Wikipedia Portal A/B Test: Preferred Language Detection and Primary Link Reordering

Mikhail Popov (Analysis & Reporting), Jan Drewniak (UX Engineering), Deborah Tankersley (Product Management), Trey Jones (Review)

April 20, 2016

Executive Summary

The Wikimedia Foundation (WMF) is dedicated to encouraging the growth, development, and distribution of free, *multilingual*, educational content. WMF operates Wikipedia, a project to build free encyclopedias in *all languages of the world*. For the Discovery Department's Portal team (tasked with making the wikipedia.org a more engaging and friendlier portal to free knowledge), it is important that all visitors, regardless of language, receive a more welcoming experience.

In this A/B test, the test group were presented with primary links that were dynamically filled in according to the users' preferred languages. Users who received these dynamic primary links were more likely to engage with those primary links than the users who received the default, static experience, albeit not by a lot. The biggest impact is actually found in ***where*** those users went to from the Portal.

Users were 7.5-16.1 times more likely to visit a Wikipedia in their most preferred language (or one of their preferred languages) when they were presented with primary links that reflected their preferred languages — 7.5 times in the case of multilingual users, 16.1 times in the case of users whose Accept-Language did not include English. We believe this is evidence of localization having a positive effect on the users' experience and engagement with the Portal.

How the Wikipedia Portal would look like to someone whose preferred languages are (in order): French, Arabic, Russian, German, Swahili, and English.

Introduction

The Wikimedia Foundation (WMF) is dedicated to encouraging the growth, development, and distribution of free, *multilingual*, educational content. WMF operates Wikipedia, a project to build free encyclopedias in *all languages of the world*. For the Discovery Department's Portal team (tasked with making wikipedia.org a more engaging and friendlier portal to free knowledge), it is important that all visitors, regardless of language, receive a more welcoming experience.

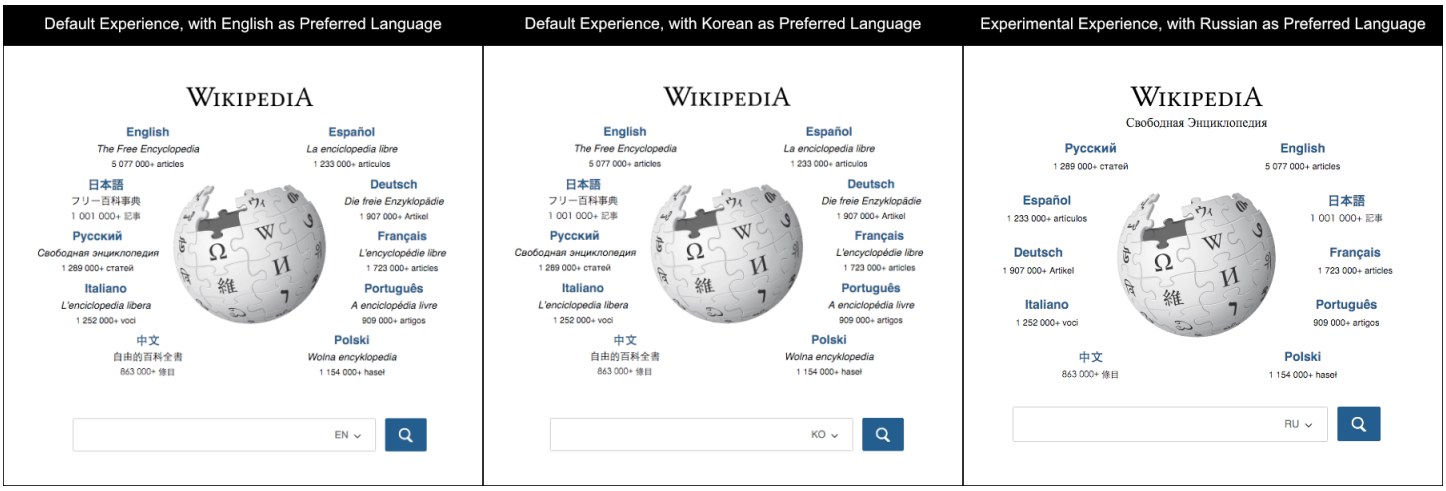

Figure 1: The three types of Wikipedia Portal interfaces we currently show our visitors. The language to search in is automatically set by the page according to user's preferred language. The third interface was the A/B test.

Methods

The code for the exploratory data analysis and the confirmatory data analysis can be found on the GitHub repository for this analysis.

Data Collection

In general, we randomly select 1 in 200 users to be included in the Wikipedia Portal EventLogging schema. Of those, we randomly selected 1 in 10 to be included in the A/B test. If selected for the test, the user is assigned either variation A (the default experience) or variation B (the experimental experience, see demo) with 50%/50% probability. The test was deployed 22 March 2016 (see T124112) and ran for 22 days, a period during which we logged 35K users' clickthroughs and abandonments.

