

Extended Process Similarity Review Panel

(EPSRP)

Report for

Country: European Union

A-Label: xn--qxa6a

U-Label: ευ

Unicode Code Points: U+03B5 U+03C5

String in English: eu

String Language: Greek, Modern (1453)

Language Scripts: Greek

September 2014

## Contents

Executive Summary.....	3
1 Background .....	4
2 Methodology.....	4
3 Panel Members and Research Team .....	5
4 Information on string to evaluate.....	6
5 Documents provided to the panel by ICANN.....	6
6 Research Report Summary .....	6
6.1 Stimuli for Candidate: eu/ EY in Greek.....	7
6.2 Results.....	7
6.2.1 DMTS.....	7
6.2.2 Same/different go/no-go task.....	8
7 Analysis by panel members .....	8
8 Recommendations of the EPSRP.....	9
Annex A - Results of the Research Team Experimentation .....	11
Annex B - Final Report of the EPSRP for the application for EU in Greek.....	38

## Executive Summary

The Extended Process Similarity Review Panel (EPSRP) presents its recommendations on the following IDN ccTLD application:

Corresponding ISO3166 Entry: EU  
A-Label: xn--qxa6a  
U-Label: ευ  
Unicode Code Points: U+03B5 U+03C5  
String in English: eu  
String Language: Greek, Modern (1453-)  
Language Scripts: Greek

The Extended Process Similarity Review Panel (EPSRP) was created under the Final Implementation Plan for IDN ccTLD Fast Track Process to provide ICANN with recommendations regarding IDN ccTLD applications being confusingly similar to ISO 3166-1 entries.

The EPSRP is composed of panel members which are internationally recognized researchers in the relevant field as well as a research team which was responsible for carrying out the experimentation.

The research team in collaboration with panel members developed an empirical evaluation methodology based on the latest scientific findings in the relevant field to determine if an applied for IDN ccTLD string should be considered confusingly similar to any ISO 3166-1 entries.

The methodology was used by the research team to establish threshold values for its tasks using ISO 3166-1 entries. All of the ISO 3166-1 are in use or potentially available as ccTLDs regardless of their potential for being confusingly similar within this group. The threshold values essentially allow for IDN ccTLD applications to be as similar as any ISO 3166-1 pair.

The methodology was then used on the applied for IDN ccTLD strings and the results compared to the threshold values to determine if they were confusingly similar or not. If the applied for IDN ccTLD in upper or lower case exceeds a threshold value for a given ISO 3166-1 comparison for both tasks then it will be considered confusingly similar.

The panel provides separate recommendations for upper and lower case versions of the applied for IDN ccTLD strings given that from a visual similarity point of view upper and lower case characters of the same letter are distinct entities.

As such the Extended Process Similarity Review Panel presents the following recommendations for this application:

- The panel recommends that the IDN ccTLD application in upper case should be considered confusingly similar to the following ISO 3166-1 entries: EV and EY.
- The panel recommends that the IDN ccTLD application in lower case should not be considered confusingly similar to any ISO 3166-1 entries.

## 1 Background

The **Final Implementation Plan for IDN ccTLD Fast Track Process**

(<http://www.icann.org/en/resources/idn/fast-track/idn-ccTLD-implementation-plan-05nov13-en.pdf>) instituted the **Extended Process Similarity Review Panel (EPSRP)**.

The guidelines for the EPSRP were published on 4 December 2013 and can be found at <http://www.icann.org/en/resources/idn/fast-track/epsrp-guidelines-04dec13-en.pdf>.

The objective of the EPSRP is described as follows in the guidelines:

*In the event a requested string is found to be confusingly similar by the DNS Stability Panel, an external and independent Extended Process Similarity Review Panel (“EPSRP”) conducts a review of the requested IDN ccTLD string, using a different framework from the DNS Stability Panel, and, only upon request of the applicant.*

## 2 Methodology

The methodology was developed by the research team and approved by the Panel after rigorous review.

