both the Gagauz and Moldovan governments as being supportive of the Moldovan language, though they tend to perceive the Moldovan government as having a higher level of support for the language (64 percent of respondents believe the Moldovan government is more supportive of Moldovan than the Gagauz government).

Results regarding perceptions of the use of the Gagauz and Russian languages in both Gagauzia and Moldova proper are similarly stark: respondents perceive the languages to be frequently used in absolute terms in Gagauzia, and more frequently used in Gagauzia than in Moldova proper (86 and 94 percent for Russian and Gagauz respectively). In contrast,

Table B.2: Perceptions of languages in Gagauzia and Moldova proper

| | Government supports language | | Language frequently used | |
|---|---|---|---|---|
| | Gagauzia | Moldova | Gagauzia | Moldova |
| Russian | 0.99 ($n = 793$) | 0.25 ($n = 724$) | 0.98 ($n = 829$) | 0.43 ($n = 722$) |
| Gagauz | 0.99 ($n = 792$) | 0.18 ($n = 681$) | 0.90 ($n = 829$) | 0.05 ($n = 702$) |
| Moldovan | 0.81 ($n = 763$) | 0.98 ($n = 754$) | 0.23 ($n = 817$) | 0.97 ($n = 741$) |

respondents clearly perceive Moldovan as being more frequently used in Moldova proper than in Gagauzia (92 percent), and overall usage of Moldovan to be infrequent in Gagauzia.

Table B.2 also provides evidence that Russian is perceived to be more widespread in Gagauzia than Gagauz: 21 percent of respondents consider Russian to be more widely used than Gagauz, compared to 12 percent who believe Gagauz use to be more widespread. (Perceptions of Gagauz government support for Gagauz and Russian are roughly equivalent: eight percent of respondents believe the Gagauz government is more supportive of Gagauz than Russian, compared to four percent who believe the opposite.) This finding indicates that Russian may be relatively privileged in Gagauzia, providing additional evidence that it could be more salient for separatist sentiment.

# C Components of ethnic identity

Enumerators asked respondents report if any of the items from a list are important to considering oneself a member of the respondent's ethnic group; respondents respond either affirmatively or negatively to each item (i.e. "yes" or "no"). This list is based on the instrument used by David Laitin (1998, 2001).

I report the proportion of respondents who responded affirmatively by region and ethnic group, with nonresponse coded as a negative response. I order items by the proportion of respondents which consider them important, and bold highlights the items most commonly considered important by ethnic group and region. While there is variation across regions and ethnic groups in terms of items considered important; patrilineal descent, respect for ethnic cultural traditions and knowledge of an ethnic language are in the top five most-commonly noted aspects of ethnic identity in both regions. In line with the notion of ethnicity being descent-based, patrilineal descent is more commonly considered important for ethnic identity than language in both regions. However, respondents from Gagauzia in particular consider linguistic knowledge as being important for membership in the group, in line with arguments that language can be an important ethnic symbol.

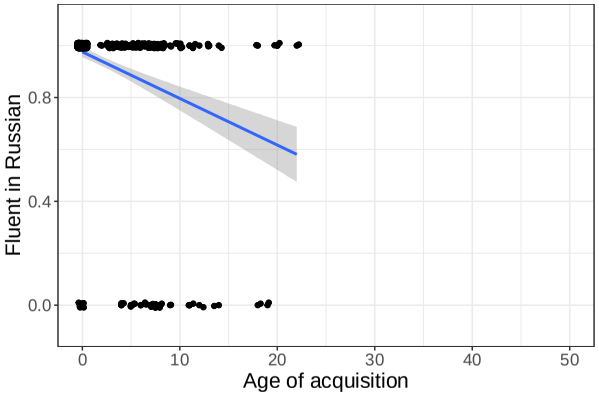
Table C.1: Components of ethnic identity

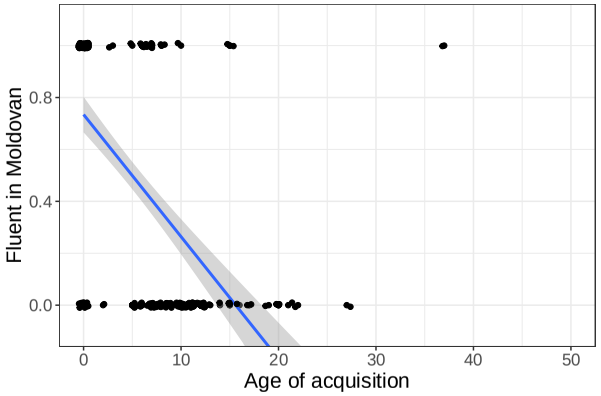
| Pridnestrovie | | | | |
|---|---|---|---|---|
| | Russian | Moldovan | Ukrainian | Overall |
| Respect for ethnic cultural traditions | **0.82** | 0.69 | 0.68 | **0.75** |
| Father's ethnicity | 0.68 | **0.70** | **0.75** | 0.70 |
| Respect for ethnic professional traditions | 0.75 | 0.64 | 0.56 | 0.67 |
| Citizenship | 0.61 | **0.70** | 0.73 | 0.65 |
| Knowledge of ethnic language | 0.59 | **0.70** | 0.63 | 0.63 |
| Mother's ethnicity | 0.57 | 0.63 | 0.66 | 0.61 |
| Teaching children ethnic language | 0.55 | 0.67 | 0.60 | 0.59 |
| Common history with other members | 0.60 | 0.60 | 0.55 | 0.58 |
| Self-assessment | 0.55 | 0.60 | 0.58 | 0.57 |
| Belief in religion of ethnic group | 0.44 | 0.57 | 0.51 | 0.49 |
| Place of residence | 0.38 | 0.54 | 0.55 | 0.47 |
| Outer appearance | 0.32 | 0.48 | 0.40 | 0.38 |
| Marriage within ethnic group | 0.26 | 0.44 | 0.36 | 0.32 |
| Gagauzia | | | | |
| | Russian | Moldovan | Gagauz | Overall |
| Father's ethnicity | 0.86 | 0.82 | **0.94** | **0.92** |
| Knowledge of ethnic language | **0.97** | **0.89** | 0.89 | 0.89 |
| Respect for ethnic cultural traditions | 0.86 | 0.85 | 0.86 | 0.86 |
| Belief in religion of ethnic group | 0.83 | 0.69 | 0.77 | 0.77 |
| Respect for ethnic professional traditions | 0.79 | 0.71 | 0.74 | 0.73 |
| Mother's ethnicity | 0.59 | 0.73 | 0.70 | 0.70 |
| Self-assessment | 0.83 | 0.69 | 0.66 | 0.66 |
| Common history with other members | 0.69 | 0.65 | 0.66 | 0.66 |
| Teaching children ethnic language | 0.83 | 0.67 | 0.65 | 0.65 |
| Citizenship | 0.41 | 0.69 | 0.57 | 0.55 |
| Place of residence | 0.28 | 0.65 | 0.54 | 0.53 |
| Outer appearance | 0.38 | 0.36 | 0.33 | 0.33 |
| Marriage within ethnic group | 0.17 | 0.13 | 0.24 | 0.23 |

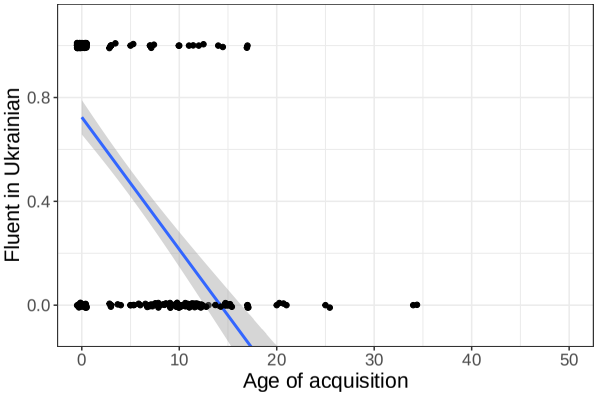
# D Language and age of acquisition

Figure D.1 presents the relationship between age of language acquisition (i.e. the age at which a respondent reports having begun learning a given language) and self-reported fluency in the language across relevant languages in Pridnestrovie and Gagauzia (measured using the standard dichotomous indicator). For all languages, there is 1) a strong negative correlation between age of acquisition and self reported fluency, and 2) stark drop in self-reported spoken fluency after the age of 20, with very few respondents who learned the language after this age reporting fluency. This relationship is strong evidence that developing fluency in a language generally requires learning the language early in life, reinforcing the claim that choice in linguistic acquisition is limited for adults.

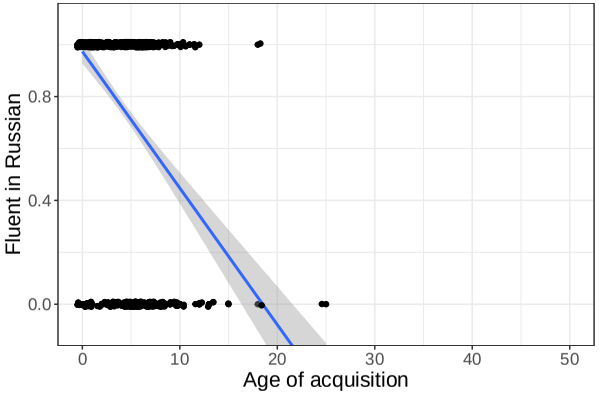
Figure D.1: Age of acquisition and fluency in different languages

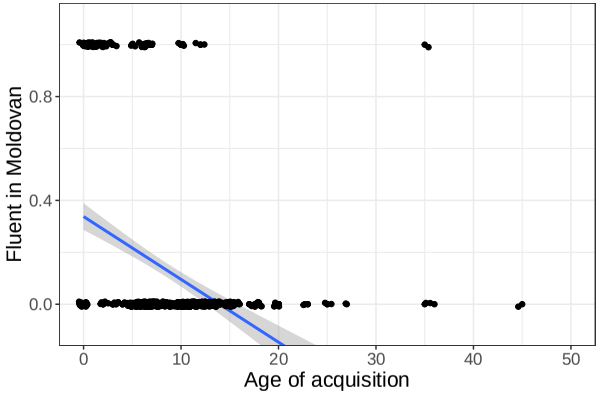


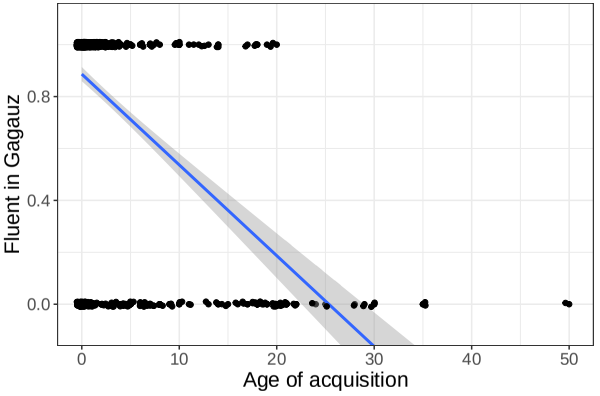




(a) Pridnestrovie




(b) Gagauzia

# E Fluency and reasons for language acquisition

Table E.1 presents the proportion of individuals who report speaking a language fluently by the primary reason they learned the language. In the case of individuals who reported learning a language because of a parental or family decision (i.e. they did not choose to learn the language themselves), I also report the primary reason respondents believe this decision was made. Note that, with the exception of Russian in Gagauzia, respondents who learned a language due to a parental or family reason are the most likely to report fluency in the language, dovetailing with the findings from Figure D.1; this reason is also the modal response for why a respondent learned a language for all languages save Russian and Gagauz in Gagauzia. (In these two cases, the primary rationale is that these languages is necessity for daily life.) Among those individuals who learned a language due to a family or parental reason, there is little evidence that those with an identity-based reason are drastically more likely to be fluent than those who learned for another reason (i.e. their parents only spoke the language and communication with others).

Table E.1: Proportion of respondents fluent in a language by reason for learning language

| Pridnestrovie | | | |
|---|---|---|---|
| | Russian | Moldovan | Ukrainian |
| Did not learn | 0.00 ($n = 2$) | 0.00 ($n = 327$) | 0.01 ($n = 328$) |
| Necessary for daily life | 0.89 ($n = 123$) | 0.32 ($n = 34$) | 0.31 ($n = 32$) |
| Necessary for education | 0.81 ($n = 37$) | 0.13 ($n = 63$) | 0.07 ($n = 67$) |
| Necessary for work | - | 0.50 ($n = 2$) | 0.00 ($n = 3$) |
| Personal desire | 0.87 ($n = 30$) | 0.21 ($n = 29$) | 0.08 ($n = 24$) |
| Parental or family decision | 0.98 ($n = 375$) | 0.83 ($n = 111$) | 0.79 ($n = 113$) |
| *If learned due to parental or family decision, proportion fluent by reason for learning* | | | |
| Parents only spoke the language | 0.95 ($n = 122$) | 0.80 ($n = 55$) | 0.86 ($n = 59$) |
| To communicate with others | 0.99 ($n = 157$) | 0.89 ($n = 37$) | 0.79 ($n = 38$) |
| Common identity with people or culture | 1.00 ($n = 72$) | 1.00 ($n = 12$) | 0.44 ($n = 9$) |
| To gain access to information or culture | 1.00 ($n = 11$) | 0.50 ($n = 4$) | 0.67 ($n = 3$) |
| Gagauzia | | | |
| | Russian | Moldovan | Gagauz |
| Did not learn | 0.00 ($n = 11$) | 0.00 ($n = 295$) | 0.00 ($n = 70$) |
| Necessary for daily life | 0.82 ($n = 525$) | 0.24 ($n = 88$) | 0.81 ($n = 485$) |
| Necessary for education | 0.53 ($n = 203$) | 0.03 ($n = 286$) | 0.25 ($n = 8$) |
| Necessary for work | 0.90 ($n = 10$) | 0.08 ($n = 38$) | 0.43 ($n = 7$) |
| Personal desire | 0.75 ($n = 8$) | 0.11 ($n = 47$) | 0.50 ($n = 28$) |
| Parental or family decision | 0.86 ($n = 76$) | 0.85 ($n = 34$) | 0.93 ($n = 216$) |
| *If learned due to parental or family decision, proportion fluent by reason for learning* | | | |
| Parents only spoke the language | 0.89 ($n = 18$) | 0.90 ($n = 10$) | 0.93 ($n = 70$) |
| To communicate with others | 0.79 ($n = 29$) | 0.50 ($n = 4$) | 0.86 ($n = 14$) |
| Common identity with people or culture | 0.91 ($n = 23$) | 0.89 ($n = 19$) | 0.94 ($n = 126$) |
| To gain access to information or culture | 1.00 ($n = 4$) | 1.00 ($n = 1$) | 0.67 ($n = 3$) |

Overall, these results indicate that identity (either their own or their family's) played a role in many respondents' decisions to learn a language. However, they also indicate that identity alone does not lead to fluency—indeed, those who reported learning a language out of personal desire tend to be among the least likely to report fluency—and that many respondents who learned a language out of necessity are among those most likely to speak the language fluently.

Note that these data do not speak directly to the identity content of language. Speaking a language fluently could lead respondents to identify more strongly with the ethnic group with which it is associated or, as I argue, with co-linguals, regardless of why they learned the language. However, they do indicate that linguistic fluency is not necessarily a function of pre-existing identities, ethinic or otherwise.

# F Survey questions

## F.1 Outcome questions

To avoid redundancy, I only show survey questions for Pridnestrovie. Questions for Gagauzia are identical, with "Gagauzia" replacing "Pridnestrovie."

### F.1.1 Observational outcomes

For observational outcome questions, enumerators presented respondents with a card representing the question response scale. Table F.1 illustrates this card.

Table F.1: Response scale to observational outcome questions

| Fully disagree | Disagree | Neither...nor... | Agree | Fully agree | Don't know | NA |
|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 8 | 9 |

Enumerators then read the following script, recording responses after each statement: *Using a scale from 1 to 5, where 1 = I fully disagree and 5 = I fully agree, please tell us how much you agree with the following statements:*

- *Pridnestrovie should be a region with no special status within Moldova.*
- *Pridnestrovie should have autonomous status within Moldova.*
- *Pridnestrovie should have confederal status with Moldova.*
- *Pridnestrovie should be an independent state.*

*Now I would like to ask you several questions about relations between Pridnestrovie and Russia. Using a scale from 1 to 5, where 1 = I fully disagree and 5 = I fully agree, please tell us how much you agree with the following statements:*

- *Pridnestrovie should be a region with no special status within Russia.*

- *Pridnestrovie should have autonomous status within Russia.*
- *Pridnestrovie should have confederal status with Russia.*

*Now I would like to ask you several questions about relations between Pridnestrovie and Ukraine. Using a scale from 1 to 5, where 1 = I fully disagree and 5 = I fully agree, please tell us how much you agree with the following statements:*

- *Pridnestrovie should be a region with no special status within Ukraine.*
- *Pridnestrovie should have autonomous status within Ukraine.*
- *Pridnestrovie should have confederal status with Ukraine.*

Note: Questions regarding Ukraine not asked in Gagauzia.

### F.1.2 Experimental outcome

- *What do you think would be the best solution for Moldovan-Pridnestrovian relations?*
  1. *Pridnestrovie should be part of Moldova, without a special status like autonomy.*
  2. *Pridnestrovie should be part of Moldova, with autonomy.*
  3. *Pridnestrovie should be part of a confederative state with Moldova.*
  4. *Moldova and Pridnestrovie should be separate states.*

## F.2 Covariate questions

- *What is your nationality (in sense of ethnicity)?*

This question was open-ended, with the pre-set categories of Russian, Moldovan, Gagauz, Ukrainian, and Bulgarian; enumerators entered other groups manually.

- *Which of the following best describes your speaking abilities in the following languages?*

Enumerators presented respondents with a card illustrating the scale (Table F.2), asking: *Which of the following best describes your abilities to understand the following languages when spoken?*. Enumerators asked about the Russian, Moldovan and then Ukrainian/Gagauz languages. I use the top category of the scale as the dichotomous measure of fluency.

Table F.2: Response scale to spoken fluency question

| I cannot speak at all | I can speak with extreme difficulty | I speak it with difficulty | I am not fluent, but speak it with ease | I am fluent | DK/ NA |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 9 |

# G Descriptive statistics

Table G.1: Independent variables

| | Pridnestrovie | | Gagauzia | |
|---|---|---|---|---|
| | Mean | N | Mean | N |
| Language | | | | |
| Not fluent in Russian | 0.07 | 576 | 0.26 | 835 |
| Fluent in Moldovan | 0.21 | 575 | 0.08 | 836 |
| Fluent in Ukrainian/Gagauz | 0.19 | 573 | 0.74 | 833 |
| Ethnic identity | | | | |
| Russian | 0.48 | 577 | 0.03 | 836 |
| Moldovan | 0.23 | 577 | 0.07 | 836 |
| Gagauz | - | - | 0.80 | 836 |
| Ukrainian | 0.25 | 577 | 0.03 | 836 |
| Bulgarian | - | - | 0.06 | 836 |
| Other | 0.04 | 577 | 0.00 | 836 |
| Controls | | | | |
| $\ln(\frac{Income}{\sqrt{N.Household.Members}})$ | 7.50 (0.34) | 577 | 6.80 (0.53) | 836 |
| ln(Age) | 3.81 (0.40) | 577 | 3.85 (0.35) | 836 |
| Male | 0.46 | 577 | 0.46 | 836 |
| Higher education | 0.29 | 558 | 0.16 | 833 |
| Urban | 0.49 | 577 | 0.29 | 836 |

Quantities in parentheses represent standard deviation of continuous responses. Income imputed; quantities refer to the mean and standard deviation of posterior mean estimates across respondents. N refers to number of responses to each question or set of questions.

Table G.2: Observational outcomes

| | Pridnestrovie | | | Gagauzia | | |
|---|---|---|---|---|---|---|
| | Median / Proportion | N | Proportion "Don't know" / Total NA | Median / Proportion | N | Proportion "Don't know" / Total NA |
| Support independence | 5 (4, 5) | 458 | 0.58 | 4 (2, 4) | 604 | 0.87 |
| Integration with Moldova | | | | | | |
| No autonomy | 1 (1, 2) | 437 | 0.57 | 2 (1, 3) | 636 | 0.90 |
| Autonomy | 1 (1, 3) | 437 | 0.57 | 4 (4, 5) | 729 | 0.90 |
| Confederation | 2 (1, 3) | 436 | 0.57 | 4 (3, 4) | 556 | 0.85 |
| At least one Moldova question | | 439 | | | 734 | |
| Integration with Russia | | | | | | |
| No autonomy | 4 (2, 4) | 403 | 0.64 | 4 (3, 4) | 711 | 0.84 |
| Autonomy | 3 (2, 4) | 401 | 0.65 | 4 (4, 4) | 723 | 0.85 |
| Confederation | 4 (3, 4) | 399 | 0.66 | 4 (3, 5) | 668 | 0.85 |
| At least one Russia question | | 406 | | | 754 | |
| Integration with Ukraine | | | | | | |
| No autonomy | 2 (1, 3) | 414 | 0.63 | | | |
| Autonomy | 2 (1, 3) | 415 | 0.63 | | | |
| Confederation | 2 (1, 3) | 416 | 0.63 | | | |
| At least one Ukraine question | | 417 | | | | |
| At least one Likert scale question | | 473 | | | 808 | |
| At least one outcome question (cross-sectional and experimental) | | 520 | | | 814 | |

Quantities in parentheses represent second and fourth quantiles of ordinal responses. N refers to number of responses to each question or set of questions.