Statistical Analysis

Under the conjugate Beta-Binomial model, each i-th group's number of successful outcomes y_i (out of a total n_i subjects in the group) follows a Binomial distribution with a Beta prior on the probability of success parameter π_i:

$$\begin{aligned} y_i &\sim \text{Binomial}(n_i, \pi_i), \; i = 1, 2; \\ \pi_i &\sim \text{Beta}(\alpha_i, \beta_i), \end{aligned}$$

which yields a Beta posterior $p(\pi_i|y_i)$. We utilized the R package *BCDA* (0.1.0) to perform a Bayesian analysis of the collected logs, assuming a Beta-Binomial model for counts of successful outcomes among the groups.

Results

Clickthrough Rates

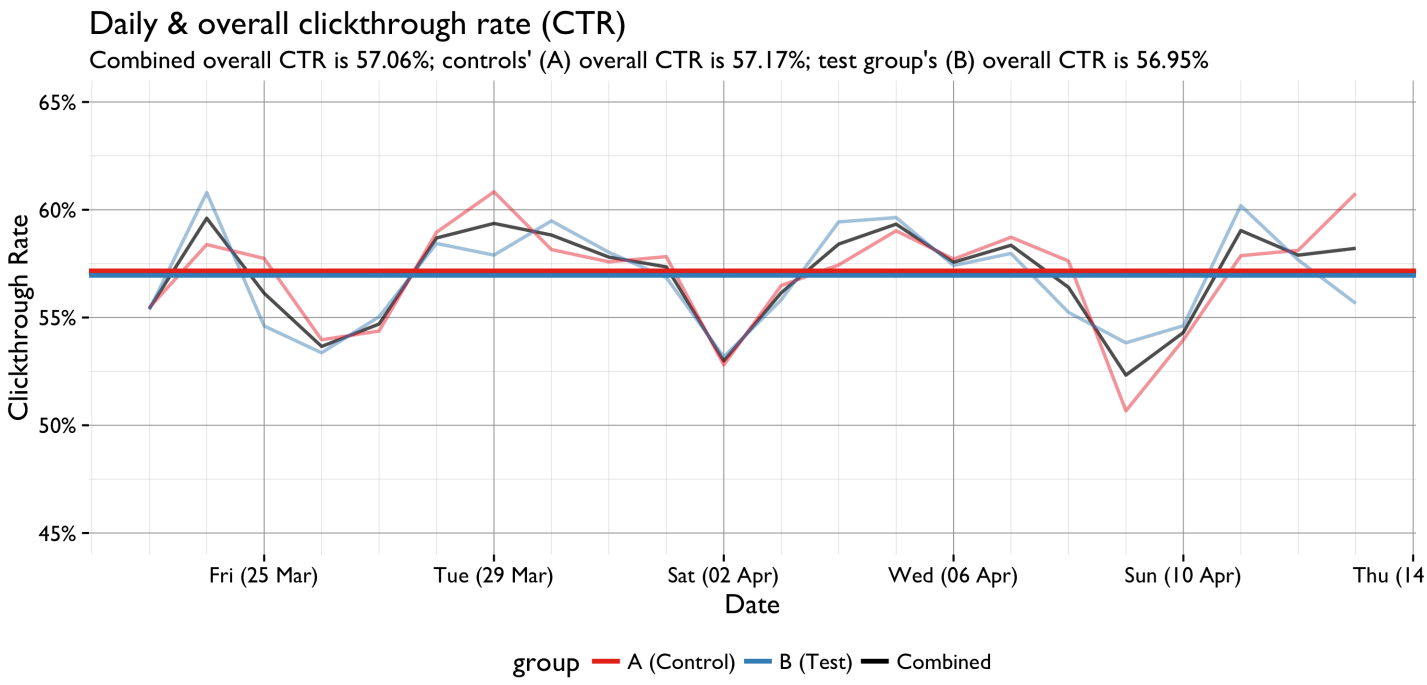

Figure 2: Daily and overall clickthrough rate for the whole dataset and broken down by group (controls vs test). If we look at just the clickthrough by group without breaking it up by section and without taking into consideration the user's preferred languages (from their Accept-Language header), the two groups behave nearly identically.

Our **previous investigation** of language preferences and engagement with Portal revealed that users whose preferences include English search more and engage with the primary links (the language links around the Wikipedia logo) more than users whose preferences do not include English. We see the same pattern in Fig. 4, where users whose preferences include English are much more likely to engage with the primary language links and the search box, while users whose preferences do not include English are more likely to engage with the secondary language links.