Two tasks were selected to evaluate visual similarity:

- **Delayed match-to sample (two-alternative forced-choice) task (DMTS).** In this task, participants briefly see one candidate pairs on the screen, after which it is masked. Then, that pair plus a foil appears after a short delay, and they must identify which option was presented.
- **Go/No-go same-different task (GNG).** In this task, participants see two pairs on the screen, left and right of center, outside their central vision. They must respond only when the two differ.

For each task two evaluations of similarity were calculated from the observations, one for response time (RT) and another for response accuracy (error rate). These evaluations combined with the tasks produce four measurements:

- DMTS inv(RT)
- DMTS error rate
- GNG inv(RT)
- GNG error rate

The basic testing procedure involved presenting test subjects with a number of visual stimuli which consist of 2 characters in various versions to obtain data on both tasks. Versions include variations on fonts, font types as well as upper and lower case.

This testing was initially performed on a set of ISO 3166-1 two character codes, all of which are delegated or admissible as ccTLDs, and focused on visually confusable entries to establish the threshold for each of the 4 measurements. The threshold values essentially allow for IDN ccTLD applications to be as confusingly similar as any ISO 3166-1 pair of entries.

The threshold values derived from this experimentation were:

- DMTS inv(RT) - values less than 0.9 would indicate the entry is confusingly similar.
- DMTS error rate - values greater than 0.14 would indicate the entry is confusingly similar.
- GNG inv(RT) - values less than 0.77 would indicate the entry is confusingly similar.
- GNG error rate - values greater than 0.34 would indicate the entry is confusingly similar.

Further testing, which included the requested IDN ccTLD string against a number of ISO 3166-1 entries (selected for their potential for confusion with the requested string – see Section 6 of this report for details), was also carried out to generate measurements for this string for each version.

For an applied for string to be considered confusingly similar, there must be evidence that the candidate is highly similar to potentially-confusing ISO 3166-1 entries for both behavioral tasks. The DMTS task assesses memory confusion after brief delays, whereas the GNG task assesses the potential confusion of simultaneous glyphs.

For a given task, highly-similar refers to one or to both measures (Inv RT and error rate) exceeding the established threshold criterion (to exceed a given threshold both the mean and the 95% confidence interval must exceed the threshold). If only one of these two measures (invRT or error rate) exceeds threshold this is sufficient evidence for rejection for this task provided that the result cannot be due to a speed-accuracy trade-off. This pattern does not need to be in same font face for the given testing pair combination in both tasks.

Notes:

- This is simply a summary of the methodology that was developed by the research team in collaboration with the Panel to evaluate the candidate strings. A complete description of the methodology and the results can be found in the annexes of this document.
- Separate recommendations for upper and lower case versions of the candidate string. The Panel was requested to consider both upper and lower case versions of the candidate strings to evaluate if it is confusingly similar to any ISO 3166-1 entry in both upper and lower case. From a visual similarity point of view upper and lower case characters of the same letter are distinct entities – as such upper and lower case versions of the candidate strings needed to be tested separately. Given there is no scientific or policy basis as to how to combine these separate results of upper and lower case for IDN ccTLDs the Panel concluded it could only provide separate recommendations for each of these.

### 3 Panel Members and Research Team

Dr. Max Coltheart (chair), Emeritus Professor, Department of Cognitive Science, Macquarie University, Australia

Dr. Jonathan Grainger, Directeur de recherches au CNRS Aix-Marseille Université, France

Dr. Kevin Larson, United States

Research Institute: Department of Cognitive and Learning Sciences, Michigan Technological University, United States ; Leader of the research team: Professor Dr. Shane T. Mueller

## 4 Information on string to evaluate

Corresponding ISO3166 Entry: EU

A-Label: xn--qxa6a

U-Label: εv

Unicode Code Points: U+03B5 U+03C5

String in English: eu

String Language: Greek, Modern (1453-)

Language Scripts: Greek

## 5 Documents provided to the panel by ICANN

Submitted to the panel by ICANN:

- euIDN-reapplication-20012014
- Annex1-eu-reapplication
- Annex2-eu-reapplication
- EU IDN Tables

Submitted by the applicant in the 30 day window following the application:

- None

Documents requested by the panel:

- None

Other documents:

- DNS Stability Evaluation results – original application

## 6 Research Report Summary

The following is a summary of the research report for the string being considered.