Table G.3: Experimental outcomes and conditions

| | Pridnestrovie | | | Gagauzia | | |
|---|---|---|---|---|---|---|
| | Median / Mean | N | Proportion "Don't know" / Total NA | Median / Mean | N | Proportion "Don't know" / Total NA |
| Ideal outcome for region-Moldova relations | 4 (4, 4) | 456 | 0.53 | 2 (2, 3) | 676 | 0.90 |
| Experimental condition | | | | | | |
| Control | 0.35 | 577 | | 0.34 | 836 | |
| $T_{Moldovan}$ | 0.21 | 577 | | 0.22 | 836 | |
| $T_{Russian}$ | 0.22 | 577 | | 0.22 | 836 | |
| $T_{Gagauz/Ukrainian}$ | 0.21 | 577 | | 0.22 | 836 | |

Quantities in parentheses represent second and fourth quantiles of ordinal responses. N refers to number of responses to each question or set of questions. Note that control condition was oversampled to facilitate observational analysis.

# H Expectations

Table H.1: Relationship between ethnicity, language and support for separatism in Pridnestrovie

| Identity | Theoretical relationship | Ethnicity/language | Support for independence | Effect of linguistic prime |
|---|---|---|---|---|
| Russian | + | Monolingual | | -.02 (-.17, .13) |
| | - | Bilingual (Moldovan) | -.02 (-.13, .07) | **-.36 (-.69, .01)** |
| | (-) | Bilingual (Ukrainian) | -.03 (-.14, .07) | .20 (-.10, .51) |
| Moldovan | - | Ethnicity | **-.12 (-.26, -.00)** | .08 (-.26, .43) |
| | | Bilingual | -.02, (-.16, .11) | **-.30 (-.52, -.05)** |
| | | Monolingual | **-.24 (-.46, -.00)** | -.26 (-.64, .15) |
| Ukrainian | (-) | Ethnicity | -.06 (-.17, .04) | **-.45 (-.75, -.11)** |
| | | Bilingual | -.03 (-.15, .08) | -.15 (-.41, .14) |
| | | Monolingual | **-.24 (-.43, -.04)** | .09 (-.45, .57) |

Results are on the posterior probability scale. Parentheses around a theoretical relationship represents ambiguous expectations. Statistics represent median and 95% credible regions over posterior draws from the analyses discussed in the text; bold represents distributions for which the 90% credible region does not overlap zero. Support for independence values represent differences between ethnic and linguistic combinations. For ethnic Russians, bilingual values are relative to a monolingual Russian-speaking Russian. For ethnic Moldovans and Ukrainians, "Ethnicity" compares the posterior probability for between a monolingual Russian-speaking member of the respective ethnic group to a Russian with the same linguistic repertoire; "Bilingual" and "Monolingual" values are relative to a monolingual Russian-speaking member of the same ethnic group. Effect of linguistic prime is difference between treatment and control condition for a relevant prime. For ethnic Russians, primes correspond to the Russian language for monolinguals, the Moldovan language for Moldovan/Russian bilinguals, and the Ukrainian language for Ukrainian/Russian bilinguals. For Moldovans and Ukrainians, all effects regard linguistic prime for the ethnic group (Moldovan for Moldovans, Ukrainian for Ukrainians).

Table H.2: Relationship between ethnicity, language and support for separatism in Gagauzia

| Identity | Theoretical relationship | Ethnicity/language | Support for independence | Effect of linguistic prime |
|---|---|---|---|---|
| Gagauz | ()/+ | Monolingual Russian | | R: -.10 (-.25, .03) |
| | + /() | | | G: **-.12 (-.26, .01)** |
| | +/() | Bilingual | .04 (-.05, .15) | R: -.01 (-.08, .07)<br>G: .01 (-.06, .08) |
| | +/(-) | Monolingual Gagauz | **-.13 (-.26, .02)** | .07 (-.02, .17) |
| Moldovan | - | Ethnicity | **-.16 (-.33, .00)** | |
| | | Bilingual | -.07 (-.21, .08) | |
| | | Monolingual | **-.22 (-.38, -.06)** | |
| Russian | (-)/() | Ethnicity | -.06 (-.26, .13) | |
| | +/() | Bilingual (Gagauz) | -.07 (-.21, .08) | |
| | - | Bilingual (Moldovan) | .04 (-.06, .14) | |

Results are on the posterior probability scale. Parentheses around a theoretical relationship represents ambiguous expectations; when ethnic and linguistic theoretical expectations diverge the left-most sign represents the ethnic expectation, the right sign the linguistic expectation. When there is no expectation, there are only parentheses. For monolingual Russian-speaking ethnic Gagauz, theoretical expectations regard only prime effects. Statistics represent median and 95% credible regions over posterior draws from the analyses discussed in the text; bold represents distributions for which the 90% credible region does not overlap zero. Support for independence values represent differences between ethnic and linguistic combinations. For ethnic Gagauz, posterior probabilities are relative to values for a monolingual Russian-speaking Gagauz. For ethnic Moldovans and Russians, "Ethnicity" is the relative value of monolingual Russian-speaking member of the respective ethnic group to a monolingual Russian-speaking Gagauz; "Bilingual" and "Monolingual" are relative to a monolingual Russian-speaking member of relevant group. The effect of the linguistic prime is the difference between treatment and control condition for a relevant prime. I report both Russian (R) and Gagauz (G) prime effects for Russian-speaking Gagauz; for monolingual Gagauz-speaking Gagauz I only report results for the Gagauz prime. I report no experimental results for Russians and Moldovans due to the low sample size.

# I Observational analysis

I analyze support for 10 separatist outcomes in Pridnestrovie and seven in Gagauzia, namely the degree to which a respondent supports regional independence and outcomes related to different forms of integration with relevant states: integration 1) without autonomy, 2) with autonomy and 3) as a confederation with each of three states: Moldova, Russia and Ukraine (in Gagauzia I only ask about integration with Moldova and Russia, since integration with Ukraine is not a plausible scenario). The ordinal probit model I use diverges from standard models in that I pool thresholds across outcomes, accounting for differences in overall support with outcome-specific intercepts. That is, I assume the distance between "Fully disagree" and "Disagree" is constant across outcomes, but respondents are generally more supportive of some outcomes than others.

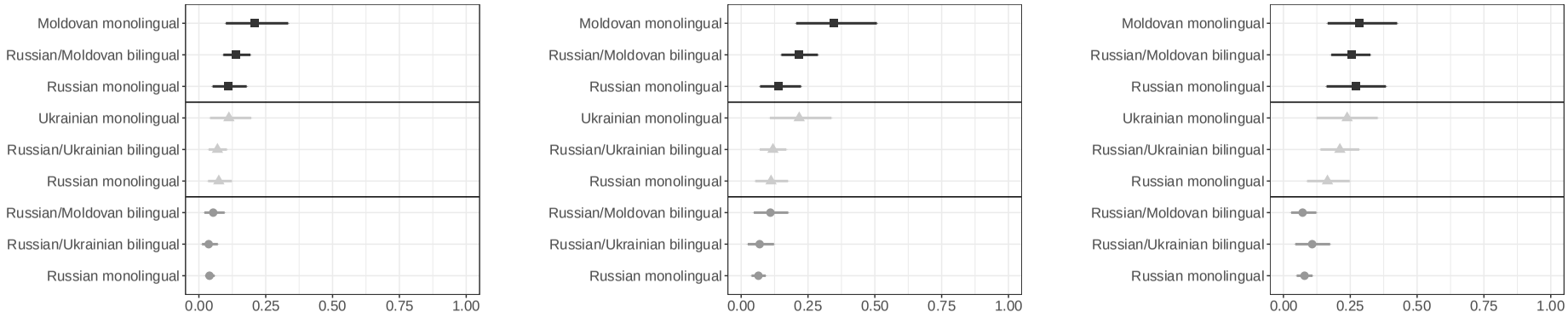
The probability that respondent $i$ provides one of the five ordinal responses for each of the $n$ questions (10 in Pridnestrovie, seven in Gagauzia) is thus $\phi\{\tau_j - (\gamma_n + X_{ik}\alpha_{kn} + Z_{im}\beta_{mn})\} - \phi\{\tau_{j-1} - (\gamma_n + X_{ik}\alpha_{kn} + Z_{im}\beta_{mn})\}$, and the response is distributed according to a categorical distribution with the resulting vector of five probabilities. $\phi$ is the CDF of the normal distribution, and $\tau$ one of $j = 1, 2, 3, 4$ thresholds. Note that since there is no theoretical reason to believe that the distance between thresholds varies across outcomes, I pool $\tau$ across outcomes, and account for generally higher levels of support for different outcomes with the outcome-specific parameter $\gamma$. $X$ is an $i \times k$ matrix of the $k$ ethnic and linguistic covariates (six in Pridnestrovie and eight in Gagauzia) and $\alpha$ a vector of $k$ coefficients; $\beta$ represents the $m = 5$ coefficients for control variables $Z$. Both $\alpha$ and $\beta$ are distributed according to a standard normal distribution distributed about zero and a precision of one.

Missing outcome values are imputed iteratively over MCMC simulations; missing right-hand side variables imputed using a population-level $Beta(1, 1)$ distribution. I impute income from the ordinal question scale using a Negative Binomial distribution with a mean value estimated based on respondent's demographic characteristics (age, gender, marital status, education), the number of working-age respondents in household, locality of residence and profession. I estimate income by fitting the estimate to both the response scale item minimum and maximum, truncating such that estimate cannot exceed each respondents' reported response scale range. I do not assess convergence for estimated income since many values are at the respondent maximum or minimum.

I run each model using four chains using *runjags*. Each chain involves 500 thousand iterations after a 50 thousand iteration burn-in; I sample 1000 draws from each chain and assess convergence with the Gelman-Rubin diagnostic (all parameters' $\hat{r} < 1.1$). Note that I model the observational analyses together with both the experimental analyses (Appendix J) and analyses of outcome missingness (Appendix K), which means that right-hand side missing values are consistent across all analyses.

## I.1 Figures for additional regression results

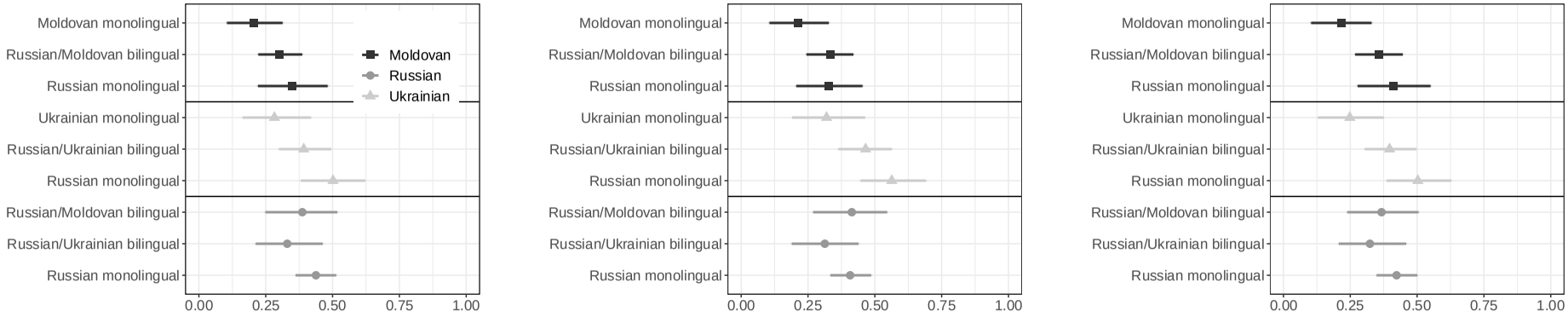
Figure I.1: Posterior probability that Pridnestrovian respondents support integration with Moldova




(a) Complete (b) Autonomy (c) Confederation



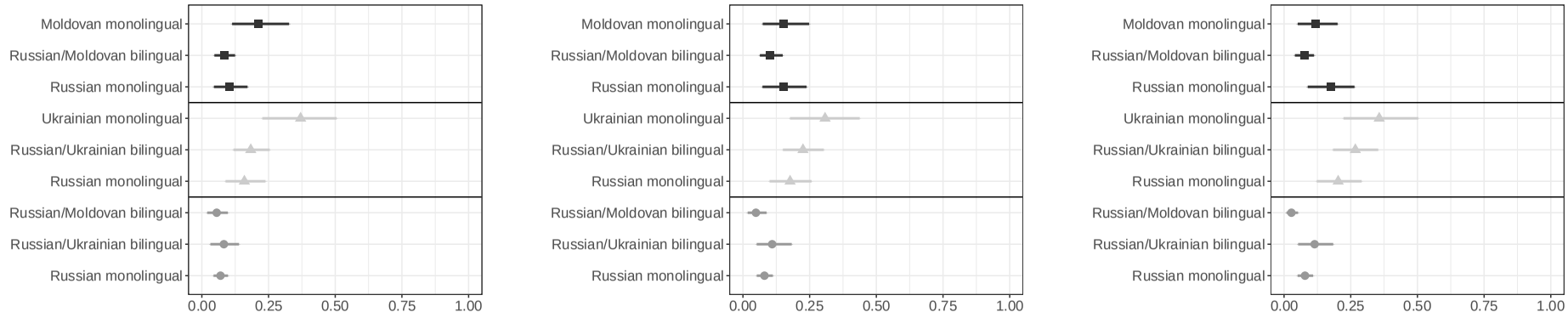
Figure I.2: Posterior probability that Pridnestrovian respondents support integration with Russia




(a) Complete (b) Autonomy (c) Confederation

Points represent posterior median and horizontal lines 90 percent credible regions over 500 thousand iterations of four MCMC chains.

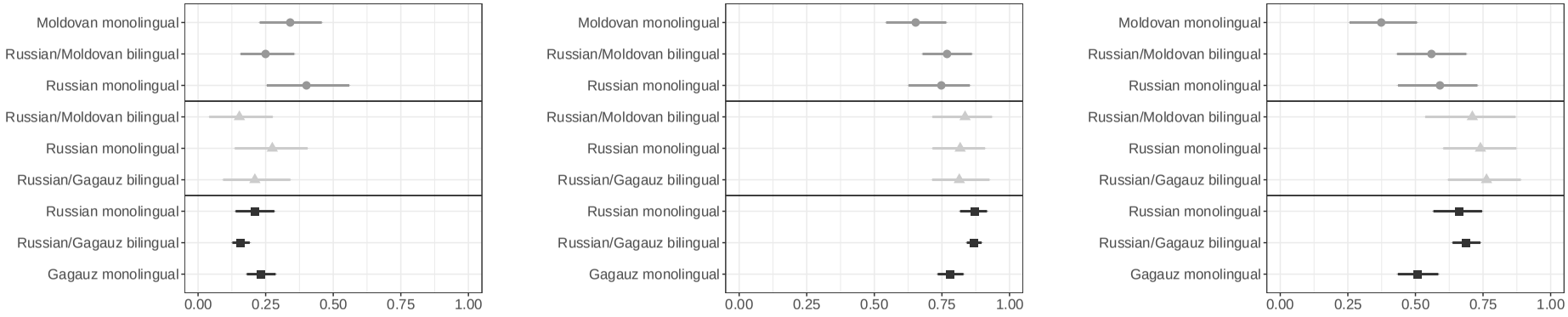
Figure I.3: Posterior probability that Pridnestrovian respondents support integration with Ukraine




(a) Complete (b) Autonomy (c) Confederation

Points represent posterior median and horizontal lines 90 percent credible regions over 500 thousand iterations of four MCMC chains.

Figure I.4: Posterior probability that Gagauz espondents support integration with Moldova

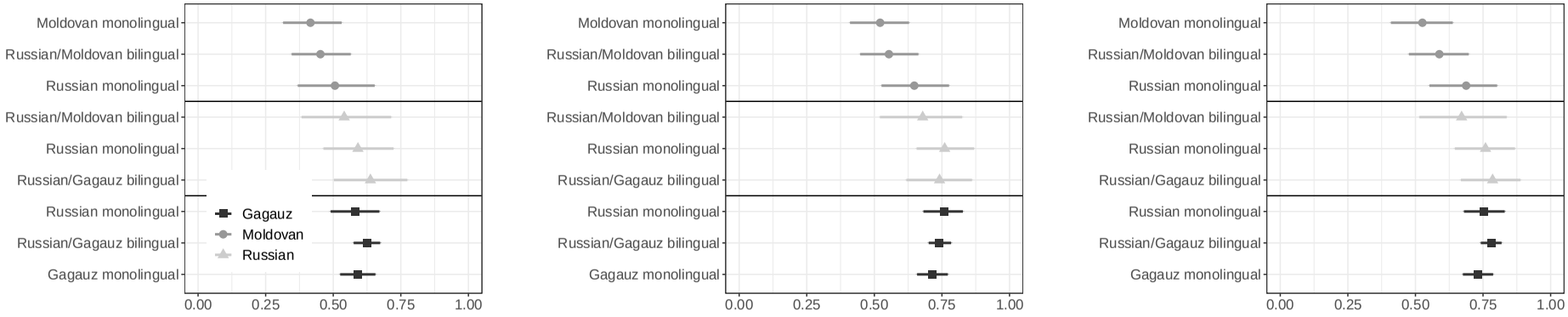



(a) Complete (b) Autonomy (c) Confederation

Points represent posterior median and horizontal lines 90 percent credible regions over 500 thousand iterations of four MCMC chains.



Figure I.5: Posterior probability that Gagauz respondents support integration with Russia




(a) Complete (b) Autonomy (c) Confederation

Points represent posterior median and horizontal lines 90 percent credible regions over 500 thousand iterations of four MCMC chains.

## I.2 Regression tables

Table I.1: Support for different separatist outcomes (Gagauzia).

| | Integration with Moldova | | | |
|---|---|---|---|---|
| | w/o Autonomy | Autonomy | Confederation | Independence |
| Intercept | -1.43 (-2.75, -0.22) | -0.29 (-1.46, 1.10) | 0.15 (-1.13, 1.43) | 0.83 (-0.56, 2.03) |
| Not fluent in Russian | 0.27 (0.04, 0.46) | -0.34 (-0.54, -0.14) | -0.47 (-0.69, -0.24) | -0.44 (-0.66, -0.23) |
| Fluent in Moldovan | -0.42 (-0.87, 0.01) | 0.07 (-0.31, 0.46) | -0.09 (-0.48, 0.34) | -0.19 (-0.55, 0.21) |
| Fluent in Gagauz | -0.20 (-0.46, 0.07) | -0.01 (-0.25, 0.27) | 0.07 (-0.19, 0.33) | 0.11 (-0.12, 0.39) |
| Not Gagauz | 0.54 (0.07, 1.03) | -0.46 (-0.88, -0.03) | -0.18 (-0.62, 0.27) | -0.40 (-0.86, 0.01) |
| Russian | -0.35 (-0.97, 0.24) | 0.24 (-0.31, 0.79) | 0.41 (-0.20, 1.00) | 0.25 (-0.32, 0.81) |
| Bulgarian | -0.37 (-0.96, 0.22) | 0.63 (0.01, 1.16) | -0.06 (-0.69, 0.62) | -0.26 (-0.89, 0.35) |
| Ukrainian | -0.12 (-0.64, 0.41) | 0.02 (-0.48, 0.52) | -0.13 (-0.67, 0.38) | 0.12 (-0.39, 0.62) |
| Other | -1.10 (-2.30, 0.07) | 0.79 (-0.40, 1.95) | 0.28 (-0.91, 1.54) | 0.38 (-0.71, 1.50) |
| ln(Income) | -0.15 (-0.32, -0.01) | 0.05 (-0.10, 0.19) | -0.25 (-0.41, -0.08) | -0.28 (-0.45, -0.11) |
| ln(Age) | 0.10 (-0.14, 0.33) | -0.05 (-0.26, 0.18) | 0.17 (-0.07, 0.41) | 0.02 (-0.23, 0.24) |
| Male | 0.01 (-0.15, 0.18) | -0.11 (-0.27, 0.05) | -0.01 (-0.19, 0.18) | 0.10 (-0.06, 0.28) |
| Higher education | -0.17 (-0.43, 0.06) | 0.12 (-0.11, 0.35) | 0.24 (-0.01, 0.50) | 0.11 (-0.13, 0.36) |
| Urban | 0.004 (-0.18, 0.20) | -0.13 (-0.31, 0.04) | -0.03 (-0.23, 0.17) | -0.05 (-0.25, 0.14) |
| | Integration with Russia | | | |
| | w/o Autonomy | Autonomy | Confederation | |
| Intercept | 0.03 (-1.19, 1.21) | 0.35 (-0.85, 1.60) | 0.26 (-1.06, 1.48) | |
| Not fluent in Russian | -0.09 (-0.31, 0.10) | -0.08 (-0.28, 0.13) | -0.16 (-0.36, 0.05) | |
| Fluent in Moldovan | -0.13 (-0.50, 0.26) | -0.24 (-0.61, 0.12) | -0.26 (-0.66, 0.11) | |
| Fluent in Gagauz | 0.12 (-0.13, 0.37) | -0.06 (-0.31, 0.19) | 0.09 (-0.17, 0.34) | |
| Not Gagauz | -0.19 (-0.62, 0.23) | -0.32 (-0.73, 0.08) | -0.20 (-0.65, 0.22) | |
| Russian | 0.22 (-0.31, 0.74) | 0.32 (-0.17, 0.89) | 0.22 (-0.36, 0.78) | |
| Bulgarian | -0.02 (-0.56, 0.57) | -0.10 (-0.66, 0.46) | -0.20 (-0.80, 0.40) | |
| Ukrainian | 0.86 (0.38, 1.37) | 0.85 (0.37, 1.36) | 0.51 (0.02, 1.03) | |
| Other | 0.03 (-1.02, 1.09) | 0.08 (-1.13, 1.35) | 0.04 (-1.23, 1.21) | |
| ln(Income) | -0.31 (-0.47, -0.15) | -0.27 (-0.43, -0.11) | -0.30 (-0.46, -0.13) | |
| ln(Age) | 0.25 (0.06, 0.49) | 0.23 (0.02, 0.46) | 0.30 (0.07, 0.54) | |
| Male | 0.08 (-0.09, 0.23) | 0.04 (-0.11, 0.21) | -0.01 (-0.18, 0.15) | |
| Higher education | -0.27 (-0.50, -0.04) | -0.35 (-0.57, -0.11) | -0.18 (-0.42, 0.09) | |
| Urban | 0.22 (0.04, 0.40) | 0.18 (0.004, 0.37) | -0.02 (-0.22, 0.17) | |
| | Thresholds | | | |
| $\gamma_1$ | -2.69 (-3.65, -1.86) | | | |
| $\gamma_2$ | -1.70 (-2.62, -0.85) | | | |
| $\gamma_3$ | -1.29 (-2.22, -0.44) | | | |
| $\gamma_4$ | -0.04 (-0.95, 0.83) | | | |

Posterior median and 90 percent credible regions over 500 thousand iterations of four MCMC chains (burn-in 50 thousand, thinned at 500 iterations).