We can also see in Figure 4 and Table 1 that in both sets of users, the test group is slighly more likely to engage with the primary links (the part affected by the A/B test) than the control group. Specifically, when considering only the users whose preferences included English (where we observed the largest difference), the test group has a 1.32% higher primary link clickthrough and is 1.064 times more likely to clickthrough than the control group.

Unfortunately, these differences aren't "statistically significant" – the 95% Bayesian confidence intervals include 0 for the difference of proportions and 1 for the relative risk, meaning we don't have evidence of big impact. The same numbers for the users who did not include English are: 0.84% difference and 1.053 times more more likely to click on a primary link. However, as we will see in the following section, the real difference is hidden in the relationship between the language of the Wikipedia the user goes to and their preferred language(s).

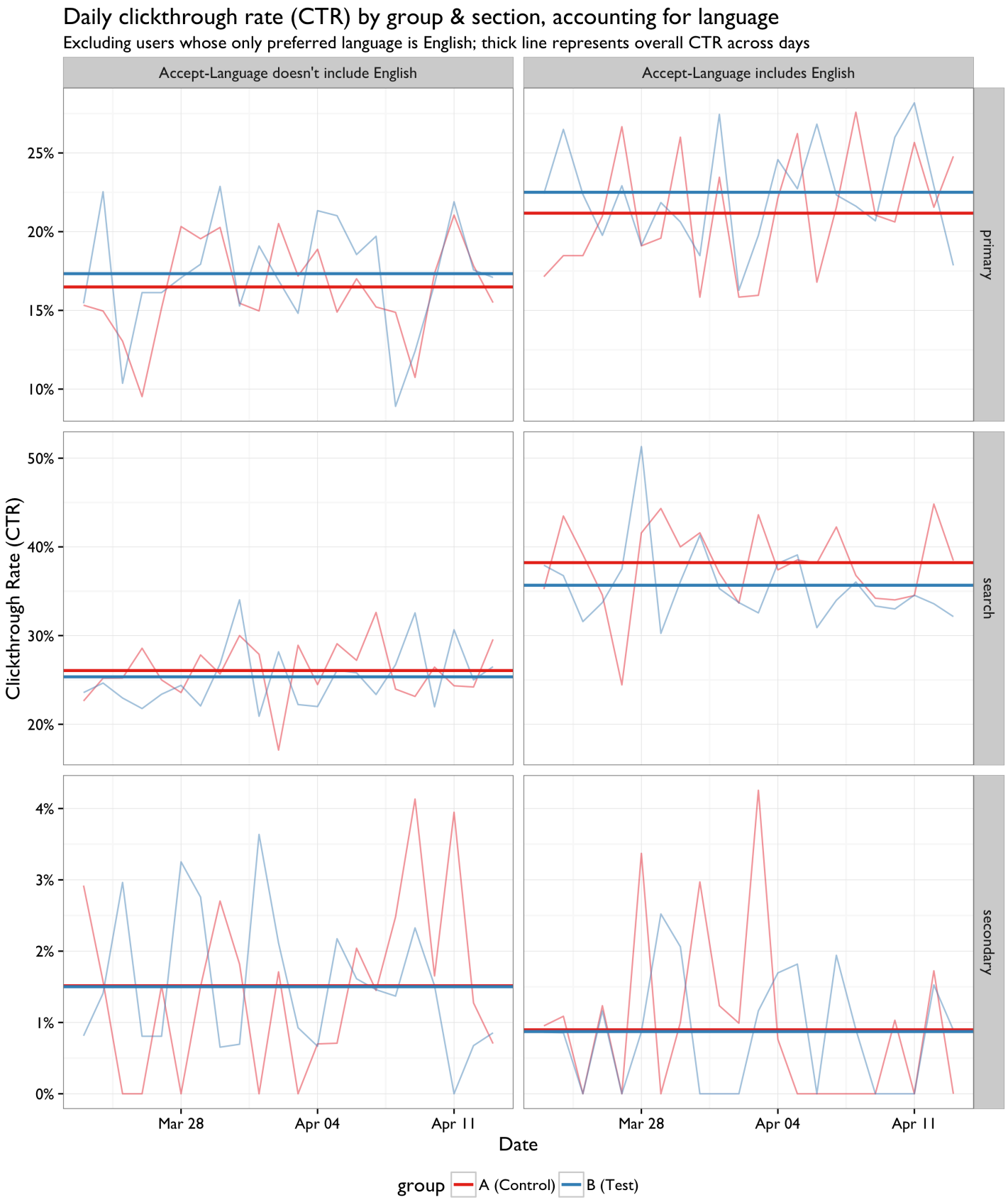

Figure 3: Daily and overall clickthrough rate broken down by group (controls vs test) and section used (primary links around the globe, search box, secondary links) after excluding users who would not be affected by the test (users whose only preferred language is English).

Table 1: Clickthrough rate by accept-language, link, and group. This is the table version of Figure 4. Increases in primary link clickthrough rate cannibalize the clickthrough rates of other sections.