The complete research report, which was submitted to the EPSRP by Dr. Mueller can be found in Annex A of this document.

The following is a listing of the version information as well as the characters used in the experimentation for this application:

## 6.1 Stimuli for Candidate: ευ/ EY in Greek

	<b>Serif lowercase Times New Roman</b>	<b>Sans serif lowercase Segoe UI</b>
<b>Evaluation target</b>	<b>ΕΥ</b>	<b>ευ</b>
Similar Latin	eu en ev	eu en ev
Dissimilar Latin comparisons:	xk zi ar	xk zi ar
Highly similar comparisons	εω	εω

<b>Evaluation Target</b>	<b>Serif uppercase Times new roman</b>	<b>Sans serif uppercase Segoe UI Uppercase</b>
	EY	EY
Similar Latin	EY FY EV	EY FY EV
Dissimilar Latin comparisons:	OS,AQ ,NG	OS,AQ,NG
Highly similar comparisons	ΞΥ,ΕΨ,ΞΨ	ΞΥ,ΕΨ,ΞΨ

Note: Some non-Latin character pairs were tested in early experimentation but these were not considered in the final analysis.

## 6.2 Results

The following is a summary of the results obtained.

### 6.2.1 DMTS

**Summary of invRT below threshold (if both are below 0.9 then the result is a fail - bold)**

<b>Pair:</b>	<b>Fontface</b>	<b>Mean</b>	<b>Confidence interval</b>
<i>eu</i>	<i>Sans Lowercase</i>	<i>0.871</i>	<i>0.935</i>
<b>EY</b>	<b>Sans Uppercase</b>	<b>0.545</b>	<b>0.66</b>
<b>FY</b>	<b>Sans Uppercase</b>	<b>0.78</b>	<b>0.847</b>
<b>EV</b>	<b>Sans Uppercase</b>	<b>0.817</b>	<b>0.896</b>
<b>EY</b>	<b>Serif Uppercase</b>	<b>0.7</b>	<b>0.82</b>
<b>FY</b>	<b>Serif Uppercase</b>	<b>0.802</b>	<b>0.872</b>
<b>EV</b>	<b>Serif Uppercase</b>	<b>0.789</b>	<b>0.866</b>

Italic indicates mean exceeds threshold. Bold indicates mean significantly exceeds threshold.

**Summary of Error rate above threshold (if both are greater than 0.14 then the result is a fail - bold)**

<b>Pair:</b>	<b>Fontface</b>	<b>Mean</b>	<b>Confidence interval</b>
<b>EY</b>	<b>Sans Uppercase</b>	<b>0.469</b>	<b>0.341</b>
<b>EY</b>	<b>Serif Uppercase</b>	<b>0.438</b>	<b>0.304</b>

Italic indicates mean exceeds threshold. Bold indicates mean significantly exceeds threshold.

### 6.2.2 Same/different go/no-go task

Summary of invRT below threshold (if both are below 0.77 then the result is a fail - bold)

<b>Pair:</b>	<b>Fontface</b>	<b>Mean:</b>	<b>Confidence interval</b>
<b>EY</b>	<b>Sans Uppercase</b>	<b>0.434</b>	<b>0.568</b>
<i>FY</i>	<i>Sans Uppercase</i>	<i>0.684</i>	<i>0.816</i>
<b>EV</b>	<b>Sans Uppercase</b>	<b>0.534</b>	<b>0.624</b>
<b>EY</b>	<b>Serif Uppercase</b>	<b>0.381</b>	<b>0.47</b>
<i>FY</i>	<i>Serif Uppercase</i>	<i>0.702</i>	<i>0.826</i>
<b>EV</b>	<b>Serif Uppercase</b>	<b>0.635</b>	<b>0.747</b>

Italic indicates mean exceeds threshold. Bold indicates mean significantly exceeds threshold.