Table I.2: Support for different separatist outcomes (Pridnestrovie),

| | Integration with Moldova | | | |
|---|---|---|---|---|
| | w/o Autonomy | Autonomy | Confederation | Independence |
| Intercept | -0.73 (-2.16, 0.65) | -1.16 (-2.51, 0.34) | -1.25 (-2.72, 0.10) | 1.40 (-0.14, 2.83) |
| Not fluent in Russian | 0.26 (-0.11, 0.65) | 0.39 (0.003, 0.75) | 0.09 (-0.26, 0.44) | -0.56 (-0.91, -0.16) |
| Fluent in Moldovan | 0.14 (-0.16, 0.49) | 0.29 (-0.05, 0.61) | -0.05 (-0.38, 0.28) | -0.07 (-0.39, 0.27) |
| Fluent in Ukrainian | -0.04 (-0.39, 0.30) | 0.04 (-0.30, 0.37) | 0.17 (-0.15, 0.50) | -0.11 (-0.44, 0.22) |
| Not Russian | 0.53 (0.21, 0.86) | 0.44 (0.10, 0.74) | 0.81 (0.49, 1.14) | -0.43 (-0.75, -0.09) |
| Ukrainian | -0.22 (-0.65, 0.17) | -0.14 (-0.53, 0.28) | -0.37 (-0.76, 0.04) | 0.22 (-0.22, 0.60) |
| Other | -0.02 (-0.55, 0.45) | 0.06 (-0.48, 0.54) | -0.18 (-0.68, 0.31) | 0.31 (-0.22, 0.79) |
| ln(Income) | -0.05 (-0.25, 0.16) | 0.03 (-0.17, 0.25) | 0.07 (-0.14, 0.29) | -0.16 (-0.36, 0.06) |
| ln(Age) | -0.14 (-0.37, 0.08) | -0.14 (-0.38, 0.08) | -0.15 (-0.38, 0.07) | 0.24 (0.03, 0.49) |
| Male | 0.05 (-0.12, 0.24) | -0.01 (-0.20, 0.16) | -0.04 (-0.22, 0.14) | 0.07 (-0.11, 0.25) |
| Higher education | -0.15 (-0.37, 0.07) | -0.22 (-0.43, 0.01) | -0.18 (-0.40, 0.03) | 0.34 (0.11, 0.56) |
| Urban | -0.07 (-0.25, 0.12) | -0.20 (-0.39, -0.03) | -0.23 (-0.40, -0.03) | 0.20 (0.02, 0.38) |
| | Integration with Russia | | | |
| | w/o Autonomy | Autonomy | Confederation | |
| Intercept | 1.07 (-0.28, 2.55) | 0.58 (-0.83, 1.96) | 0.24 (-1.07, 1.68) | |
| Not fluent in Russian | -0.30 (-0.66, 0.04) | -0.38 (-0.75, -0.03) | -0.42 (-0.77, -0.05) | |
| Fluent in Moldovan | -0.13 (-0.46, 0.23) | 0.02 (-0.31, 0.36) | -0.15 (-0.47, 0.21) | |
| Fluent in Ukrainian | -0.28 (-0.60, 0.03) | -0.25 (-0.57, 0.07) | -0.26 (-0.56, 0.06) | |
| Not Russian | -0.23 (-0.56, 0.11) | -0.21 (-0.53, 0.12) | -0.03 (-0.35, 0.31) | |
| Ukrainian | 0.39 (0.001, 0.80) | 0.61 (0.22, 0.99) | 0.23 (-0.18, 0.63) | |
| Other | 0.27 (-0.23, 0.77) | 0.58 (0.10, 1.08) | 0.38 (-0.13, 0.87) | |
| ln(Income) | -0.20 (-0.41, -0.01) | -0.11 (-0.31, 0.09) | -0.13 (-0.32, 0.07) | |
| ln(Age) | 0.10 (-0.11, 0.32) | 0.03 (-0.18, 0.25) | 0.16 (-0.06, 0.39) | |
| Male | 0.11 (-0.06, 0.28) | 0.03 (-0.15, 0.20) | 0.15 (-0.03, 0.33) | |
| Higher education | 0.28 (0.07, 0.48) | 0.28 (0.09, 0.50) | 0.53 (0.32, 0.74) | |
| Urban | 0.13 (-0.04, 0.30) | -0.01 (-0.18, 0.18) | 0.12 (-0.05, 0.31) | |
| | Integration with Ukraine | | | |
| | w/o Autonomy | Autonomy | Confederation | |
| Intercept | -0.44 (-1.75, 0.99) | 0.06 (-1.31, 1.38) | 0.37 (-0.93, 1.78) | |
| Not fluent in Russian | 0.58 (0.24, 0.93) | 0.26 (-0.09, 0.59) | 0.25 (-0.09, 0.61) | |
| Fluent in Moldovan | -0.12 (-0.45, 0.20) | -0.25 (-0.60, 0.07) | -0.50 (-0.82, -0.18) | |
| Fluent in Ukrainian | 0.09 (-0.26, 0.37) | 0.17 (-0.13, 0.49) | 0.21 (-0.08, 0.52) | |
| Not Russian | 0.22 (-0.13, 0.52) | 0.38 (0.03, 0.69) | 0.48 (0.18, 0.81) | |
| Ukrainian | 0.27 (-0.12, 0.66) | 0.09 (-0.32, 0.48) | 0.10 (-0.26, 0.50) | |
| Other | 0.17 (-0.30, 0.68) | 0.06 (-0.44, 0.54) | -0.15 (-0.64, 0.34) | |
| ln(Income) | -0.14 (-0.34, 0.06) | -0.15 (-0.34, 0.07) | -0.21 (-0.41, 0.002) | |
| ln(Age) | 0.02 (-0.21, 0.24) | -0.07 (-0.30, 0.15) | -0.02 (-0.24, 0.19) | |
| Male | -0.02 (-0.19, 0.16) | -0.07 (-0.24, 0.11) | 0.06 (-0.11, 0.24) | |
| Higher education | -0.04 (-0.25, 0.18) | -0.03 (-0.23, 0.18) | 0.02 (-0.19, 0.23) | |
| Urban | 0.12 (-0.06, 0.30) | 0.08 (-0.12, 0.25) | 0.12 (-0.06, 0.29) | |
| | Thresholds | | | |
| $\gamma_1$ | -1.33 (-2.30, -0.19) | | | |
| $\gamma_2$ | -0.64 (-1.68, 0.43) | | | |
| $\gamma_3$ | 0.10 (-0.94, 1.17) | | | |
| $\gamma_4$ | 0.93 (-0.08, 2.02) | | | |

Posterior median and 90 percent credible regions over 500 thousand iterations of four MCMC chains (burn-in 50 thousand, thinned at 500 iterations)

# J Additional experimental data

## J.1 T-test results

### J.1.1 Pridnestrovie

Table J.1 presents results from t-test analyses of experimental data from Pridnestrovie. Since theory predicts heterogenous treatment effects, I both report the population-level effects (first row), and the effects across groups with different levels of fluency in the relevant languages and forms of ethnic identification. More precisely, I dichotomize the four-point Likert scale question regarding the ideal outcome for Pridnestrovie-Moldovan relations (1 =separate states is the best outcome). I operationalize fluency in a given language and ethnicity in the same fashion as in regression analyses in text.

As expected given strong *a priori* evidence of heterogenous treatment effects, the treatments have minimal effect at the overall population level. However, results from t-tests on different population subgroups indicate that there are heterogenous treatment effects based on both ethnic identification and linguistic fluency. $T_{Moldovan}$ significantly reduces support for separatism among individuals who are fluent in Moldovan, ethnic Moldovans, and individuals who are not ethnic Russians. The populations of these subgroups obviously overlap: ethnic Moldovans are not Russian, and are also the residents of Pridnestrovie who are most likely to be fluent in Moldovan.

Results regarding $T_{Russian}$ and $T_{Ukrainian}$ are more in line with ethnic explanations of support for separatism. Both primes reduce support for Pridnestrovian independence among non-Russians; $T_{Russian}$ decreases support among ethnic Moldovans, while $T_{Ukrainian}$ reduces support for separatism among ethnic Ukrainians. However, $T_{Russian}$ also induces a lower level of support for separatism among respondents who are not fluent in Russian and who speak Moldovan, while $T_{Ukrainian}$ substantially reduces support for separatism non-speakers of Russian, as well as speakers of Moldovan and Ukrainian. Though the treatment effects on linguistic subgroups are not significant, they are evidence that the treatment results demand more rigorous analysis.

### J.1.2 Gagauzia

I report results from t-test analyses of the Gagauz data in a similar manner as with Pridnestrovie: I dichotomize the outcome variable to represent whether or not a respondent's preferred outcome for Gagauzia-Moldovan relations is separate statehood, and subdivide the population based on linguistic proficiency and primary ethnic identification. Table J.2 presents these results. Again, the treatments show little effect at the overall population level in terms of the odds that a respondent would support Gagauzia independence. However, there is strong evidence that the Gagauz language prime affects support for separatism among certain linguistic subgroups. Namely, $T_{Gagauz}$ decreases support for separatism among individuals who are fluent in Moldovan and individuals who are not fluent in Gagauz. Given that individuals who are fluent in Moldovan are unlikely to be fluent in Gagauz, disentangling the role of different types of linguistic fluency again demands explanation.

Table J.1: Proportion of respondents with different linguistic repertoires and ethnic backgrounds in Pridnestrovie who support independence.

| | Control | $T_{Moldovan}$ | | $T_{Russian}$ | | $T_{Ukrainian}$ | |
|---|---|---|---|---|---|---|---|
| | Percentage | Percentage | T-Statistic | Percentage | T-Statistic | Percentage | T-Statistic |
| Overall | 0.81 | 0.73 | 1.49 (0.14) | 0.75 | 1.05 (0.30) | 0.73 | 1.30 (0.20) |
| Language | | | | | | | |
| Fluent in Russian | 0.82 | 0.74 | 1.33 (0.18) | 0.79 | 0.49 (0.62) | 0.74 | 1.40 (0.16) |
| Not fluent in Russian | 0.67 | 0.25 | 1.45 (0.21) | 0.33 | 1.52 (0.15) | 0.50 | 0.32 (0.80) |
| Fluent in Moldovan | 0.72 | 0.35 | **2.71** (0.01) | 0.56 | 1.03 (0.31) | 0.60 | 0.86 (0.40) |
| Not fluent in Moldovan | 0.83 | 0.81 | 0.22 (0.83) | 0.79 | 0.73 (0.47) | 0.77 | 0.95 (0.35) |
| Fluent in Ukrainian | 0.70 | 0.62 | 0.60 (0.55) | 0.67 | 0.67 (0.26) | 0.50 | 1.40 (0.17) |
| Not fluent in Ukrainian | 0.82 | 0.75 | 1.22 (0.23) | 0.77 | 0.98 (0.33) | 0.80 | 0.47 (0.64) |
| Ethnicity | | | | | | | |
| Ethnic Russian | 0.89 | 0.88 | 0.05 (0.96) | 0.89 | -0.10 (0.92) | 0.90 | -0.26 (0.80) |
| Not ethnic Russian | 0.73 | 0.58 | **1.87** (0.06) | 0.59 | **1.68** (0.10) | 0.56 | **2.01** (0.05) |
| Ethnic Moldovan | 0.68 | 0.38 | **2.28** (0.03) | 0.43 | **1.91** (0.06) | 0.61 | 0.52 (0.61) |
| Not ethnic Moldovan | 0.84 | 0.81 | 0.54 (0.59) | 0.84 | 0.00 (1.00) | 0.76 | 1.37 (0.17) |
| Ethnic Ukrainian | 0.77 | 0.65 | 1.03 (0.30) | 0.71 | 0.49 (0.63) | 0.38 | **3.11** (0.00) |
| Not ethnic Ukrainian | 0.82 | 0.75 | 1.12 (0.26) | 0.76 | 0.99 (0.33) | 0.84 | -0.29 (0.77) |

The outcome is the dichotomized response from the experimental question regarding the ideal outcome for Pridnestrovie-Moldovan relations, with a one representing that the respondent supports Pridnestrovian independence above all other outcomes. Results highlighted in bold represent significant results at the $\alpha = 0.1$ level.

Table J.2: Proportion of respondents with different linguistic repertoires and ethnic backgrounds who support Gagauzia independence.

| | Control | $T_{Moldovan}$ | | $T_{Russian}$ | | $T_{Gagauz}$ | |
|---|---|---|---|---|---|---|---|
| | Percentage | Percentage | T-Statistic | Percentage | T-Statistic | Percentage | T-Statistic |
| Overall | 0.19 | 0.16 | 0.75 (0.45) | 0.17 | 0.50 (0.62) | 0.18 | 0.16 (0.87) |
| Language | | | | | | | |
| Fluent in Russian | 0.22 | 0.19 | 0.50 (0.62) | 0.19 | 0.53 (0.60) | 0.20 | 0.32 (0.75) |
| Not fluent in Russian | 0.11 | 0.06 | 0.91 (0.37) | 0.09 | 0.28 (0.78) | 0.13 | -0.26 (0.80) |
| Fluent in Moldovan | 0.21 | 0.09 | 0.90 (0.37) | 0.14 | 0.39 (0.70) | 0.00 | **2.19** (0.04) |
| Not fluent in Moldovan | 0.19 | 0.17 | 0.53 (0.59) | 0.17 | 0.40 (0.69) | 0.20 | -0.31 (0.76) |
| Fluent in Gagauz | 0.21 | 0.17 | 0.74 (0.46) | 0.20 | 0.25 (0.80) | 0.23 | -0.41 (0.68) |
| Not fluent in Gagauz | 0.14 | 0.13 | 0.24 (0.81) | 0.11 | 0.53 (0.60) | 0.03 | **2.20** (0.03) |
| Ethnicity | | | | | | | |
| Ethnic Russian | 0.08 | 0.33 | -1.14 (0.29) | 0.00 | 1 (0.34) | 0.00 | 1.00 (0.34) |
| Not ethnic Russian | 0.20 | 0.15 | 1.09 (0.28) | 0.17 | 0.54 (0.59) | 0.19 | 0.27 (0.79) |
| Ethnic Moldovan | 0.07 | 0.10 | -0.23 (0.82) | 0.00 | 1.00 (0.34) | 0.00 | 1.00 (0.34) |
| Not ethnic Moldovan | 0.20 | 0.16 | 0.80 (0.43) | 0.18 | 0.53 (0.59) | 0.19 | 0.11 (0.91) |
| Ethnic Gagauz | 0.22 | 0.17 | 1.10 (0.27) | 0.18 | 0.74 (0.46) | 0.20 | 0.29 (0.77) |
| Not ethnic Gagauz | 0.07 | 0.13 | -0.83 (0.41) | 0.13 | -0.88 (0.38) | 0.05 | 0.38 (0.70) |

The outcome is the dichotomized response from the experimental question regarding the ideal outcome for Gagauz-Moldovan relations, with a one representing that the respondent supports Gagauz independence above all other outcomes. Results highlighted in bold represent significant results at the $\alpha = 0.1$ level.

## J.2 Hierarchical modeling strategy

I model individual $i$'s support for separatism $y$ as an ordinal probit model in a fashion similar to that which I employed for the observational analysis. The probability that a respondent offers a given ordinal response is $\phi\{\tau_j - (\gamma_t + X_{ik}\alpha_{k_t} + Z_{im}\beta_m)\} - \phi\{\tau_{j-1} - (\gamma_t + X_{ik}\alpha_{k_t} + Z_{im}\beta_m)\}$, where $\phi$ is the CDF of a standard normal distribution and $\tau$ represents one of $j = 1, 2, 3$ thresholds, corresponding to the cut-off points for each of the four ordinal responses. The model diverges from a standard ordinal probit analysis in that $\alpha$ represents $t = 4$ vectors of $k$ coefficients each, which correspond to each of four experimental conditions $t$ (the control and three treatments, $T_{Moldovan}$, $T_{Russian}$ and $T_{Ukrainian/Gagauz}$). In turn, $\alpha_{kt} \sim N(\zeta_k, 1)$, where $\zeta$ is a vector of $k$ hyperparameter values for each $\alpha_t$ (each $\zeta_k \sim N(0, 1)$). $X$ remains an $i \times k$ matrix of the linguistic and ethnic covariates. Similarly, $\gamma$ represents treatment specific intercepts, each distributed $N(\psi, 1)$ with $\psi \sim N(0, 1)$.

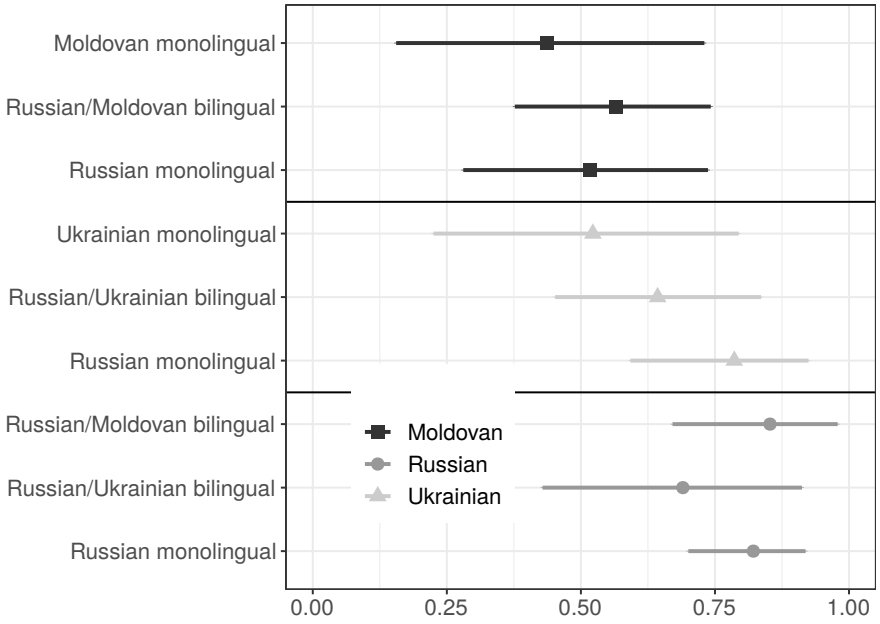
Finally, $\beta$ represents the $m = 5$ treatment-invariant coefficients for standard survey control variables $Z$, each distributed according to a standard normal distribution.

I run the experimental analyses in the same model as the observational and NA analyses, meaning that imputed right-hand side values are the same in each set of analyses. Again, each model is analysed using *runjags*. Each chain involves 500 thousand iterations and a 50 thousand iteration burn-in; I sample 1000 draws from each chain and assess convergence with the Gelman-Rubin diagnostic (all parameters' $\hat{r} < 1.1$).

## J.3 Additional results for experiment in Pridnestrovie

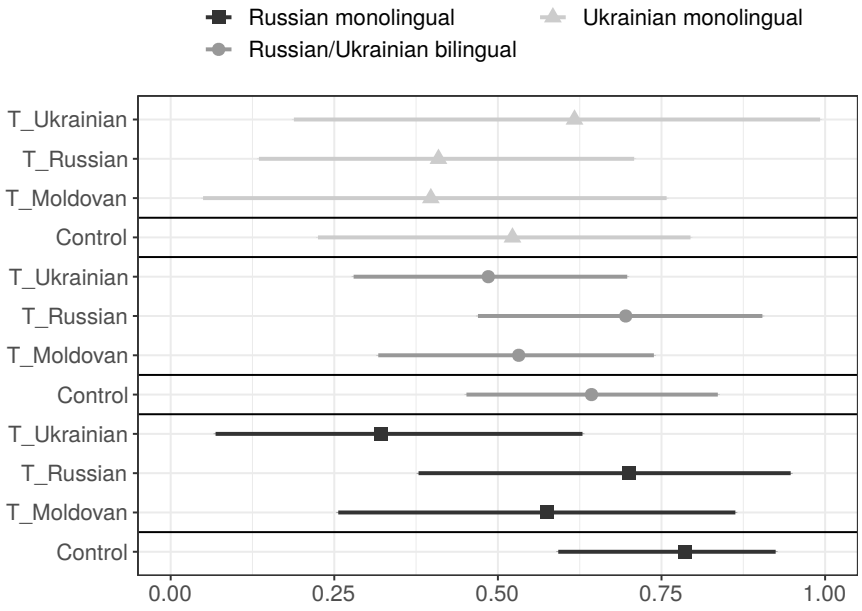
While expectations for the Ukrainian prime are ambiguous since Ukrainians are a secondary peripheral group, $T_{Ukrainian}$ has very different effects across Ukrainians: it substantially decreases support for separatism among monolingual Russian speakers, while appears to increase support for separatism among monolingual Ukrainian speakers. This heterogeneity is difficult to interpret, except insofar as it is evidence that the prime has no consistent effect across co-ethnics.

Figure J.1: Posterior probability of supporting Pridnestrovian independence in control condition.