Includes English	**Link**	**Group**	**Clickthrough Rate**	
Accept-Language doesn't include English	primary	A (Control)	16.49%	
Accept-Language doesn't include English	primary	B (Test)	17.33%	↑
Accept-Language doesn't include English	search	A (Control)	26.06%	
Accept-Language doesn't include English	search	B (Test)	25.35%	↓
Accept-Language doesn't include English	secondary	A (Control)	1.51%	
Accept-Language doesn't include English	secondary	B (Test)	1.50%	↓
Accept-Language includes English	primary	A (Control)	21.18%	
Accept-Language includes English	primary	B (Test)	22.50%	↑
Accept-Language includes English	search	A (Control)	38.22%	
Accept-Language includes English	search	B (Test)	35.67%	↓
Accept-Language includes English	secondary	A (Control)	0.90%	
Accept-Language includes English	secondary	B (Test)	0.87%	↓

Destination

Figures 4–7 and Tables 2–5 below summarize the statistical comparisons of the test group against the controls. "% Diff" refers to — for example — "π_B who visited Wikipedia in a preferred language ***minus*** π_A who visited Wikipedia in a preferred language", with positive numbers indicating that the proability of a successful outcome is greater in B than A. Relative risk is interpreted as "X times more likely to visit Wikipedia in a preferred language" ($RR = \pi_B/\pi_A$) and odds ratio is interpreted as "the odds of group B users visiting Wikipedia in a preferred language are X times the odds of group A users" ($OR = (\pi_B/(1-\pi_B))/(\pi_A/(1-\pi_A))$).

Most notably (in Figure 4 and Table 2) 16.1% more users whose Accept-Language header did not include English visited Wikipedia in one of their preferred languages in group B than in group A, and were 1.3 times more likely to go to a Wikipedia in one of their preferred languages. Furthermore, 15.5% more B users visited a Wikipedia in their *most* preferred language than A users, and were 1.3 times more likely to visit such a Wikipedia.

The other two comparisons imply that (1) the test does not appear to have had an effect on users whose Accept-Language header included English, and (2) the results for secondary links mirror the results for the primary links, but negative, which implies that people were indeed using the dynamic primary links to visit a Wikipedia in their language rather than the secondary links.

In Figure 5 and Table 3, where we excluded users whose most preferred language was English, we see the some of the same results as in Table 2 — since the users whose Accept-Language doesn't include English would be in both subpopulations — but we also notice that users whose Accept-Language *does* include English (just not as the first language) were 1.25 times more likely to visit a Wikipedia in their most preferred language, with 11.3% more users in the test group visiting a Wikipedia in their most preferred language than the control group.

We see a similar pattern in Figure 6 and Table 4, where we excluded users whose only preferred language is English. In Figure 7 and Table 5, where we only looked at multilingual users, users were 1.16 times more likely to visit a Wikipedia in their most preferred language.

Conclusion/Discussion

In summary, users who saw the primary links that were dynamically filled in according to the users' preferred languages were more likely to engage with those primary links than the users who received the default, static experience, albeit not by a lot. The biggest impact is actually found in ***where*** those users went to from the Portal.

Users were 7.5-16.1 times more likely to visit a Wikipedia in their most preferred language (or one of their preferred languages) when they were presented with primary links that reflected their preferred languages — 7.5 times in the case of multilingual users, 16.1 times in the case of users whose Accept-Language did not include English. We believe this is evidence of localization having a positive effect on the users' experience and engagement with the Portal.

Figure 4: The types of Wikipedias that all users visited from the Portal.