Summary of Error rate above threshold (if both are above 0.34 then the result is a fail - bold)

<b>Pair:</b>	<b>Fontface</b>	<b>Mean:</b>	<b>Confidence interval</b>
<b>EY</b>	<b>Sans Uppercase</b>	<b>0.792</b>	<b>0.664</b>
<b>EV</b>	<b>Sans Uppercase</b>	<b>0.625</b>	<b>0.468</b>
<b>EY</b>	<b>Serif Uppercase</b>	<b>0.896</b>	<b>0.789</b>

Italic indicates mean exceeds threshold. Bold indicates mean significantly exceeds threshold.

## 7 Analysis by panel members

The panel reviewed the research report and was satisfied that it met the requirements it set out initially.

The additional material supplied by the applicant provided an interesting review of Letter recognition and confusability including articles published by Panel members. The recommendation of using *a single alphabet as a general strategy in the composition of .eu domain names* (in Greek) unfortunately could not be considered by the panel given it was not tasked to look into domain names preceding the IDN ccTLD.

This recommendation may be more relevant to ICANN when it considers how to implement the panel recommendations for this candidate string.



The panel was requested to consider both upper and lower case versions of the candidate string to evaluate if it is confusingly similar to any ISO 3166-1 entry in both upper and lower case. From a visual similarity point of view upper and lower case characters of the same letter are distinct entities or glyphs – as such upper and lower case versions of the candidate strings needed to be tested separately. Given there is no scientific or policy basis as to how to combine these separate results of upper and lower case for IDN ccTLDs the Panel concluded it could only provide separate recommendations for each of these.

For an applied for string to be considered confusingly similar, there must be evidence that the candidate is highly similar to potentially-confusing ISO 3166-1 entries for both behavioral tasks. The DMTS task assesses memory confusion after brief delays, whereas the GNG task assesses the potential confusion of simultaneous glyphs.

For a given task, highly-similar refers to one or to both measures (Inv RT and error rate) exceeding the established threshold criterion (to exceed a given threshold both the mean and the 95% confidence interval must exceed the threshold). If only one of these two measures (invRT or error rate) exceeds threshold this is sufficient evidence for rejection for this task provided that the result cannot be due to a speed-accuracy trade-off. This pattern does not need to be in same font face for the given testing pair combination in both tasks.

The established threshold criteria are:

- DMTS inv(RT) - values less than 0.9 would indicate the entry is confusingly similar.
- DMTS error rate - values greater than 0.14 would indicate the entry is confusingly similar.
- GNG inv(RT) - values less than 0.77 would indicate the entry is confusingly similar.
- GNG error rate - values greater than 0.34 would indicate the entry is confusingly similar.

The panel considered the research results for upper case and noted that the candidate string generated results which exceeded the thresholds in both DMTS and GNG for the ISO 3166-1 entries of EY and EV.

The panel also considered the research results for lower case and noted that the candidate string generated no results which exceeded the thresholds for both the mean and a 95% confidence interval.

The panel therefore concludes that the IDN ccTLD application in upper case should be considered confusingly similar to the following ISO 3166-1 entries: EV and EY.

The panel also concludes that the IDN ccTLD application in lower case should not be considered confusingly similar to any ISO 3166-1 entries.

Note: The full report of the EPSRP can be found in Annex B

## 8 Recommendations of the EPSRP

For the candidate string:

Corresponding ISO3166 Entry: EU  
A-Label: xn--qxa6a  
U-Label: ευ  
Unicode Code Points: U+03B5 U+03C5  
String in English: eu  
String Language: Greek, Modern (1453)  
Language Scripts: Greek

The panel recommends that the IDN ccTLD application in upper case should be considered confusingly similar to the following ISO 3166-1 entries: EV and EY.

The panel recommends that the IDN ccTLD application in lower case should not be considered confusingly similar to any ISO 3166-1 entries.