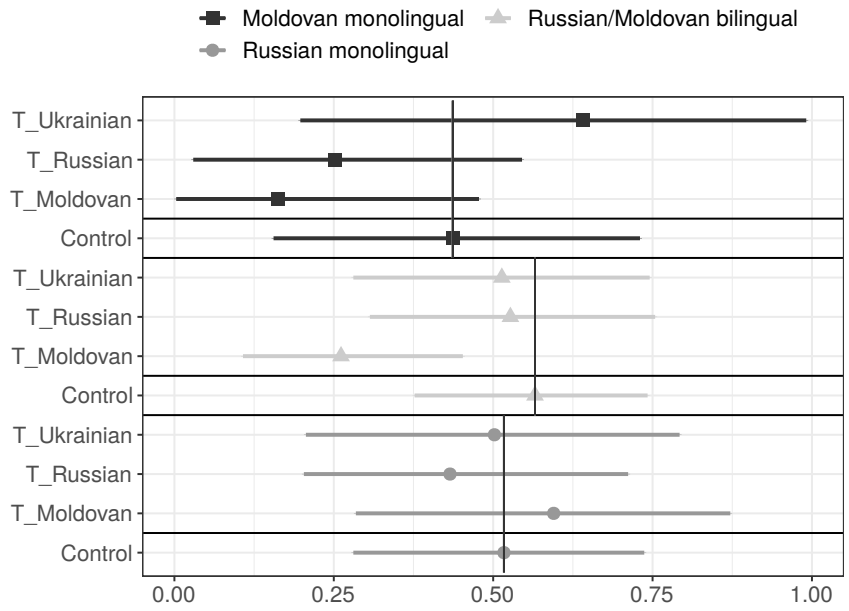

Points represent posterior median and horizontal lines 95 percent credible regions over 500 thousand iterations of four MCMC chains. Shading represents estimates for different ethnic groups; rows linguistic repertoires.

Figure J.2: Posterior probability ethnic Ukrainian supports Pridnestrovian independence across experimental conditions.



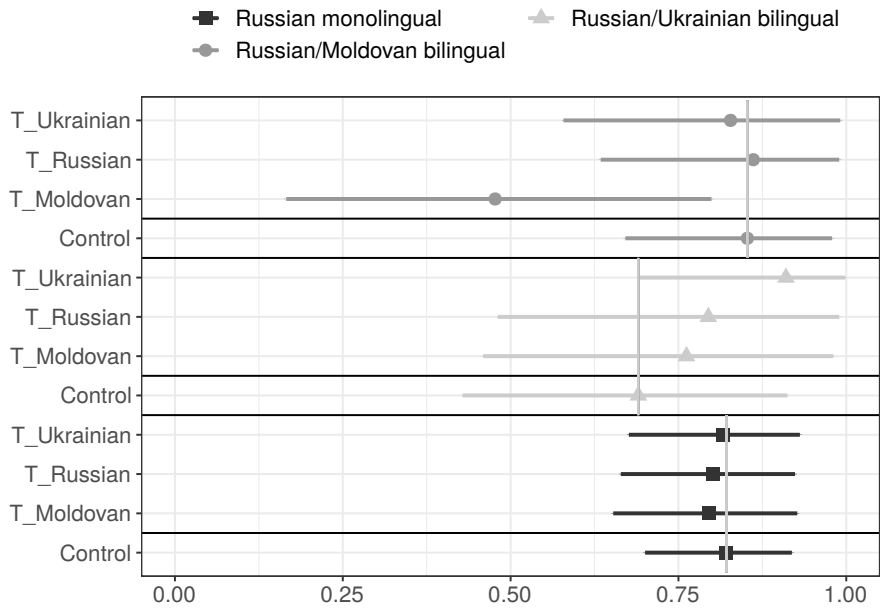

Points represent posterior median and horizontal lines 95 percent credible regions over 500 thousand iterations of four MCMC chains. Shading represents estimates for different linguistic repertoires; rows experimental conditions.

Figure J.3: Posterior probability ethnic Moldovan supports Pridnestrovian independence across experimental conditions (including $T_{Ukrainian}$).




Points represent posterior median and horizontal lines 95 percent credible regions over 500 thousand iterations of four MCMC chains. Shading represents estimates for different linguistic repertoires; rows experimental conditions.

Figure J.4: Posterior probability ethnic Russian supports Pridnestrovian independence across experimental conditions (including $T_{Ukrainian}$).



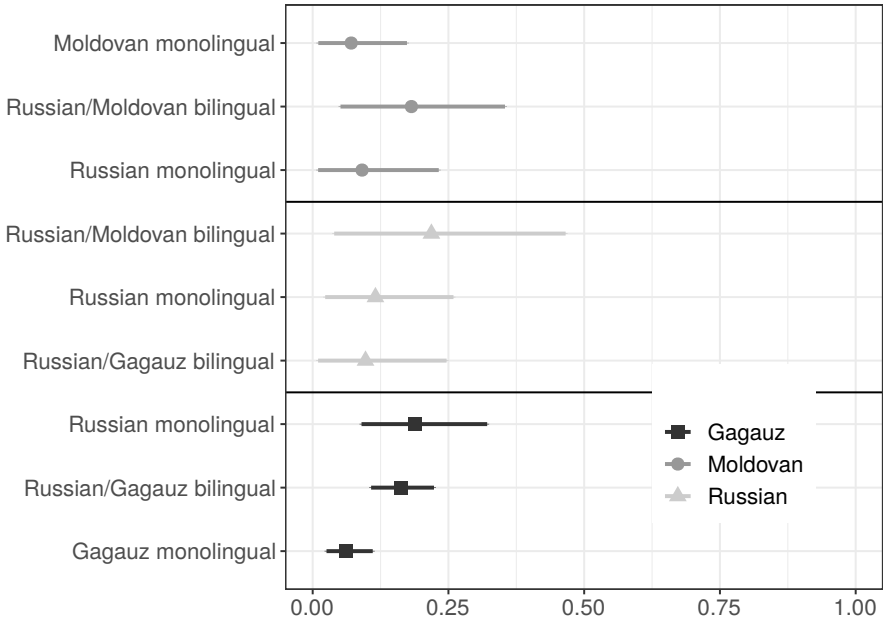

Points represent posterior median and horizontal lines 95 percent credible regions over 500 thousand iterations of four MCMC chains. Shading represents estimates for different linguistic repertoires; rows experimental conditions.

## J.4 Additional results for experiment in Gagauzia

Experimental results are generally low in magnitude and certainty with regard to ethnic Moldovans. In line with results regarding monolingual Russian-speaking Gagauz, $T_{Gagauz}$ decreases the median posterior probability that a Moldovan who is bilingual would support Gagauz independence, from 0.18 to 0.05; the probability is greater in 93 percent of MCMC draws. In addition, the effect of $T_{Moldovan}$ on monolingual Russian-speaking Moldovans is highly suggestive: it roughly doubles the median posterior probability of support for separatism, from 0.09 in the control to 0.17; the posterior probability is greater in $T_{Moldovan}$ in 76 percent of MCMC draws. This result is wholly in line with a linguistic theory of separatism: an ethnic Moldovan who only speaks Russian would become more aware of her limited opportunities in Moldova proper once primed to consider her lack of fluency in Moldovan.

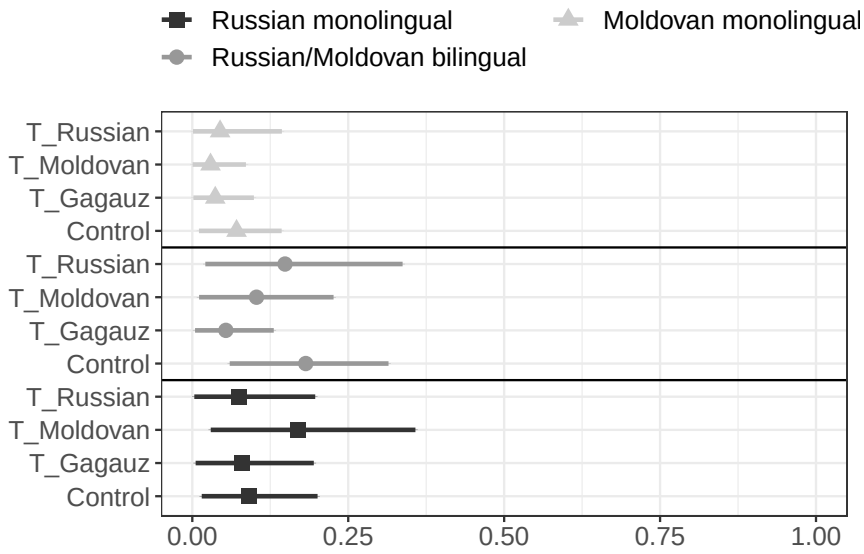
Again, though the sample size of Russians is low, results are in line with those from ethnic Moldovans. In particular, while $T_{Moldovan}$ increases support for separatism among all ethnic Russians, it does so to the greatest extent among non-speakers of Moldovan.

Figure J.5: Posterior probability respondents support regional independence in control condition.

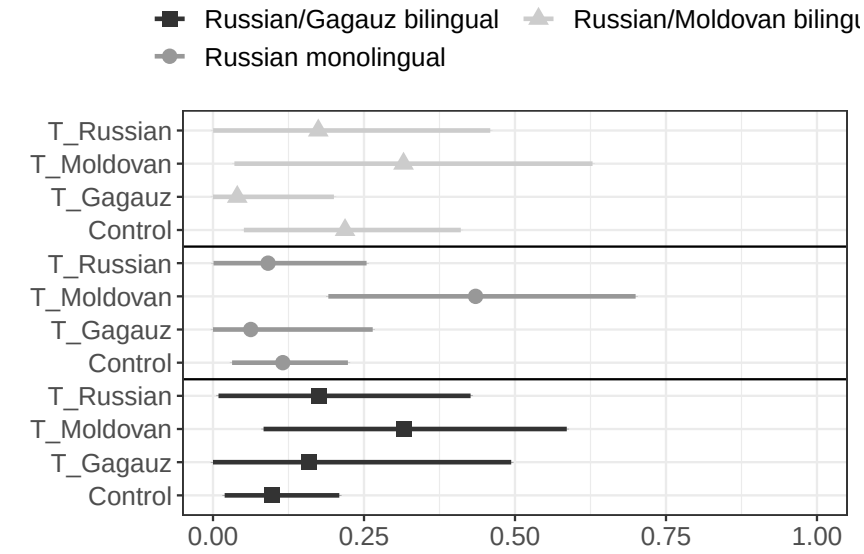



Points represent posterior median and horizontal lines 95 percent credible regions over 500 thousand iterations of four MCMC chains. Shading represents estimates for different ethnic groups; rows linguistic repertoires.

Figure J.6: Posterior probability of an ethnic Moldovan and Russian supporting Gagauz independence across experimental outcomes.




(a) Ethnic Moldovan




(b) Ethnic Russian

Points represent posterior median and horizontal lines 90 percent credible regions over 500 thousand iterations of four MCMC chains.

## J.5 Experiment regression tables

Table J.3: Ideal outcome for Pridnestrovian-Moldovan relations

| | Control | $T_{Moldovan}$ | $T_{Russian}$ | $T_{Ukrainian}$ |
|---|---|---|---|---|
| Intercept | -0.22 (-2.06, 1.30) | -0.32 (-2.01, 1.34) | -0.30 (-1.87, 1.45) | -0.25 (-1.91, 1.42) |
| Not fluent in Russian | -0.32 (-0.90, 0.30) | -0.34 (-1.19, 0.56) | -0.75 (-1.45, -0.10) | 0.34 (-0.79, 1.47) |
| Fluent in Moldovan | 0.13 (-0.42, 0.65) | -0.88 (-1.62, -0.25) | 0.24 (-0.42, 0.94) | 0.05 (-0.65, 0.67) |
| Fluent in Ukrainian | -0.42 (-1.01, 0.09) | -0.11 (-0.87, 0.66) | -0.02 (-0.79, 0.80) | 0.42 (-0.31, 1.18) |
| Not Russian | -0.89 (-1.43, -0.37) | -0.59 (-1.30, 0.15) | -1.01 (-1.62, -0.36) | -0.92 (-1.62, -0.20) |
| Ukrainian | 0.76 (0.10, 1.39) | -0.06 (-0.90, 0.78) | 0.69 (-0.13, 1.60) | -0.46 (-1.36, 0.36) |
| Other | 0.37 (-0.75, 1.53) | 1.33 (-0.02, 2.58) | 1.37 (-0.01, 2.75) | 1.54 (0.39, 2.89) |
| ln(Income) | -0.39 (-0.73, -0.002) | | | |
| ln(Age) | 0.01 (-0.30, 0.32) | | | |
| Male | -0.04 (-0.26, 0.18) | | | |
| Higher education | 0.43 (0.15, 0.73) | | | |
| Urban | 0.51 (0.27, 0.75) | | | |
| $\gamma_1$ | -5.66 (-9.09, -2.21) | | | |
| $\gamma_2$ | -4.71 (-8.34, -1.46) | | | |
| $\gamma_3$ | -4.09 (-7.70, -0.86) | | | |

Posterior median and 90 percent credible regions over 500 thousand iterations of four MCMC chains (burn-in 50 thousand, thinned at 500 iterations)

Table J.4: Ideal outcome for Gagauz-Moldovan relations

| | Control | $T_{Moldovan}$ | $T_{Russian}$ | $T_{Gagauz}$ |
|---|---|---|---|---|
| Intercept | 0.27 (-1.54, 1.91) | 0.46 (-1.31, 2.18) | -0.27 (-2.03, 1.50) | -0.35 (-2.16, 1.31) |
| Not fluent in Russian | -0.56 (-0.91, -0.19) | -0.62 (-1.07, -0.18) | -0.65 (-1.11, -0.21) | -0.18 (-0.59, 0.24) |
| Fluent in Moldovan | 0.42 (-0.19, 1.06) | -0.31 (-1.11, 0.57) | 0.39 (-0.58, 1.26) | -0.20 (-0.93, 0.52) |
| Fluent in Gagauz | -0.10 (-0.51, 0.34) | -0.32 (-0.91, 0.27) | 0.39 (-0.11, 0.96) | 0.56 (0.02, 1.08) |
| Not Gagauz | -0.44 (-1.23, 0.30) | -0.27 (-1.17, 0.51) | -0.02 (-0.91, 0.85) | 0.09 (-0.79, 0.99) |
| Russian | 0.13 (-0.71, 0.97) | 0.79 (-0.25, 1.83) | 0.10 (-1.07, 1.33) | -0.14 (-1.56, 1.16) |
| Bulgarian | 0.02 (-1.04, 0.98) | 0.22 (-0.90, 1.20) | 0.07 (-0.81, 1.06) | -0.16 (-1.69, 1.56) |
| Ukrainian | -0.03 (-0.96, 0.85) | -0.52 (-1.56, 0.49) | -0.13 (-1.05, 0.84) | -0.47 (-1.45, 0.51) |
| Other | 0.06 (-2.34, 2.34) | 0.07 (-2.19, 2.36) | 0.27 (-1.09, 1.70) | -0.05 (-1.70, 1.68) |
| ln(Income) | -0.05 (-0.22, 0.12) | | | |
| ln(Age) | -0.22 (-0.48, 0.04) | | | |
| Male | 0.28 (0.11, 0.46) | | | |
| Higher education | -0.06 (-0.31, 0.19) | | | |
| Urban | 0.15 (-0.06, 0.33) | | | |
| $\gamma_1$ | -2.42 (-4.72, -0.24) | | | |
| $\gamma_2$ | -0.54 (-2.74, 1.74) | | | |
| $\gamma_3$ | -0.03 (-2.28, 2.20) | | | |

Posterior median and 90 percent credible regions over 500 thousand iterations of four MCMC chains (burn-in 50 thousand, thinned at 500 iterations)

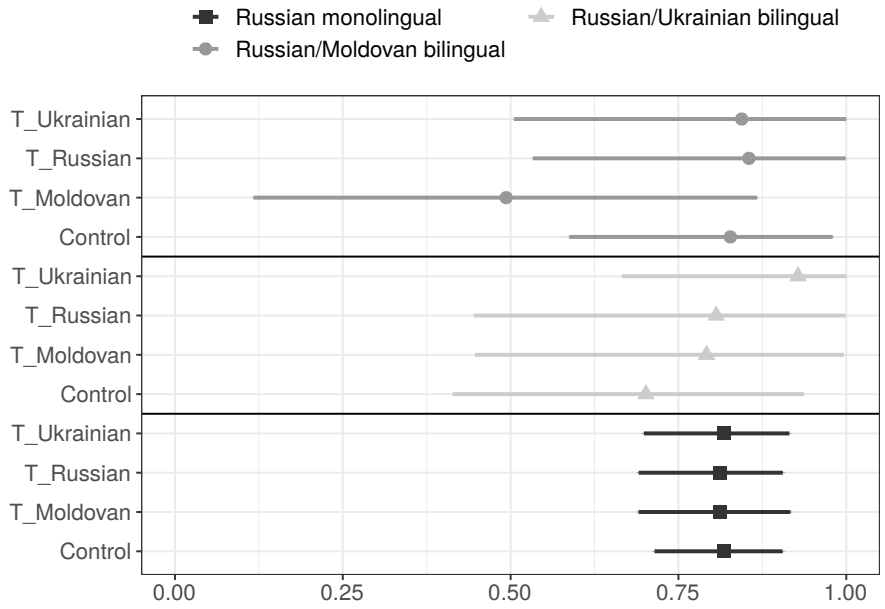
## J.6 Models with random hierarchical clustering variation

As a robustness check, I also ran models in which the variation about mean values for experimental coefficients was allowed to vary, as opposed to being set to one. Thus experimental

coefficients $\alpha$ for $k$ covariate and condition $t$ are distributed as follows: $\alpha_{kt} \sim N(\zeta_k, \lambda_k)$, where $\lambda_k \sim U(0, 100)$ as opposed to $\lambda_k = 1$. Results are robust to this parameterization.

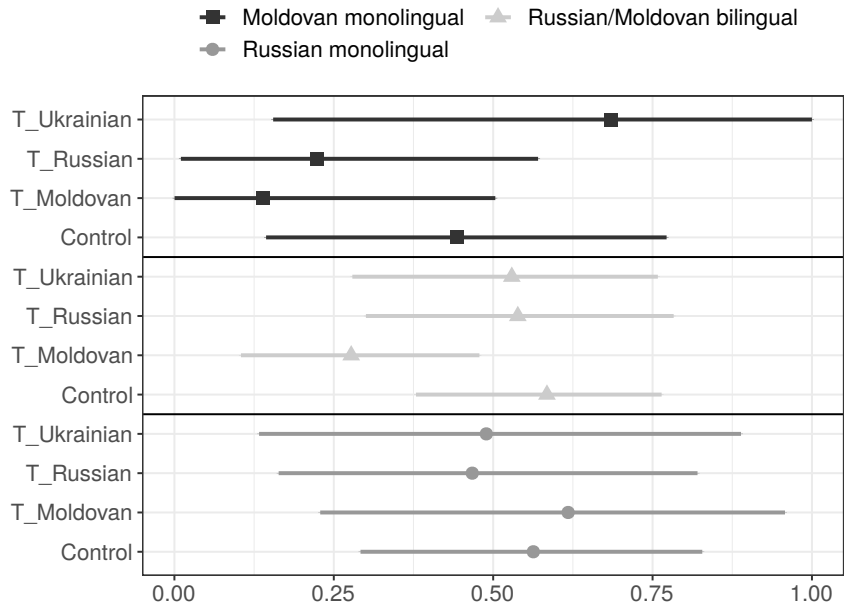
### J.6.1 Pridnestrovie results

Figure J.7: Posterior probability ethnic Russian supports Pridnestrovian independence across experimental conditions.



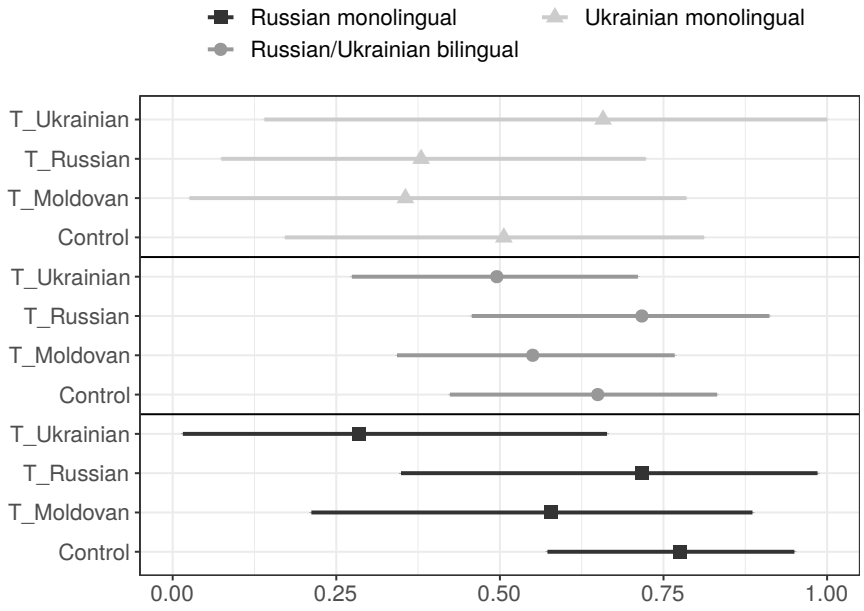

Points represent posterior median and horizontal lines 95 percent credible regions over 500 thousand iterations of four MCMC chains. Shading represents estimates for different linguistic repertoires; rows experimental conditions.

Figure J.8: Posterior probability ethnic Moldovan supports Pridnestrovian independence across experimental conditions.




Points represent posterior median and horizontal lines 95 percent credible regions over 500 thousand iterations of four MCMC chains. Shading represents estimates for different linguistic repertoires; rows experimental conditions.