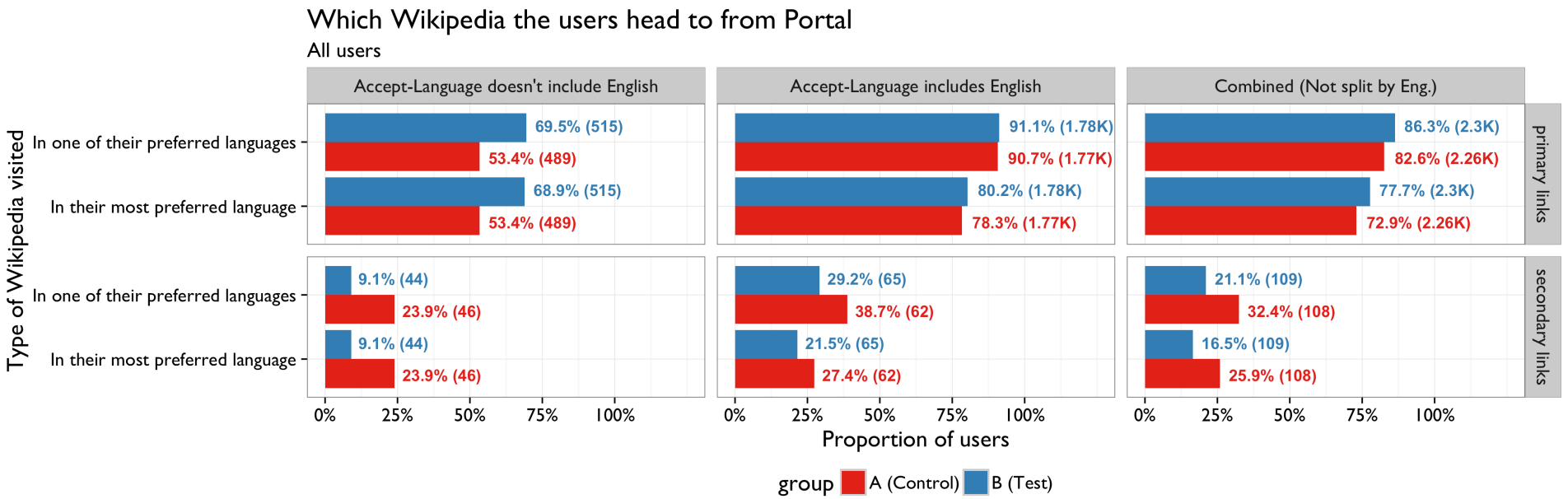

Table 2: Statistical comparisons of test group vs controls with respect to language of the Wikipedia they visited through a primary or secondary link, not excluding any users. Values in parentheses represent 95% Bayesian Confidence Intervals – meaning there is a 95% probability the value of interest is inside the interval. ("Prim" = primary link, "Sec" = secondary link)

Link	Includes English	Wikipedia language	n_B	n_A	π_B (%)	π_A (%)	% Diff (B vs A)	Relative Risk	Odds Ratio
Prim	A-L doesn't include En	one of preferred	515	489	69.4 (65.4, 73.3)	53.3 (49.0, 57.7)	16.1 (9.9, 21.8)	1.30 (1.18, 1.44)	2.01 (1.51, 2.54)
Prim	A-L doesn't include En	most preferred	515	489	68.9 (64.8, 72.8)	53.3 (49.0, 57.7)	15.5 (9.4, 21.2)	1.29 (1.17, 1.43)	1.95 (1.47, 2.46)
Prim	A-L includes En	one of preferred	1.8K	1.8K	91.1 (89.7, 92.4)	90.7 (89.3, 92.0)	0.4 (-1.5, 2.3)	1.00 (0.98, 1.03)	1.06 (0.83, 1.31)
Prim	A-L includes En	most preferred	1.8K	1.8K	80.2 (78.3, 82.0)	78.3 (76.4, 80.2)	1.9 (-0.7, 4.6)	1.02 (0.99, 1.06)	1.13 (0.95, 1.31)
Sec	A-L doesn't include En	one of preferred	44	46	10.9 (2.8, 19.8)	25.0 (13.4, 36.9)	-14.1 (-28.8, 1.5)	0.46 (0.09, 0.93)	0.41 (0.06, 0.92)
Sec	A-L doesn't include En	most preferred	44	46	10.9 (2.8, 19.8)	25.0 (13.4, 36.9)	-14.1 (-28.8, 1.5)	0.46 (0.09, 0.93)	0.41 (0.06, 0.92)
Sec	A-L includes En	one of preferred	65	62	29.9 (19.3, 41.0)	39.1 (27.7, 51.1)	-9.2 (-25.1, 6.9)	0.78 (0.42, 1.17)	0.71 (0.26, 1.25)
Sec	A-L includes En	most preferred	65	62	22.4 (12.5, 32.2)	28.1 (17.2, 38.7)	-5.7 (-19.7, 10.0)	0.83 (0.38, 1.35)	0.80 (0.27, 1.48)

*Figure 5: The types of Wikipedias that non-**primarily**-English-speaking users visited from the Portal.*