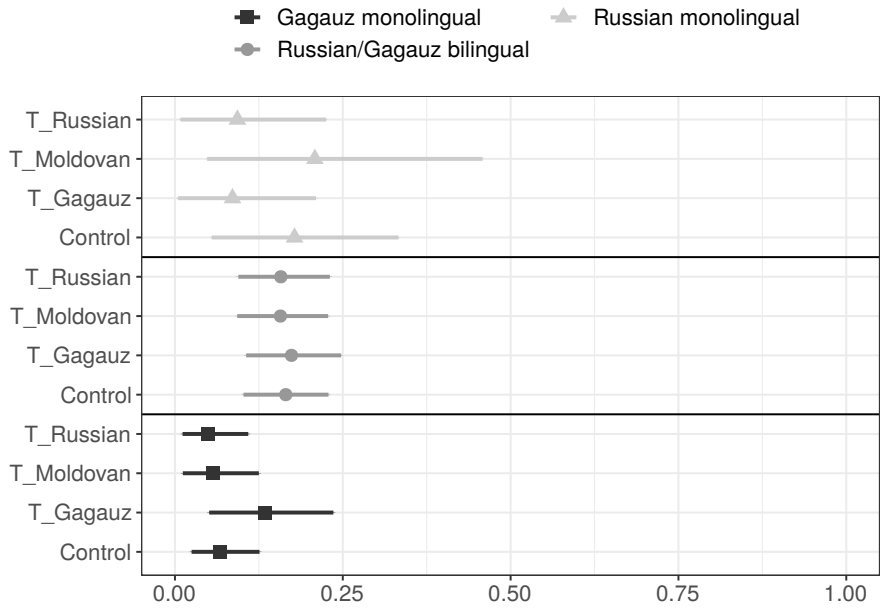
Figure J.9: Posterior probability ethnic Ukrainian supports Pridnestrovian independence across experimental conditions.




Points represent posterior median and horizontal lines 95 percent credible regions over 500 thousand iterations of four MCMC chains. Shading represents estimates for different linguistic repertoires; rows experimental conditions.

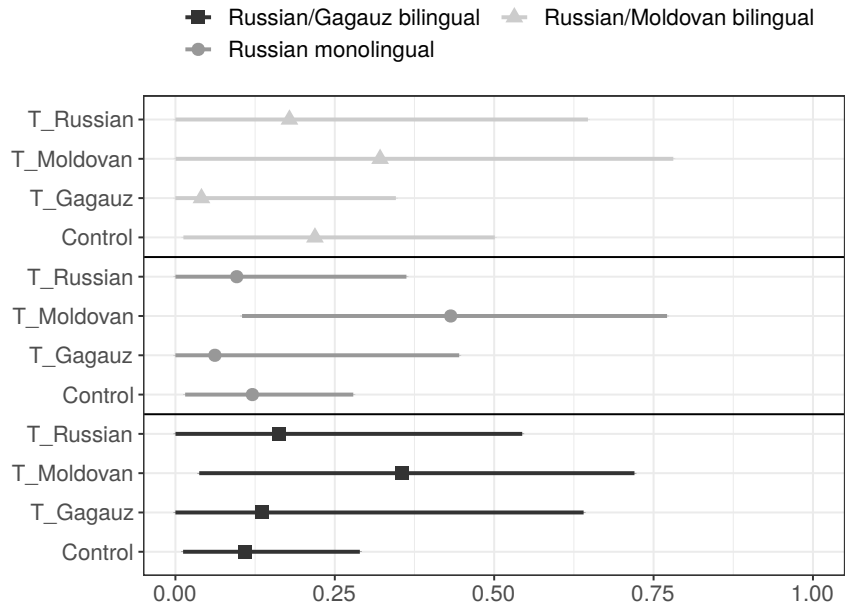
### J.6.2 Gagauzia results

Figure J.10: Posterior probability ethnic Gagauz supports Gagauz independence across experimental conditions.



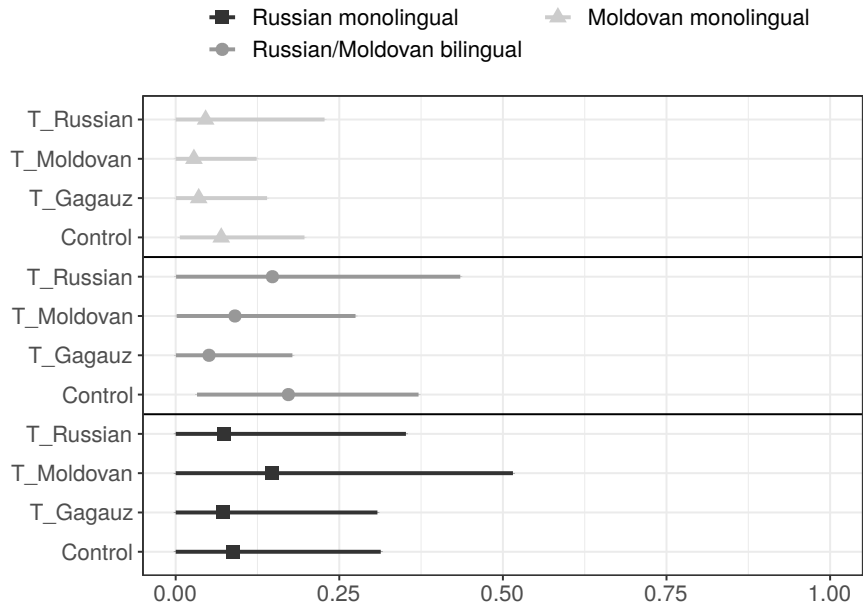

Points represent posterior median and horizontal lines 95 percent credible regions over 500 thousand iterations of four MCMC chains. Shading represents estimates for different linguistic repertoires; rows experimental conditions.

Figure J.11: Posterior probability ethnic Russian supports Gagauz independence across experimental conditions.




Points represent posterior median and horizontal lines 95 percent credible regions over 500 thousand iterations of four MCMC chains. Shading represents estimates for different linguistic repertoires; rows experimental conditions.

Figure J.12: Posterior probability ethnic Moldovan supports Gagauz independence across experimental conditions.




Points represent posterior median and horizontal lines 95 percent credible regions over 500 thousand iterations of four MCMC chains. Shading represents estimates for different linguistic repertoires; rows experimental conditions.

# K Analysis of nonresponse

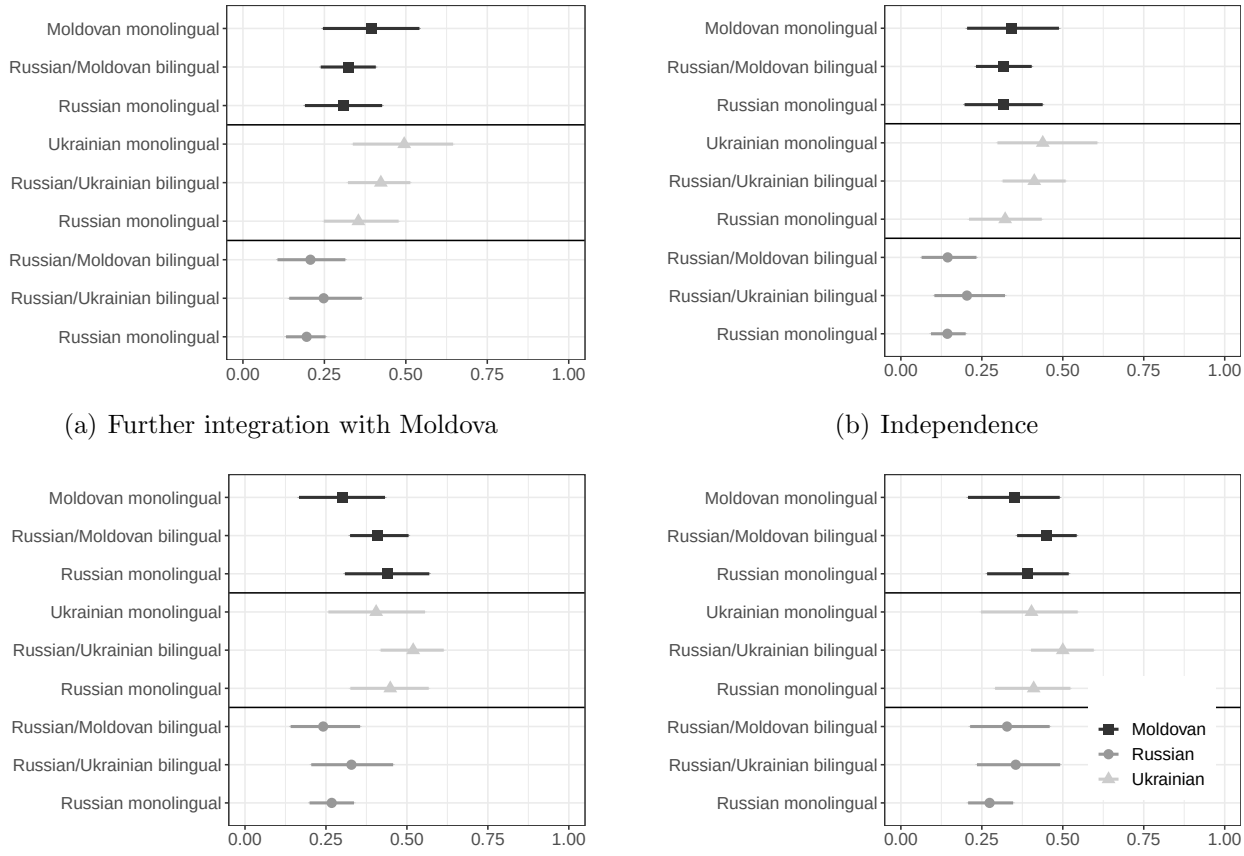
Item nonresponse on observational questions varies based on outcome question and region, ranging from 21 (support for independence) to 30 percent (support for confederation with Ukraine) in Pridnestrovie and 13 (support for autonomy in Moldova) to 31 percent (support for confederation with Moldova) in Gagauzia. However, only 10 percent of respondents in Pridnestrovie and three percent of respondents in Gagauzia refused to respond to all questions related to separatism (both observational and experimental), which indicates that the analysis of multiple outcomes ameliorates some concerns regarding item non-response. Furthermore, a substantial amount of nonresponse may be attributable to a lack of political knowledge or indifference, as opposed to political sensitivity. A high proportion of nonresponse took the form of "Don't know" responses, which were coded as distinct from declining to respond (Appendix G). "Don't know" rates ranged from 85 to 90 percent of all nonresponse to ordinal outcomes in Gagauzia. In Pridnestrovie, the equivalent percent range is 57 to 66 percent.

I analyze item nonresponse regarding both observational and experimental data for both Pridnestrovie and Gagauzia. Specifically, I present the posterior probability that members of different ethnic groups with different linguistic repertoires would not respond to questions regarding the separatist outcomes which I report in the main text. To do so, I conduct Bayesian probit analyses of the relationship between the variables analyzed in the text (linguistic and ethnic characteristics, controls) and a dichotomous indicator of item nonresponse for each variable. The models are thus of a similar form to those in the paper, with nonresponse as the outcome as opposed to the Likert-scale support for separatism variables.

In general, patterns of nonresponse indicate that, if anything, the reported findings may underestimate the relationship between language and support for separatism, especially in Gagauzia.

I run these analyses together with the observational and experiment analyses. Each of the fours MCMC chains thus involves 500 thousand iterations after a 50 thousand iteration burn-in; I sample 1000 draws from each chain and assess convergence with the Gelman-Rubin diagnostic (all parameters' $\hat{r} < 1.1$).

Figure K.1: Posterior probability of nonresponse regarding different separatist outcomes in Pridnestrovie




(a) Further integration with Moldova

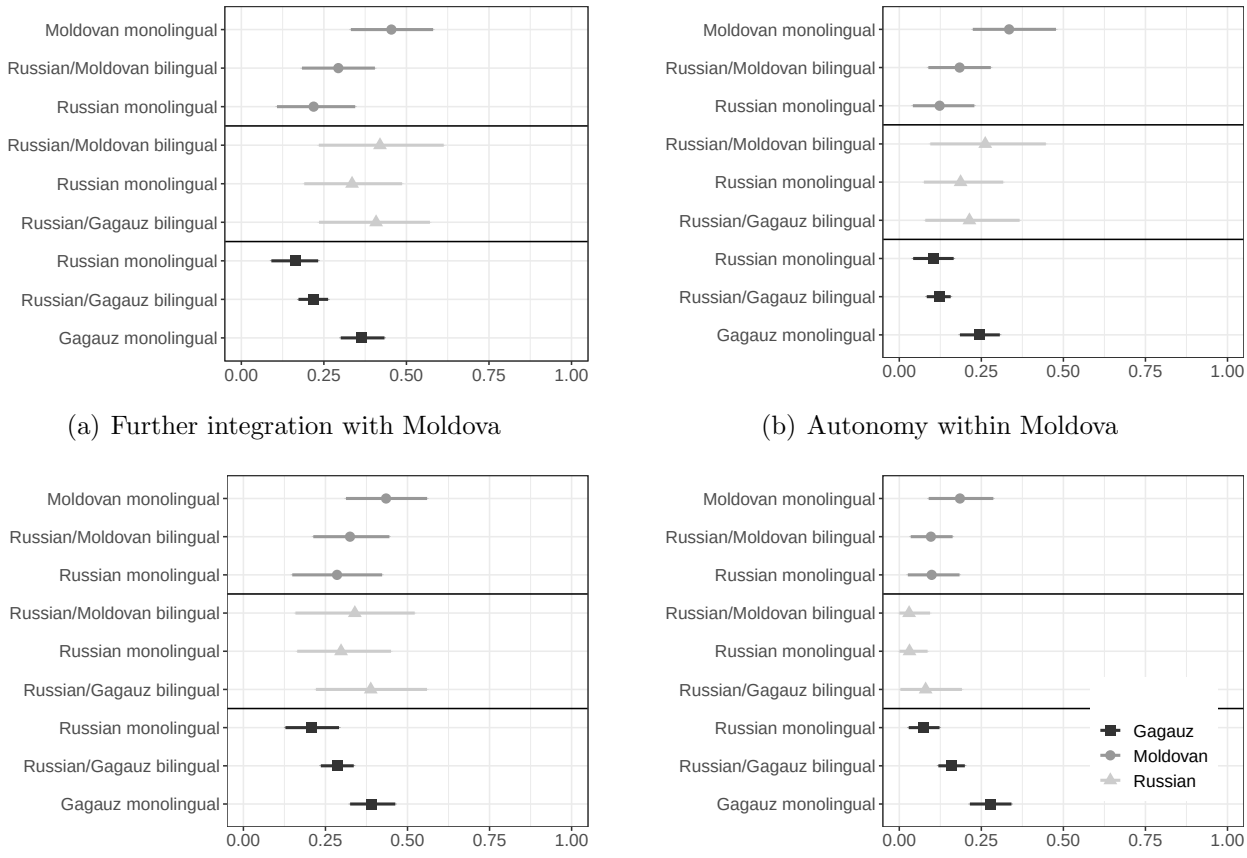
(b) Independence

(c) Integration with Ukraine

(d) Integration with Russia

Points represent posterior median and horizontal lines 90 percent credible regions over 500 thousand iterations of four MCMC chains.

Figure K.2: Posterior probability of nonresponse regarding different separatist outcomes in Gagauzia




(a) Further integration with Moldova

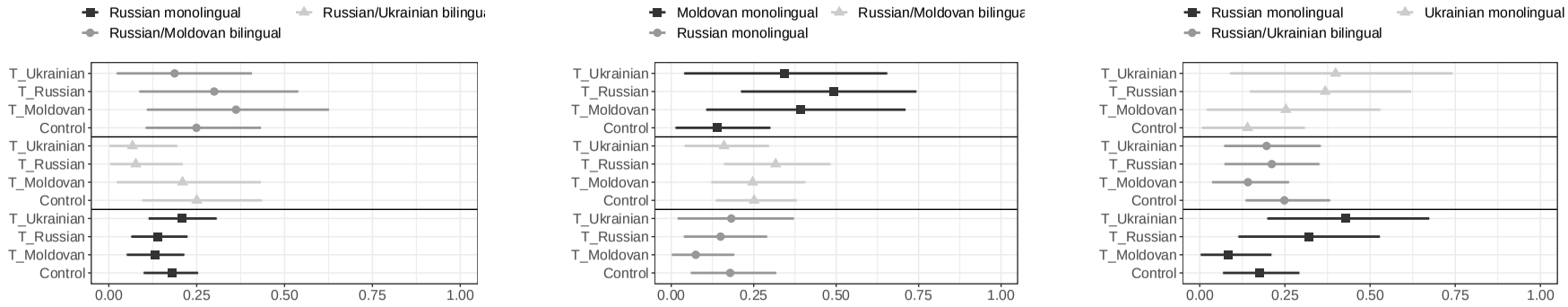
(b) Autonomy within Moldova

(c) Independence

(d) Integration with Russia

Points represent posterior median and horizontal lines 90 percent credible regions over 500 thousand iterations of four MCMC chains.

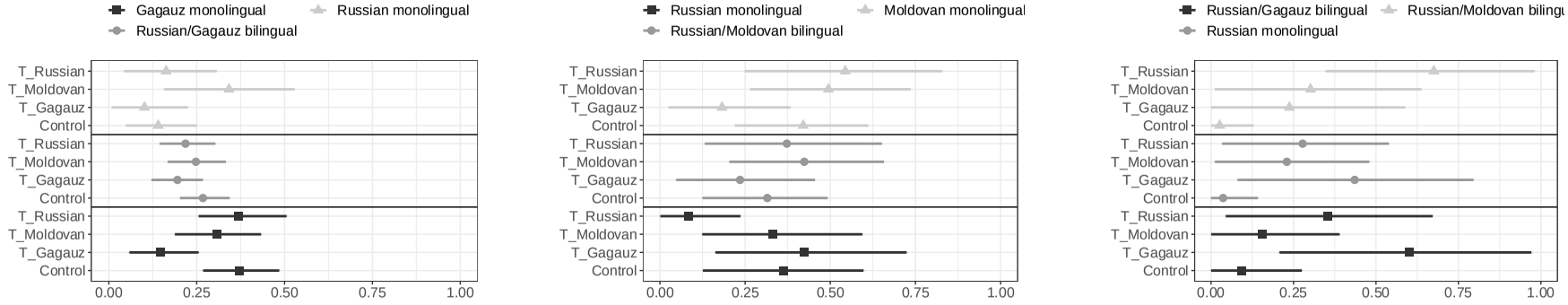
Figure K.3: Posterior probability of nonresponse regarding experimental outcome in Pridnestrovie




(a) Ethnic Russian (b) Ethnic Moldovan (c) Ethnic Ukrainian

Points represent posterior median and horizontal lines 90 percent credible regions over 500 thousand iterations of four MCMC chains.

Figure K.4: Posterior probability of nonresponse regarding experimental outcome in Gagauzia




(a) Ethnic Gagauz (b) Ethnic Moldovan (c) Ethnic Russian

Points represent posterior median and horizontal lines 90 percent credible regions over 500 thousand iterations of four MCMC chains.

# L Russian influence

Russia plays an integral role in the politics of both Gagauzia and Pridnestrovie, with many residents of both regions traveling to Russia as students and guest workers, and many residents of Pridnestrovie holding Russian citizenship. Such ties could influence support for separatism in three ways. First, the residents of these regions who support different forms of separatism may, in fact, support these outcomes as a means to facilitate political incorporation into Russia. Second, spending time in Russia could both increase proficiency in Russian and support for separatism. Third, it is possible that language is just a proxy for consumption of Russian media, which is generally pro-Russia. I discuss the first two concerns here, while Appendix M discusses the relationship between media consumption and support for separatism.

## L.1 Support for integration with Russia

Integration with Russia is one of the most preferred outcomes for regional status in both regions: 51 percent of respondents in Pridnestrovie support integration with Russia without autonomy, while 64 percent of respondents in Gagauzia support this outcome. However, a higher proportion of respondents support the status quo over integration with Russia (84 percent of Pridnestrovian respondents support independence; 88 percent of Gagauz respondents support autonomy in Moldova), which indicates that the potential benefits of greater political integration with Russia are perceived to be offset by other costs.

That said, I also investigate the correlates of support for integration with Russia (Appendix I, Figures I.2 and I.3). Particularly in case of Pridnestrovie, the results are wholly line with linguistic explanations of separatism. Since Russian is the main language of Russia, fluent Russian speakers should be the most supportive of integration with Russia since they would have the greatest opportunities there. Indeed, Russian proficiency is the strongest correlate of support for integration with Russia; though respondents who are proficient in either Moldovan and Ukrainian also tend to be less supportive of this outcome.

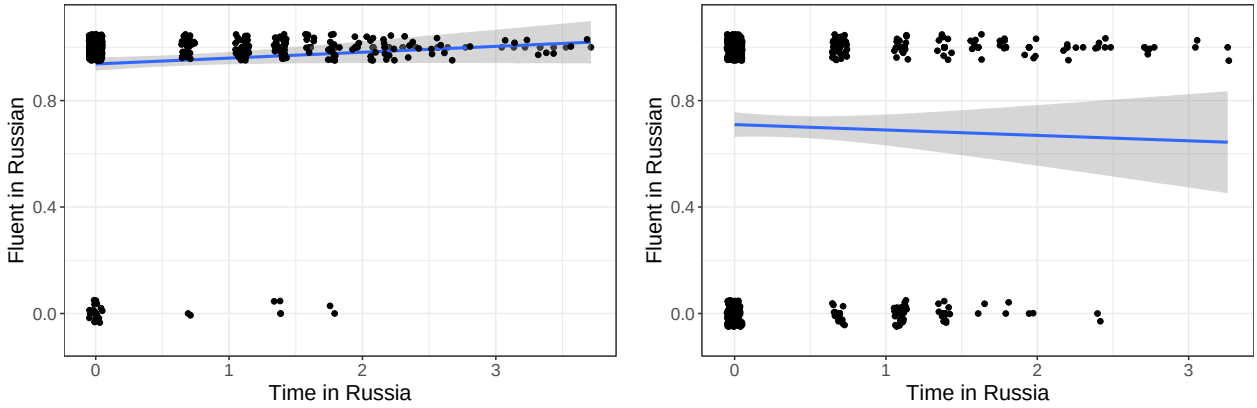
## L.2 Effect of time spent in Russia

Since it is possible that spending time in Russia increases both proficiency in Russian and support for separatism, I investigate these relationships using data from the 2013 survey. The analyses unambiguously demonstrate that there is little relationship between time spent in Russia and self-reported fluency in the Russian language, but are more ambiguous with regard to the relationship between time spent in Russia and support for separatism.

Figure L.1 illustrates the relationship between the log-transformed amount of time a respondent reported spending in Russia and her self-reported fluency in Russian.[1] There is little apparent relationship in either case, an unsurprising result since Russian is widely used in both Gagauzia and Pridnestrovie and respondents would thus have ample opportunities to develop Russian-language fluency in both regions.

[1]Respondents reported both if they cumulatively spent six months or more in Russia and, if so, how much time they spent there; I transform these data using the natural logarithm of years plus one.

Figure L.1: Relationship between time spent in Russia and fluency in Russian

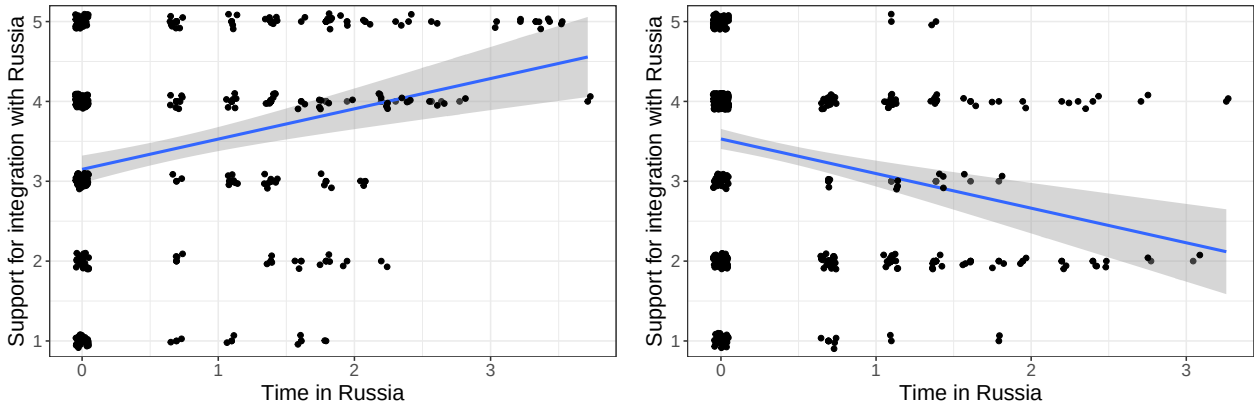



(a) Pridnestrovie (b) Gagauzia

Figure L.2 shows the relationship between time spent in Russia and support for different separatist outcomes: integration with Russia without autonomy (top row) and independence (bottom row). Pridnestrovian respondents are in the left column, those from Gagauzia on the right. Two aspects of these results are of particular interest. First, while the relationship between time in Russia and support for different forms of separatism is positive in Pridnestrovie, in Gagauzia it is negative. This result indicates that spending time in Russia does not necessarily make an individuals from separatist regions more supportive of separatism.

Second, even the case of Pridnestrovie, where time spent in Russia has a positive relationship with support for separatism, the strength of this relationship varies. Specifically, in Pridnestrovie time spent in Russia has a stronger relationship with support for integration with Russia (top row) than with support for regional independence (bottom row). In other words, the relationship between time spent in Russia and support for separatism is not uniform across separatist outcomes and has the strongest relationship with the outcome with which it is most closely linked.

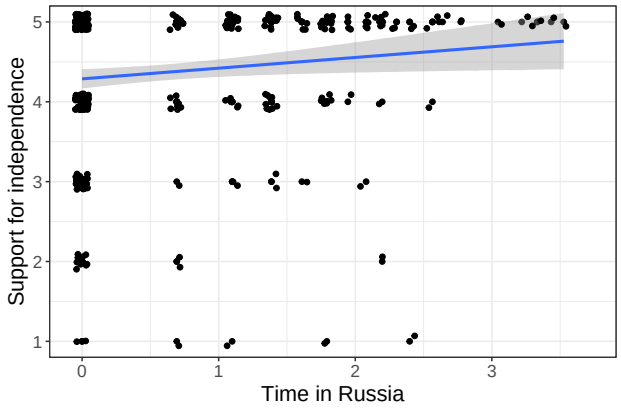
Figure L.2: Relationship between time spent in Russia and support for separatist outcomes



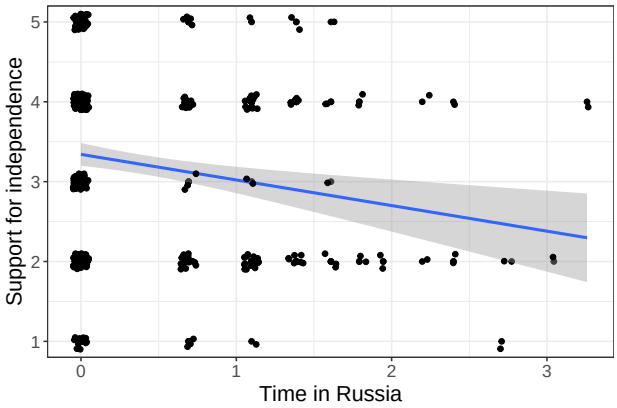

(a) Pridnestrovie




(b) Gagauzia




(c) Pridnestrovie




(d) Gagauzia

# M Television consumption analysis

It is possible that the relationship between language and support for separatism is due to media consumption: consumption of pro-Russian media makes respondents more supportive of separatist outcomes congruent with Russian interests. In this scenario, Russian speakers are more supportive of separatism because they consume Russian media. In line with this argument, Peisakhin & Rozenas (2018) demonstrate that exposure to Russian media significantly increases support for pro-Russian parties in Ukraine.[2]

To analyze whether or not media consumption is driving the results in the text, I conduct three analyses. First, I compare patterns of media consumption to linguistic proficiency. Second, I rerun the observational analyses, replacing measures of linguistic proficiency with measures of media consumption. Third, I use boxplots to illustrate the relationship between Russian fluency, media consumption, and support for separatist outcomes.

The analyses indicate that there is a relationship between the language in which a respondent primarily consumes media and support for separatism. However, this relationship appears to be distinct from that of linguistic fluency. Given the relatively small sample size and high correlation between linguistic fluency and language-of-media-consumption, these conclusions are highly tentative and represent an interesting avenue for future research.

## M.1 Media consumption and linguistic abilities

Table M.1 shows the relationship between self-reported fluency in a given language and whether or not a respondent reported watching television most frequently in that language.[3] Several aspects of the table are of special note. First, and unsurprisingly, it is very rare for respondents to report watching media most frequently in a language in which they are not fluent: with the exception of Russian in Gagauzia, where 24 percent of the population does not speak Russian fluently but still watches Russian most frequently,[4] between one and four percent of the sample watches television most frequently in a language in which they are not fluent. Second, fluent speakers of languages other than Russian are more likely to watch media in another language (almost always Russian). For example, 14 percent of the population of Pridnestrovie is fluent in Moldovan but watches media primarily in another language, compared to six percent of the population which is both fluent in Moldovan and primarily watches media in Moldovan. This relationship is likely due to the fact that most residents of both Pridnestrovie and Gagauzia are fluent in Russian, allowing them to watch

[2]Peisakhin & Rozenas (2018) also demonstrate that the relationship between media consumption and political attitudes is conditional upon language usage: there is little relationship between consuming Russian media and pro-Russian sentiment among Ukrainians who mostly speak Ukrainian.

[3]The survey included a five-point Likert scale question about the frequency which a respondent reported watching television in each of the three relevant regional languages. I code the language in which the respondent reported watching television most frequently as her preferred media language. In the event that a respondent reported watching television in two or more languages with equal frequency, I code all as being the most frequently watched. The only exception to this rule regards respondents who reported never watching television, whom I coded as having no preferred media language.

[4]This relatively large percentage may be due to the fact that Gagauz-language media is relatively undeveloped, and most residents of Gagauzia do not speak Moldovan well, leaving Russian-language media as the only option.

Table M.1: Relationship between language abilities and media consumption

| Pridnestrovie | | | | | | |
|---|---|---|---|---|---|---|
| | Russian media | | Moldovan media | | Ukrainian media | |
| | N | Y | N | Y | N | Y |
| Not fluent in language | 0.02 | 0.04 | 0.78 | 0.01 | 0.79 | 0.02 |
| Fluent in language | 0.03 | 0.90 | 0.14 | 0.06 | 0.12 | 0.07 |
| Gagauzia | | | | | | |
| | Russian media | | Moldovan media | | Gagauz media | |
| | N | Y | N | Y | N | Y |
| Not fluent in language | 0.02 | 0.24 | 0.91 | 0.01 | 0.25 | 0.01 |
| Fluent in language | 0.02 | 0.73 | 0.05 | 0.03 | 0.64 | 0.09 |

media in that language.

Overall, the table indicates that media consumption is unsurprisingly correlated with linguistic abilities (fluent speakers of a language are the most likely to consume media primarily in that language), but the correlation is far from perfect.

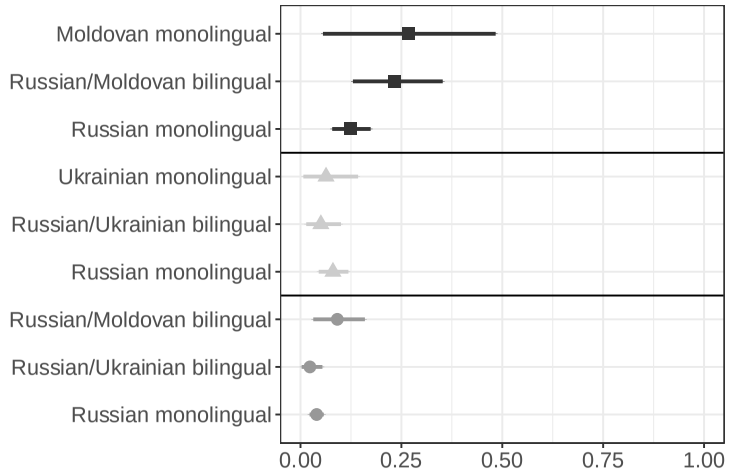
## M.2 Media consumption and support for separatism

Figures M.1 and M.3 present a replication of the observational analyses from the text, replacing indicators of proficiency in different languages with indicators of primary language of media consumption. In this context, "monolingual" represents a respondent who has only one primary language of media consumption, whereas "bilingual" represents a respondent with two primary languages of media consumption. As with the analyses in the text, the graphics are divided first by ethnicity, then by language of media consumption.

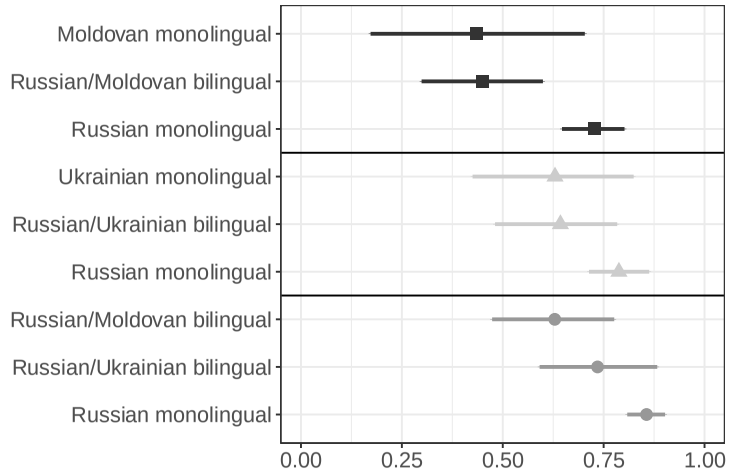
### M.2.1 Pridnestrovie

In Pridnestrovie (Figure M.1), the relationship between media consumption and support for separatist outcomes diverges somewhat from the relationship between linguistic fluency and support for separatism. Whereas *not* being fluent in Russian was the strongest predictor of support for separatism in the observational analyses in the text, in these models watching television primarily in a language other than Russian (or together with Russian), tends to be the strongest predictor of separatist sentiment. Relative to respondents who primarily watch television only in Russian, respondents who primarily watch television in Moldovan tend to be more supportive of integration with Moldova, less supportive of regional independence, and (surprisingly and with high levels of uncertainty) more supportive of integration with Russia. The first two results are in line with theory, while the third runs opposite to theory: watching Moldovan media should make respondents more supportive of integration with Moldova and opposed to both independence and integration with Russia.

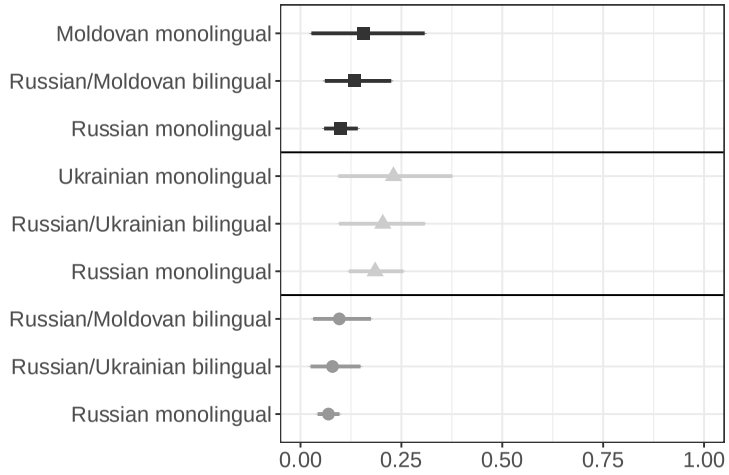
Figure M.1: Posterior probability that Pridnestrovian respondents with different patterns of media consumption support different separatist outcomes



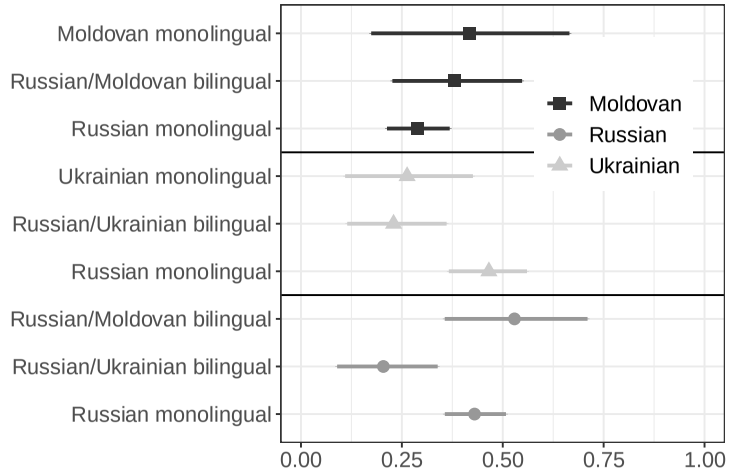

(a) Integration with Moldova




(b) Independence




(c) Integration with Ukraine




(d) Integration with Russia

Points represent posterior median and horizontal lines 90 percent credible regions over 500 thousand iterations of four MCMC chains.

Similarly, respondents who primarily consume media in Ukrainian tend to be less supportive of independence and integration with Russia, relative to respondents who primarily consume media only in Russian. Both of these results are in line with theoretical expectations—even prior to the annexation of Crimea, Ukrainian media was more skeptical of Russia than Russian media. Interestingly, there seems to be little relationship between media consumption and support for integration with Ukraine, perhaps because there is little support in Ukraine for this outcome.

Parsing these results vis-á-vis linguistic capabilities is made difficult due to the relatively small sample size and correlation between linguistic fluency and TV consumption. However, Figure M.2 presents some highly tentative evidence suggesting that TV consumption and linguistic fluency do not necessarily have the same relationship with support for separatist outcomes. The figure in the top row shows the relationship between Russian fluency, Moldovan-language media consumption, and support for greater integration with Moldova. In the case of this outcome, it appears that both fluency in Russian and consuming Moldovan-language media have a relationship with support for separatism: though there are too few observations to make claims about respondents who both are not fluent in Russian and watch TV in Moldovan, fluent Russian speakers are less supportive of integration with Moldova than non-fluent speakers; among fluent speakers of Russian, those that watch TV in Moldovan are more supportive of integration with Moldova.

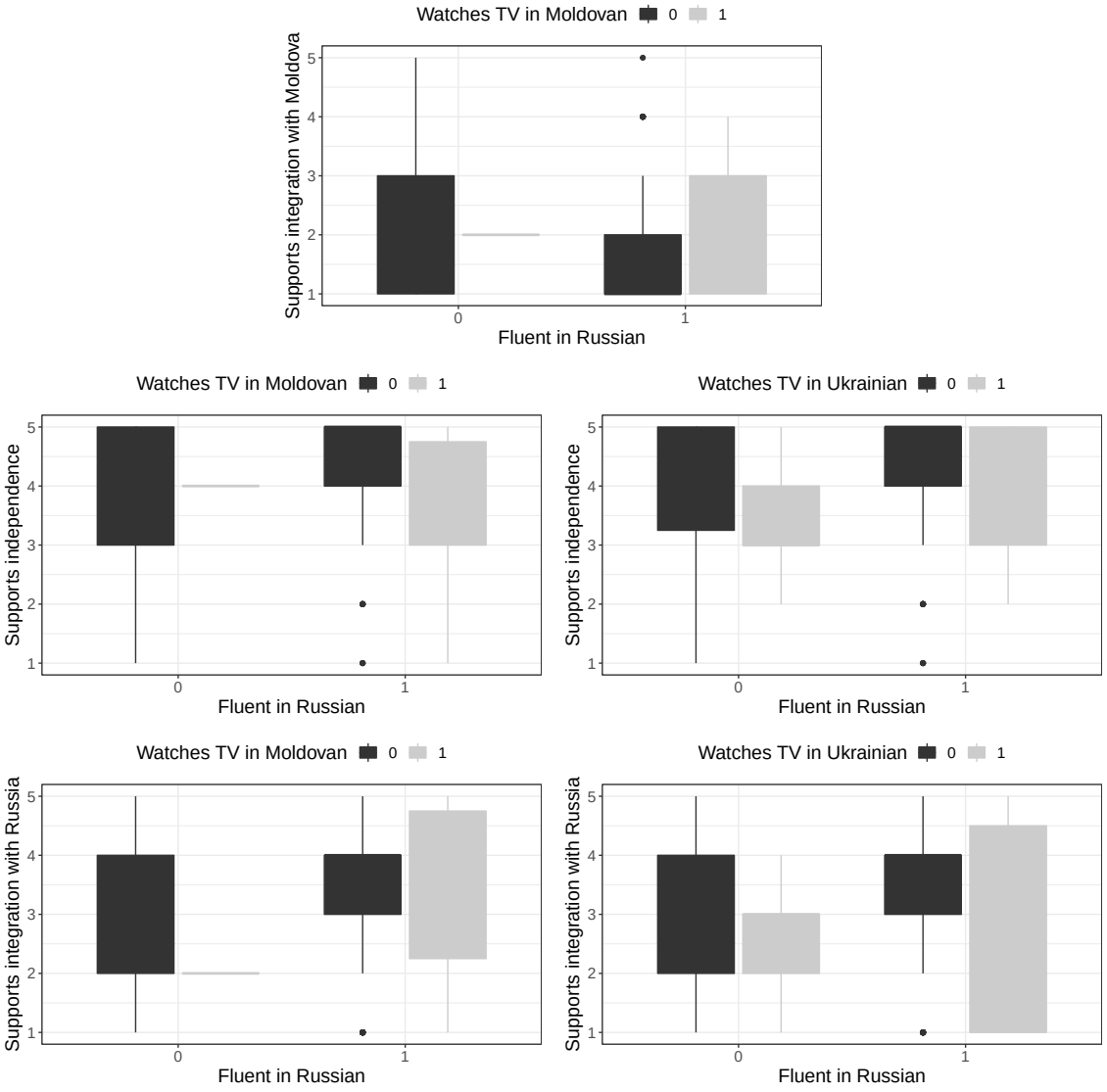
The middle row shows similar graphics for support for Pridnestrovian independence, by Moldovan-language media consumption (left) and Ukrainian-language media consumption (right). With regard to Moldovan-language media consumption, the results are largely in line with those for support for integration with Moldova: fluent speakers of Russian are more supportive of this separatist outcome than non-speakers and, among fluent speakers of Russian, those who consume Moldovan-language media are less supportive of independence.

The results for Moldovan largely coincide with those that regard to Ukrainian-language media consumption: respondents who watch Ukrainian-language TV tend to be less supportive of this separatist outcome, regardless of Russian fluency; fluent Russian speakers tend to be more supportive of Pridnestrovian independence than respondents who are not fluent.

The bottom row uses the same set of graphics, but with support for integration with Russia as the outcome. Results for both Moldovan- and Ukrainian-language consumption are relatively ambigious. However, respondents fluent in Russian are, again, more supportive of this outcome than those who are not fluent in Russian.

Taken as a whole, these results indicate that both media consumption and Russian-language abilities can influence support for separatism, and that these factors are not proxies for each other. However, all results are highly tentative and analysis with a larger sample size would be necessary for a more thorough analysis.