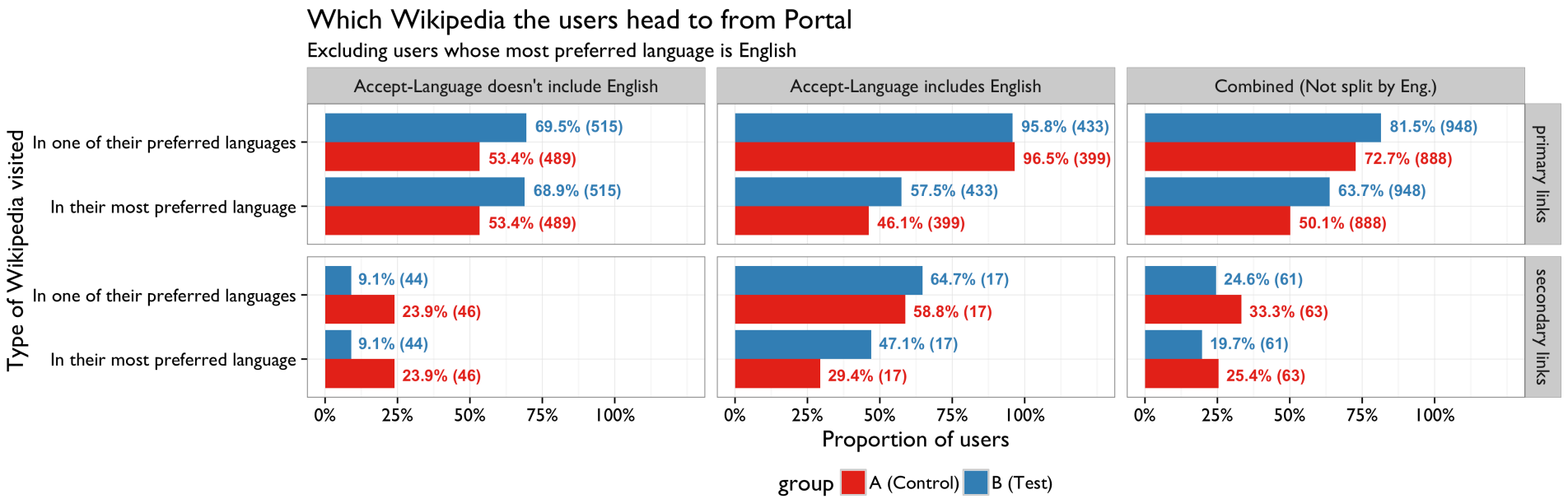

Table 3: Statistical comparisons of test group vs controls with respect to language of the Wikipedia they visited through a primary or secondary link, excluding users whose most preferred language is English. Values in parentheses represent 95% Bayesian Confidence Intervals – meaning there is a 95% probability the value of interest is inside the interval. ("Prim" = primary link, "Sec" = secondary link)

Link	Includes English	Wikipedia language	n_B	n_A	π_B (%)	π_A (%)	% Diff (B vs A)	Relative Risk	Odds Ratio
Prim	A-L doesn't include En	one of preferred	515	489	69.4 (65.4, 73.3)	53.3 (49.0, 57.7)	16.1 (9.9, 21.8)	1.30 (1.18, 1.44)	2.01 (1.51, 2.54)
Prim	A-L doesn't include En	most preferred	515	489	68.9 (64.8, 72.8)	53.3 (49.0, 57.7)	15.5 (9.4, 21.2)	1.29 (1.17, 1.43)	1.95 (1.47, 2.46)
Prim	A-L includes En	one of preferred	433	399	95.6 (93.6, 97.5)	96.3 (94.3, 97.9)	-0.6 (-3.3, 2.1)	0.99 (0.97, 1.02)	0.90 (0.34, 1.55)
Prim	A-L includes En	most preferred	433	399	57.5 (52.8, 62.1)	46.2 (41.3, 50.9)	11.3 (4.6, 18.2)	1.25 (1.09, 1.42)	1.59 (1.17, 2.04)
Sec	A-L doesn't include En	one of preferred	44	46	10.9 (2.8, 19.8)	25.0 (13.4, 36.9)	-14.1 (-28.8, 1.5)	0.46 (0.09, 0.93)	0.41 (0.06, 0.92)
Sec	A-L doesn't include En	most preferred	44	46	10.9 (2.8, 19.8)	25.0 (13.4, 36.9)	-14.1 (-28.8, 1.5)	0.46 (0.09, 0.93)	0.41 (0.06, 0.92)
Sec	A-L includes En	one of preferred	17	17	63.1 (42.4, 84.1)	57.8 (36.0, 78.5)	5.3 (-24.5, 35.1)	1.14 (0.59, 1.77)	1.60 (0.16, 3.99)
Sec	A-L includes En	most preferred	17	17	47.4 (26.3, 69.8)	31.6 (12.5, 52.0)	15.8 (-14.5, 44.3)	1.70 (0.56, 3.32)	2.59 (0.21, 6.43)

*Figure 6: The types of Wikipedias that not-**just**-English-speaking users visited from the Portal.*