Figure M.2: Relationship between media consumption, Russian fluency and support for separatist outcomes in Pridnestrovie



### M.2.2 Gagauzia

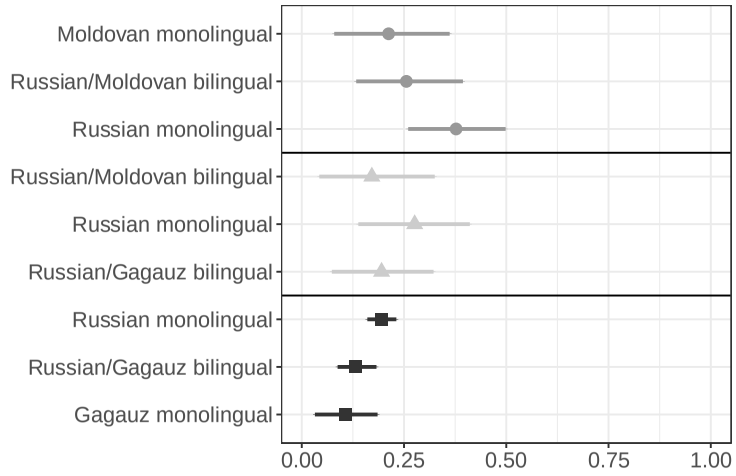
Figure M.3 presents a replication of the analyses of the Gagauz data in the text, using indicators of primary media consumption. In Gagauzia, the results more closely align with those in the text than they do for Pridnestrovie: not consuming media in Russian is the primary correlate of lower support for autonomy within Moldova and integration with Russia. However, consuming Moldovan media also tends to correlate strongly with lower support for independence, in contrast to the relatively weak observational results observed for Moldovan proficiency. Similarly, speaking Gagauz appears to be correlated with lower support for integration with Moldova.

Figure M.4 probes the relationship between Russian fluency, media consumption and support for separatist outcomes. The first row uses integration with Moldova as an outcome, the second autonomy within Moldova, the third independence, and the fourth integration with Russia. The top row shows the interactive relationship of fluency with Russian and Moldovan-language media consumption (left) and Gagauz-language media consumption (right). The right-hand graphic shows an expected set of relationships: fluent speakers of Russian tend to be less supportive of integration with Moldova than non-fluent speakers, regardless of media consumption; Gagauz-language media consumers tend to be less supportive of integration with Moldova than those who do not consume Gagauz media, again regardless of Russian fluency. The Moldovan media consumption graphic is perhaps more puzzling: respondents who watch TV in Moldovan and are fluent in Russian are more supportive of integration with Moldova than those who are not fluent in Russian. Interestingly, a similar result is clear with support for Gagauz independence: there appears to be an interactive relationship between Russian fluency and watching TV and Moldovan vis-a-vis support for Gagauz independence.

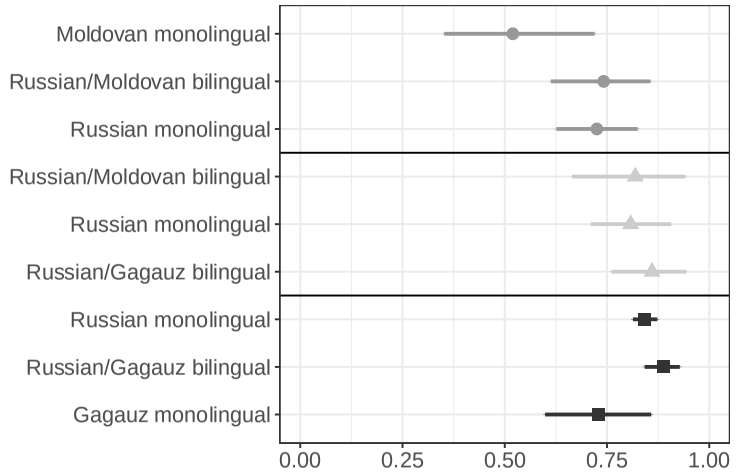
Results with regard to support for autonomy within Moldova and integration with Russia (second and fourth rows) are similarly indicative of a need for future research: in the case fof support for autonomy within Moldova, respondents who both watch TV in Russian and are fluent in the language are the most supportive of that outcome; in the case of support for integration with Russia, it is respondents who watch TV in Russian who most support this outcome, regardless of Russian fluency.

As with the Pridnestrovian analyses, these results indicate that further research with a larger sample will be necessary to further parse out the relationship between language, media, and support for separatism.

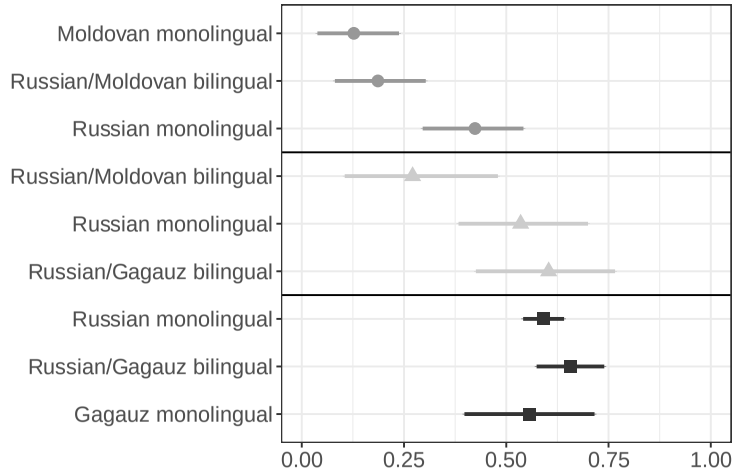
Figure M.3: Posterior probability that Gagauzia respondents with different patterns of media consumption support different separatist outcomes



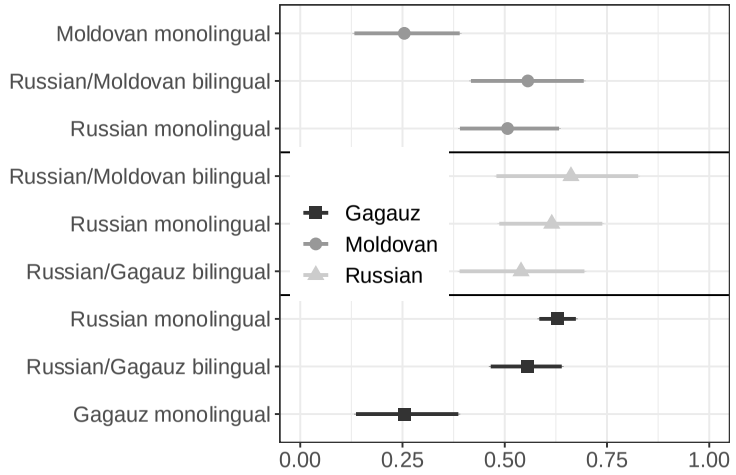

(a) Further integration with Moldova




(b) Autonomy within Moldova



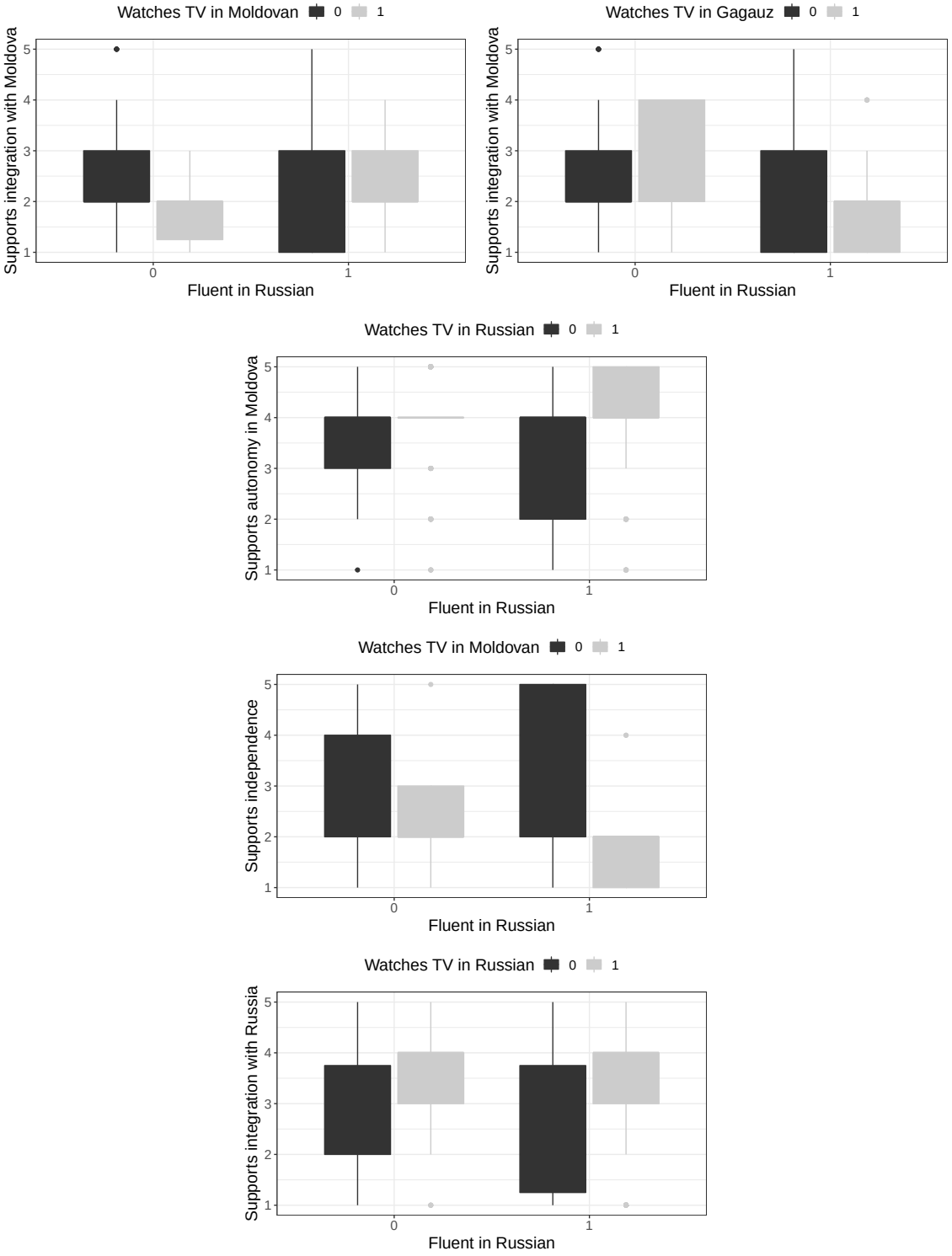

(c) Independence




(d) Integration with Russia

Points represent posterior median and horizontal lines 90 percent credible regions over 500 thousand iterations of four MCMC chains.

Figure M.4: Relationship between media consumption, Russian fluency and support for separatist outcomes in Gagauzia



# N Latent variable analysis

I use a Bayesian latent variable analysis to 1) validate the indicator of linguistic fluency used in the text—a dichotomous indicator representing the top level of a five-point scale on self-reported spoken proficiency—and 2) test the robustness of the regression analyses of support for separatism to a broader operationalization of the concept of linguistic proficiency. Specifically, I use an ordinal Item Response Theory (IRT) model to estimate a composite measure of spoken proficiency in relevant languages (Russian, Moldovan and Ukrainian/Gagauz), then rerun the analyses of support for separatism using these estimates. This analysis yields two main conclusions. First, the latent variable parameters demonstrate that the measure of fluency used in the text is a valid measurement of high-level speaking capabilities. Second, the analyses of support for separatism indicate that, if anything, the operationalization in the text underestimates the relationship between proficiency in Russian and Moldovan and support for separatism. This second result reinforces the importance of language in these cases.

The analyses in the text use a dichotomous indicator of self-reported fluency in the relevant languages to proxy spoken linguistic proficiency. While this approach is theoretically well-grounded and pleasantly simple, it is also a clear simplification of the more complicated concept (spoken proficiency in relevant languages, which is itself a specific aspect of an individual's overall linguistic repertoire). This simplification could entail substantial costs: there is substantial reason to believe that someone who cannot speak a language at all may diverge substantially in her political preferences from someone who speaks the language adequately, but not fluently. Indeed, in some contexts this difference may be more politically relevant than the difference between fluency and high proficiency (Hu & Liu 2020).

Equally importantly, it is possible that the specific indicator of fluent vs. not fluent may not even accurately measure the concept of spoken proficiency. For example, individuals who are not fluent may report fluency for reasons of social desirability or identity (ethnic or regional); definitions of "fluent" may also vary substantially across respondents.

To both assess the validity of the dichotomous fluency indicator I use in the text and test the robustness of the results to a continuous measure of linguistic proficiency, I estimate latent variable models for linguistic capabilities in both Pridnestrovie and Gagauzia. These latent variable models contain both the five-point Likert scale measure used to estimate the indicator in the text, as well as a five-point Likert scale measure of spoken proficiency and seven "can do" measures of spoken proficiency and comprehension. The latter measures to some extent correct for differences in interpretation of the Likert-scale questions—as well as potentially for some social desirability—by asking respondents whether or not they can perform concrete tasks in a given language.[5] I use a Bayesian ordinal Item-Response Theory model to construct the latent values, using ordinal probit functional forms for the Likert-scale responses and standard probit functional forms for the dichotomous responses. Though I do not pool across regions, within regions I assume that the difficulty and discrimination

[5]The tasks are: " Introduce myself ," "talk with friends in about my day-Russian," "carry on in-depth discussions about complicated issues," "understand basic spoken directions as to how to reach a location," "understand discussions about basic topics," " understand people speaking quickly with slang and jargon," and "understand television and radio programs about topics unfamiliar to me."

parameters for different languages are the same.[6] This approach allows me to use a consistent scale within regions across latent values for proficiency in different languages.

## N.1 Latent variable analysis results

Figure N.1 illustrates latent variable model parameters, with Pridnestrovie in the top row and Gagauzia in the bottom. The left column illustrates difficulty parameters from the models (i.e. intercepts from the dichotomous manifest variables and thresholds from the ordinal variables), while the right column illustrates discrimination parameters. I discuss the difficulty and discrimination parameters in turn.

Difficulty parameters (left column) determine the probability that a given latent variable achieves a given value for a manifest variable. For example, for a dichotomous manifest variable the functional form is $Pr(y_i = 1) = \Phi(\alpha + \beta\xi_i)$, where $\alpha$ is the difficulty parameter for a given manifest variable, $\beta$ the discrimination parameter, and $\xi$ the latent variable estimate for observation $i$. Since $\beta$ is always positive by design, higher values of $\alpha$ make it more likely that $y$ will have a value of one. For example, if the latent variable $\xi$ has a value of zero, an $\alpha$ of zero would result in a .5 probability of $y$ having a value of one; if $\alpha$ is 1.96, then there is a .975 probability of $y$ being equal to one. The same logic extends to ordinal manifest variables, except instead of a single difficulty parameter there are four corresponding to the cut-offs for the five Likert scale categories.

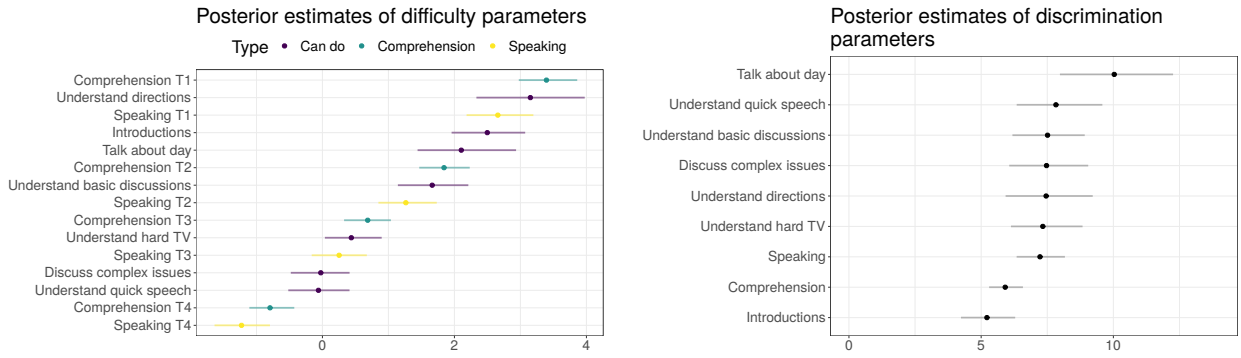
For the purpose of this paper, the key result is that the highest threshold (T4) for the Speaking/Spoken comprehension Likert scale question has the lowest difficulty in both Gagauzia and Pridnestrovie. Since this threshold corresponds to the difference between being considered "fluent" or "not fluent, but speak with ease" for the purposes of the analyses in the text, this result indicates that the measure of fluency I use in the analyses represents a high level of proficiency in the language.

Discrimination parameters (right column) represent the degree to which a given manifest variable is able to discriminate between different levels of latent capabilities (technically, it is the inverse of the variable's stochastic error variance). Higher discrimination values thus represent indicators that are less error prone than others. In Gagauzia, the measure of spoken proficiency is the most discriminatory manifest variable; in Pridnestrovie it is roughly as disciminatory as five of nine manifest variables (slightly less discriminatory than one variable, a can-do question about the ability to speak about one's day; and more discriminatory than the Likert-scale comprehension question and the ability to introduce oneself).

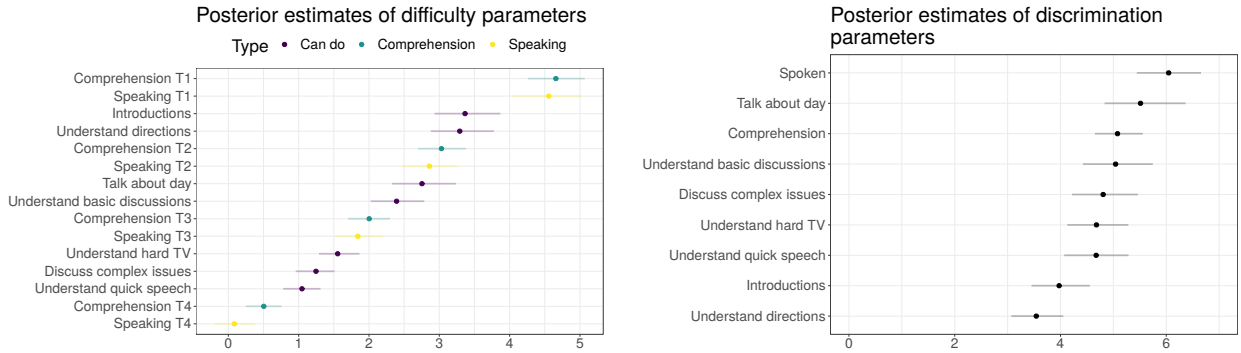
In conjunction with the results regarding the difficulty parameters, the discrimination parameter results indicate that the dichotomized indicator of spoken proficiency used in the analyses in the text is valid for differentiating between high levels of capabilities in a given language. Figure N.2 illustrates what this means in practice, showing latent variable estimates for each respondent in each language-region ordered from highest to lowest. Points represent the median over posterior draws (i.e. the best guess for a given respondent's capabilities in a language), while horizontal lines represent 95% credible regions about this estimate. Different colors in each cell represent different levels of the spoken proficiency, with the highest level (five, the level used to indicate fluency in the text) represented by

[6] I also restrict discrimination parameters to positive values for identification purposes.

Figure N.1: IRT model parameters




(a) Pridnestrovie




(b) Gagauzia

Points represent posterior median point estimates and horizontal lines 95% credible regions over 100 thousand draws from four MCMC chains.

yellow and purple the lowest level (one). The top row illustrates results from Pridnestrovie, the bottom Gagauzia; the left column Russian proficiency, the middle Moldovan, the right Ukrainian/Gagauz. Recall that values are comparable across languages within regions, but not across regions: a posterior median of -1.5 represents no proficiency and .75 complete proficiency in a language in Gagauzia; the equivalent values are -1.15 and .9 in Pridnestrovie.

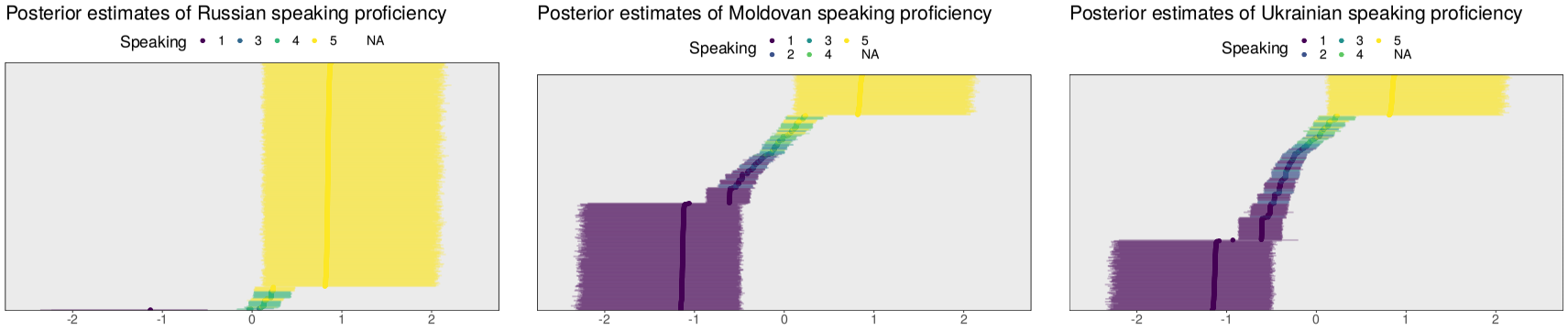
Several aspects of these results are worth noting. First, there is a clear bimodal distribution for all languages in both regions, with many respondents either reporting the highest values for all scales or the lowest. Second, there are very few respondents in both regions who are at the bottom mode for Russian: almost all respondents report at least some proficiency in the language. Again, since Russian is the main language in both regions, this result is to be expected. Finally, the top category in the spoken proficiency Likert scale (yellow) is a very good indicator of whether or not a respondent is in the top category (as the difficulty and discrimination parameters would indicate). The biggest exception to this conclusion is the Gagauz language in Gagauzia, where there is a relatively large number of respondents who report having the second highest value on the scale (four, green) and yet have higher latent values of spoken proficiency than those with who reported the highest. This indicates that some respondents who reported full fluency in the language reported not being able to do some of the tasks in the dichotomous indicators. This specific result for Gagauz may be a relic of the language's lower level of institutionalization and codification: fluent Gagauz speakers may have less exposure to the language on radio or TV; there is also continuing debate about formalizing technical terms.

## N.2 Support for separatism results

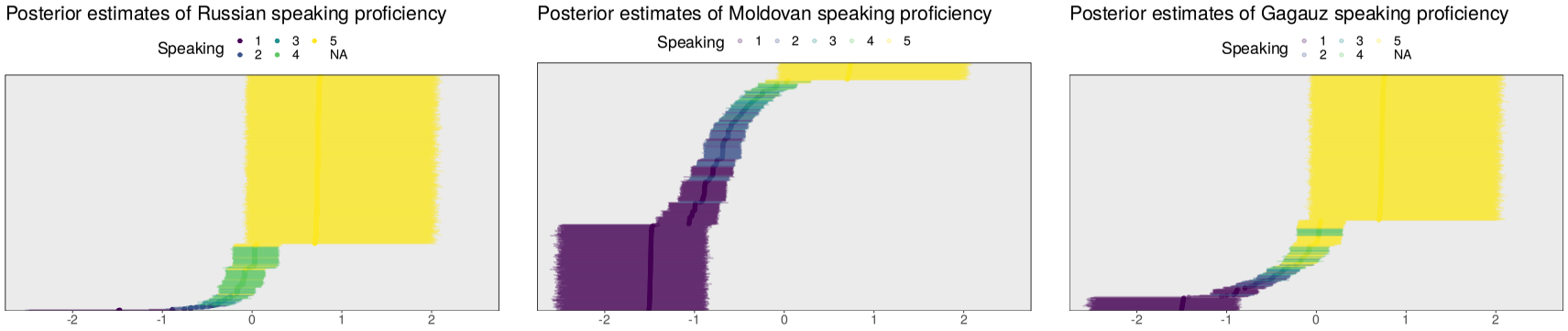
I replicate the analyses discussed in the text, using each respondent's posterior median from the latent variable models as their estimated spoken proficiency in each of the relevant languages. Since the measure of linguistic proficiency is now continuous, illustrating these results requires additional decisions. Specifically, I make predictions for individuals who are either at the top or the bottom of the latent scales. For example, a Moldovan monolingual in Pridnestrovie has a value of -1.15 for Russian proficiency and .9 for Moldovan proficiency; a Moldovan/Russian bilingual in Pridnestrovie has a values of .9 for proficiency in both languages; a Russian monolingual has a value of .9 for Russian proficiency and -1.15 for Moldovan proficiency.

Since the top and bottom of the scales represent modal values for the data, this decision has some empirical justification. However, since there are very few respondents in the bottom category for the Russian language, these results should be treated with caution. Changing the prediction algorithm so that low Russian proficiency represents individuals with relatively low— not no—proficiency in Russian still shows substantively strong—but less extreme—results.

Figure N.2: Relationship between latent spoken proficiency estimates and five-point Likert scale proficiency categories




(a) Pridnestrovie




(b) Gagauzia

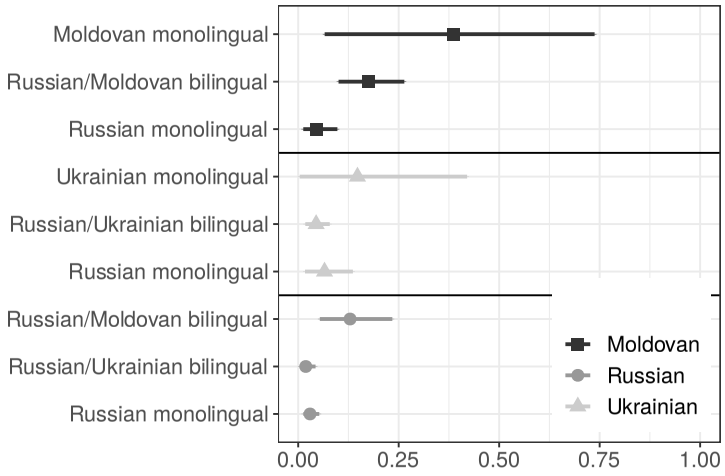
### N.2.1 Pridnestrovie

Figure N.3 illustrates the posterior-predicted probability that respondents with different ethnic and linguistic characteristics would support four different separatist outcomes. The results for support for Pridnestrovian independence (upper right cell) and integration with Ukraine (lower left cell) are roughly the same as in the analyses that use a dichotomous indicator of linguistic fluency: non-speakers of Russian are substantially less likely to support Pridnestrovian independence than speakers of the language; they are more likely to support integration with Ukraine. However, treating spoken proficiency as a continuous latent variable substantially affects results regarding support for integration with Moldova and Russia. With regard to integration with Moldova, proficiency in Moldovan shows a much stronger relationship with this outcome, congruent with linguistic explanations of conflict: since Moldovan is the language of the center, speakers of this language have greater opportunities there than non-speakers. Proficiency in both Moldovan and Ukrainian have strong negative relationships with support for integration with Russia. Given that Russian is the dominant language of Russia, it is possible to speculate that speakers of these languages would fear that the status of their languages would decrease with this outcome.

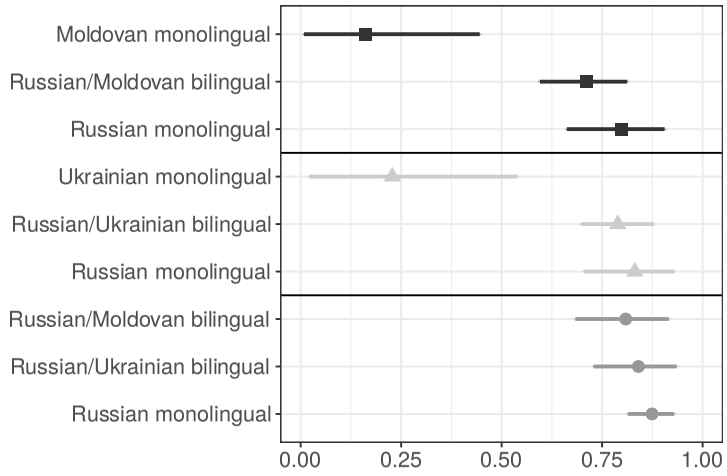
However, in the case of both of these outcomes, a lack of proficiency in Russian remains a strong predictor of lower support for separatism, in line with both linguistic explanations of separatism and the results in the text.

Figure N.4 presents experimental results, which are very much in line with those in the text. However, it is perhaps worth noting that both Ukrainian and Moldovan proficiency have a strong negative relationship with this outcome in the control condition; again in line with theoretical expectations.

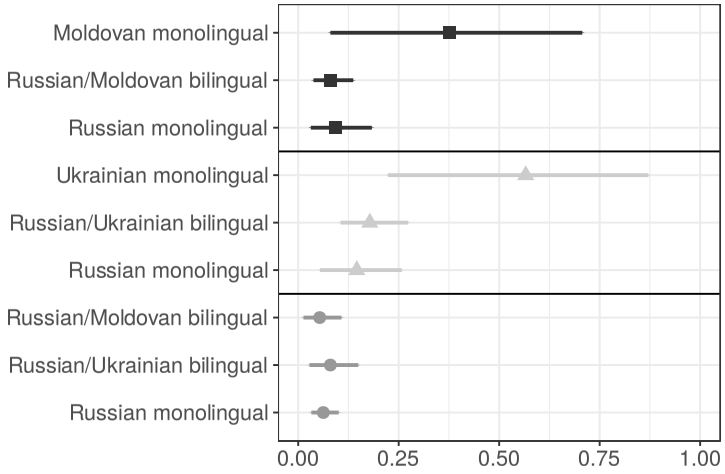
Figure N.3: Posterior probability that Pridnestrovian respondents support different separatist outcomes



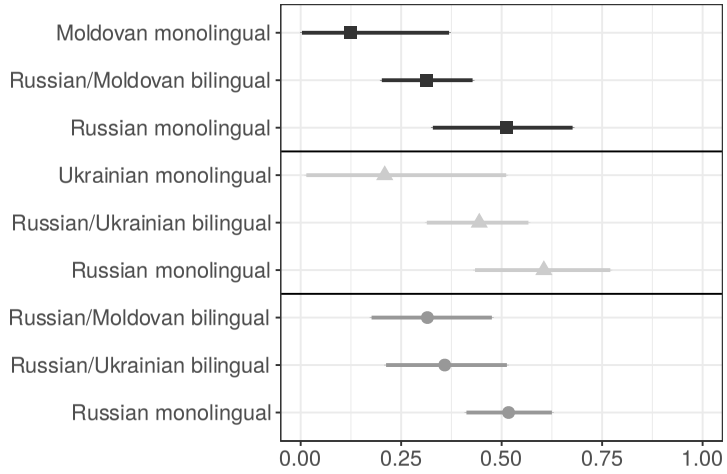

(a) Integration with Moldova




(b) Independence



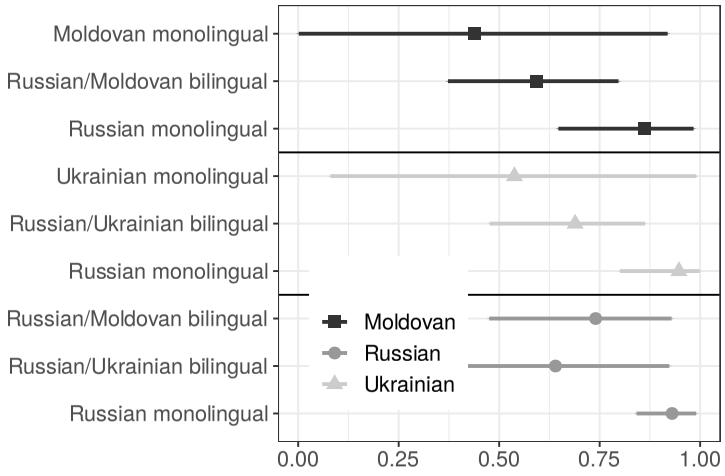

(c) Integration with Ukraine



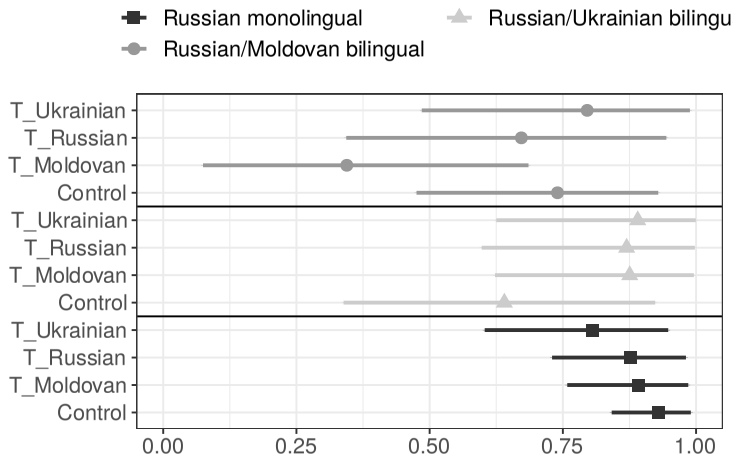

(d) Integration with Russia

Points represent posterior median and horizontal lines 95 percent credible regions over 500 thousand iterations of four MCMC chains.

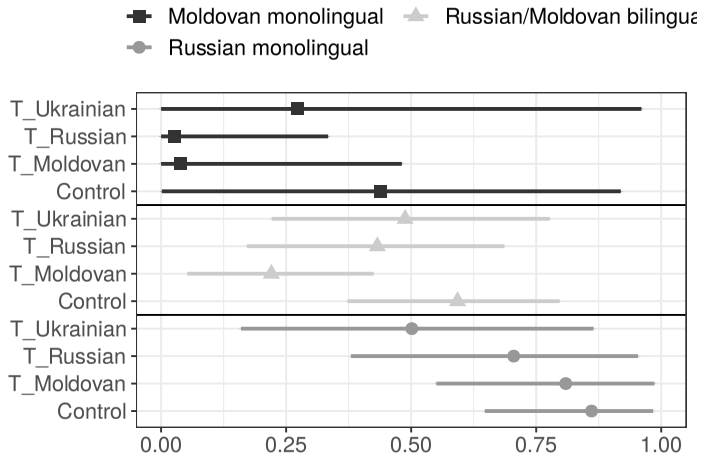
Figure N.4: Posterior probability of supporting Pridnestrovian independence in control condition and across experimental conditions by ethnic group.



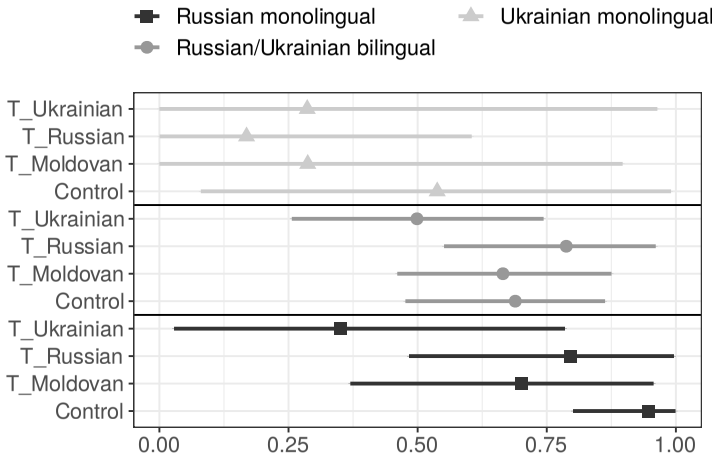

(a) Control




(b) Ethnic Russian




(c) Ethnic Moldovan




(d) Ethnic Ukrainian

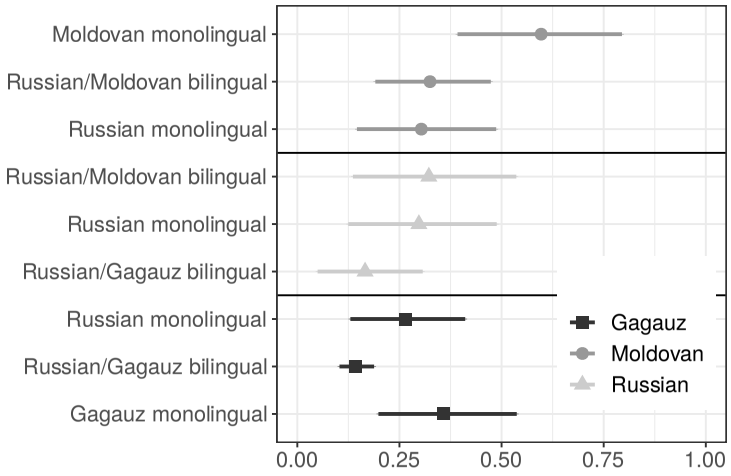
Points represent posterior median and horizontal lines 95 percent credible regions over 500 thousand iterations of four MCMC chains.

### N.2.2 Gagauzia

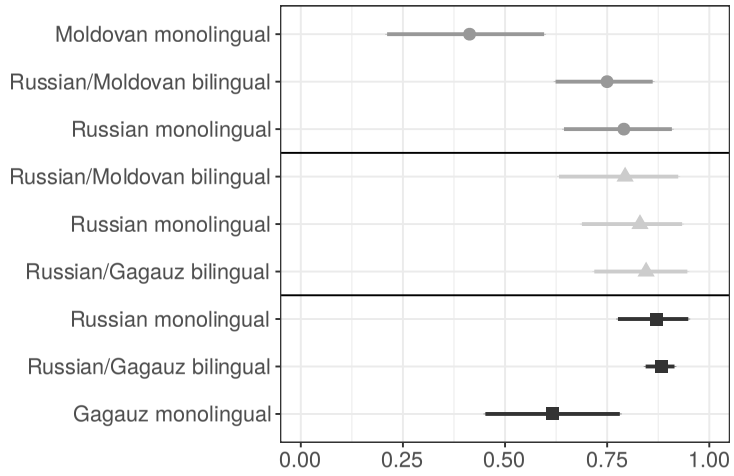
Figure N.5 presents results from analyses of four representative separatist outcomes, using the continuous measure of proficiency in the relevant languages. As with the results from Pridnestrovie, the results regarding the relevance of proficiency in Russian and support for separatism have, if anything, intensified. However, proficiency in Moldovan also has a stronger relationship with support for separatism in this model: speakers of Moldovan are substantially less likely to support Gagauz independence or integration with Russia (bottom row) than non-speakers of this central language. These analyses also emphasize that bilingual Gagauz/Russian speakers tend to be the most supportive of separatism, which is further evidence that linguistic identity vs. social mobility concerns are not necessarily either/or propositions.

Figure N.6 presents experimental results. Here, the biggest change from the results in the text is that $T_{Gagauz}$ increases support for separatism among Gagauz speakers (both mono- and bilingual), though it still reduces support for this outcome among monolingual Russian speakers. This result again indicates that the salience of the Gagauz language requires priming for speakers to consider it when making political decisions.

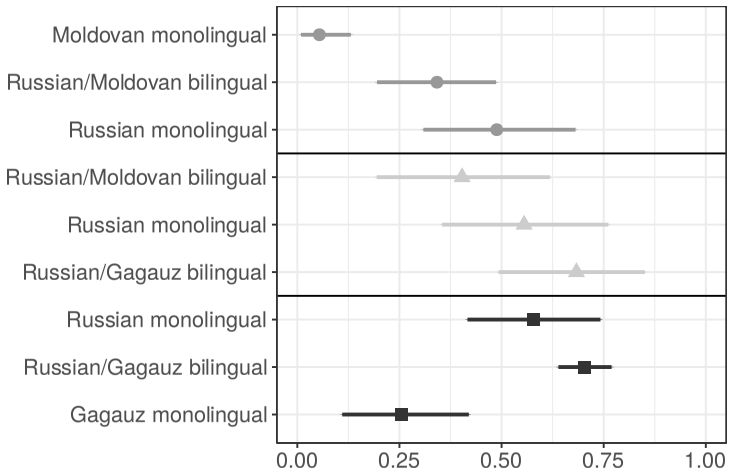
Figure N.5: Posterior probability that Gagauzia respondents support different separatist outcomes



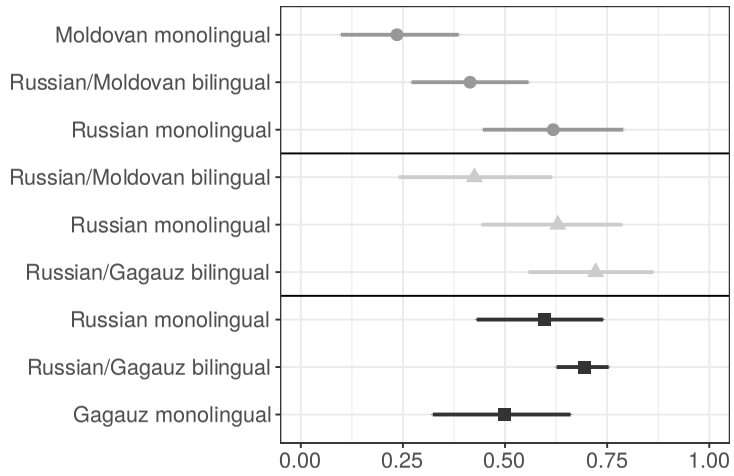

(a) Further integration with Moldova




(b) Autonomy within Moldova



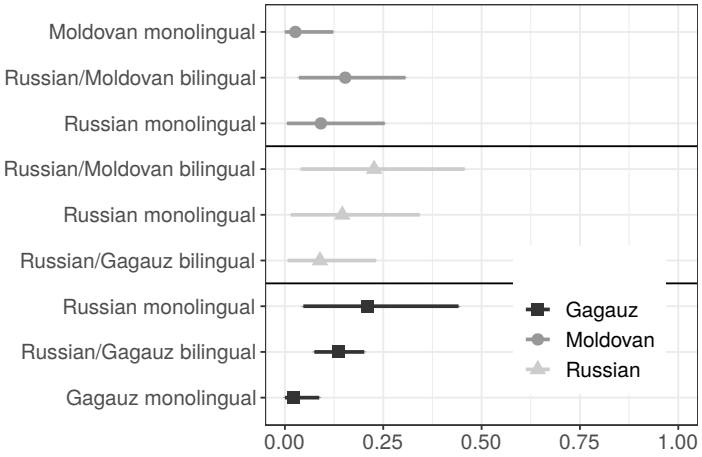

(c) Independence



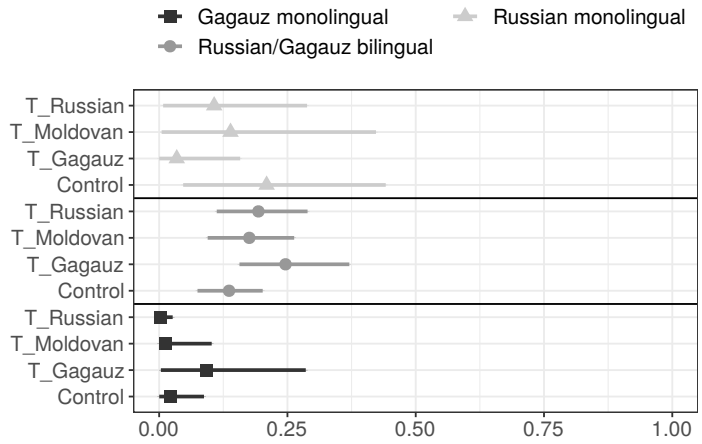

(d) Integration with Russia

Points represent posterior median and horizontal lines 95 percent credible regions over 500 thousand iterations of four MCMC chains.

Figure N.6: Posterior probability of supporting Gagauz independence in control condition and across experimental conditions for ethnic Gagauz.




(a) Control




(b) Ethnic Gagauz

Points represent posterior median and horizontal lines 95 percent credible regions over 500 thousand iterations of four MCMC chains.