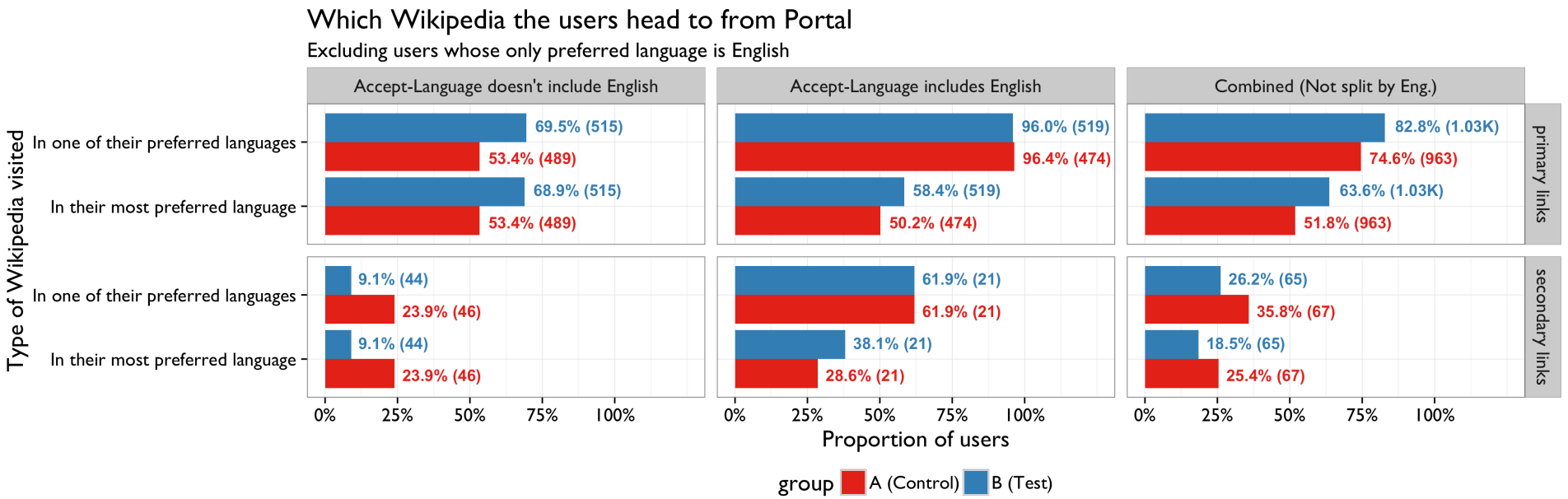

Table 4: Statistical comparisons of test group vs controls with respect to language of the Wikipedia they visited through a primary or secondary link, excluding users whose only preferred language is English. Values in parentheses represent 95% Bayesian Confidence Intervals – meaning there is a 95% probability the value of interest is inside the interval. ("Prim" = primary link, "Sec" = secondary link)

Link	Includes English	Wikipedia language	n_B	n_A	π_B (%)	π_A (%)	% Diff (B vs A)	Relative Risk	Odds Ratio
Prim	A-L doesn't include En	one of preferred	515	489	69.4 (65.4, 73.3)	53.3 (49.0, 57.7)	16.1 (9.9, 21.8)	1.30 (1.18, 1.44)	2.01 (1.51, 2.54)
Prim	A-L doesn't include En	most preferred	515	489	68.9 (64.8, 72.8)	53.3 (49.0, 57.7)	15.5 (9.4, 21.2)	1.29 (1.17, 1.43)	1.95 (1.47, 2.46)
Prim	A-L includes En	one of preferred	519	474	95.8 (94.0, 97.4)	96.2 (94.5, 97.8)	-0.4 (-2.8, 2.0)	1.00 (0.97, 1.02)	0.94 (0.40, 1.57)
Prim	A-L includes En	most preferred	519	474	58.3 (54.1, 62.6)	50.2 (45.6, 54.5)	8.2 (2.1, 14.4)	1.16 (1.03, 1.30)	1.40 (1.06, 1.77)
Sec	A-L doesn't include En	one of preferred	44	46	10.9 (2.8, 19.8)	25.0 (13.4, 36.9)	-14.1 (-28.8, 1.5)	0.46 (0.09, 0.93)	0.41 (0.06, 0.92)
Sec	A-L doesn't include En	most preferred	44	46	10.9 (2.8, 19.8)	25.0 (13.4, 36.9)	-14.1 (-28.8, 1.5)	0.46 (0.09, 0.93)	0.41 (0.06, 0.92)
Sec	A-L includes En	one of preferred	21	21	60.8 (41.5, 80.4)	60.8 (40.7, 79.1)	0.1 (-26.9, 26.7)	1.03 (0.58, 1.54)	1.21 (0.19, 2.78)
Sec	A-L includes En	most preferred	21	21	39.2 (19.6, 58.5)	30.5 (13.2, 48.9)	8.7 (-18.9, 35.0)	1.43 (0.46, 2.75)	1.85 (0.27, 4.42)

Figure 7: The types of Wikipedias that multilingual users visited from the Portal.

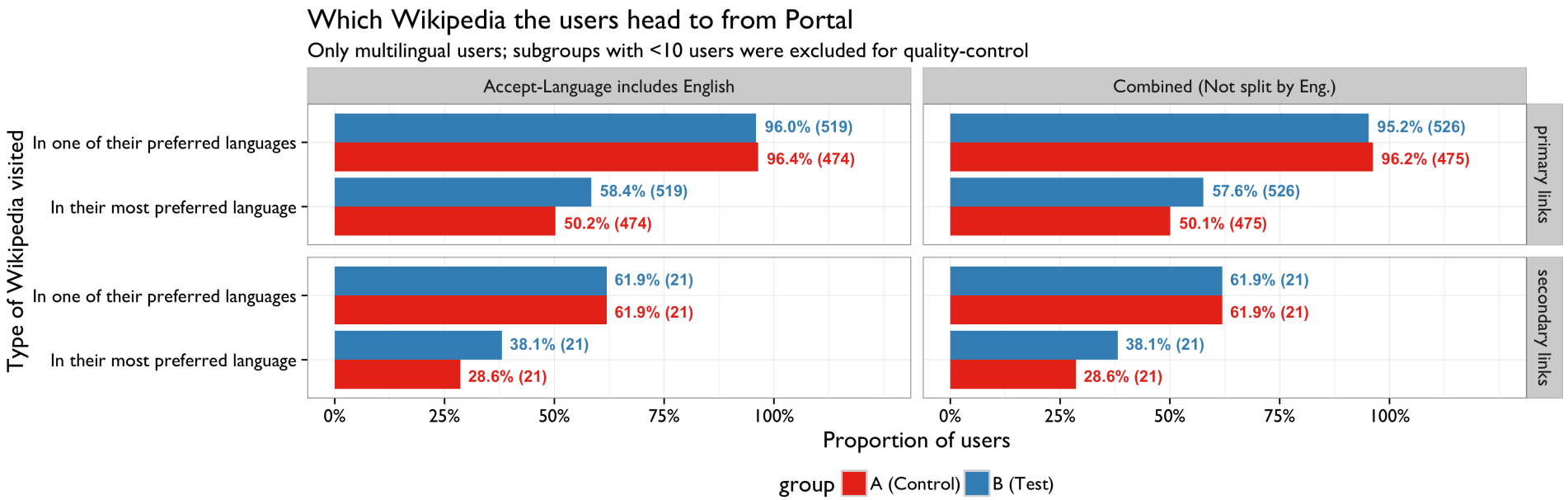

Table 5: Statistical comparisons of test group vs controls with respect to language of the Wikipedia they visited through a primary or secondary link, looking at multilingual users only. Values in parentheses represent 95% Bayesian Confidence Intervals – meaning there is a 95% probability the value of interest is inside the interval. ("Prim" = primary link, "Sec" = secondary link)

Link	Includes English	Wikipedia language	n_B	n_A	π_B (%)	π_A (%)	% Diff (B vs A)	Relative Risk	Odds Ratio
Prim	A-L includes En	one of preferred	519	474	95.8 (94.0, 97.4)	96.2 (94.5, 97.8)	-0.4 (-2.8, 2.0)	1.00 (0.97, 1.02)	0.94 (0.40, 1.57)
Prim	A-L includes En	most preferred	519	474	58.3 (54.1, 62.6)	50.2 (45.6, 54.5)	8.2 (2.1, 14.4)	1.16 (1.03, 1.30)	1.40 (1.06, 1.77)
Prim	Combined (Not split by Eng.	one of preferred	526	475	95.1 (93.2, 96.9)	96.0 (94.2, 97.7)	-0.9 (-3.5, 1.6)	0.99 (0.96, 1.02)	0.84 (0.40, 1.37)
Prim	Combined (Not split by Eng.	most preferred	526	475	57.6 (53.3, 61.8)	50.1 (45.5, 54.4)	7.5 (1.5, 13.9)	1.15 (1.02, 1.29)	1.36 (1.02, 1.70)
Sec	A-L includes En	one of preferred	21	21	60.8 (41.5, 80.4)	60.8 (40.7, 79.1)	0.1 (-26.9, 26.7)	1.03 (0.58, 1.54)	1.21 (0.19, 2.78)
Sec	A-L includes En	most preferred	21	21	39.2 (19.6, 58.5)	30.5 (13.2, 48.9)	8.7 (-18.9, 35.0)	1.43 (0.46, 2.75)	1.85 (0.27, 4.42)
Sec	Combined (Not split by Eng.	one of preferred	21	21	60.8 (41.5, 80.4)	60.8 (40.7, 79.1)	0.1 (-26.9, 26.7)	1.03 (0.58, 1.54)	1.21 (0.19, 2.78)
Sec	Combined (Not split by Eng.	most preferred	21	21	39.2 (19.6, 58.5)	30.5 (13.2, 48.9)	8.7 (-18.9, 35.0)	1.43 (0.46, 2.75)	1.85 (0.27, 4.42)