## Annex A - Results of the Research Team Experimentation

## **Results of the Research Team Experimentation**

### **Behavioral Evaluation of candidate 2-letter similarity using Match-to-sample task (DMTS)**

Candidate: EU in Greek. (epsilon upsilon)

This document evaluates the candidate with respect to its overall discriminability from other pairs, using a delayed match-to sample (two-alternative forced-choice) task. In this task, participants briefly see one candidate pairs on the screen, after which it is masked. Then, that pair plus a foil appears after a short delay, and they must identify which option was presented.

Note: Some non-Latin character pairs were tested but these were not considered in the final analysis.

#### **Presentation**

- Sans serif stimuli were displayed as rendered in the location bar of a popular internet browser running on Microsoft Windows. Serif and italic stimuli were obtained via screenshots from a word processing application using Times New Roman font face to match the size of the sans serif font (Approximately 10-11pt size, non-italic, non-bold with normal spacing).
- Participants were instructed to view the screen from a comfortable distance, to best match their naturalistic screen viewing conditions.

#### **Procedures**

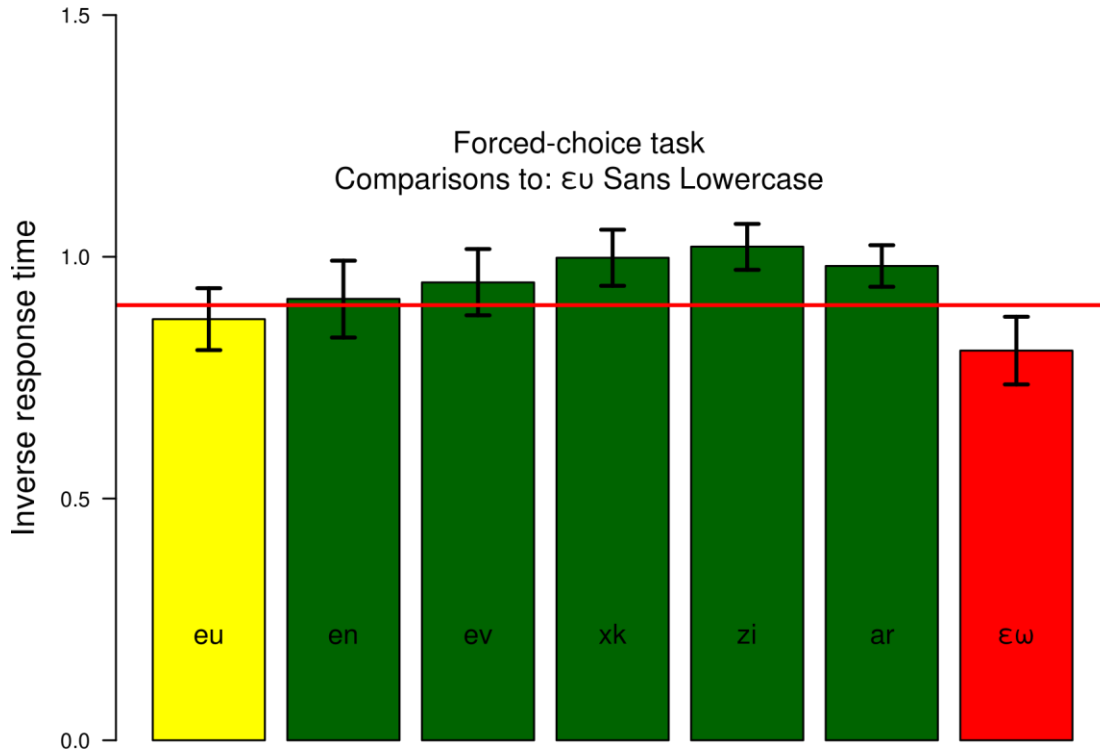
- Testing used two procedures: 1. A delayed match-to-sample forced-choice identification task, and 2. A go/no-go response same-different judgment task. The advantage of method 1 is that it tends to produce differences in response time based on confusability that are highly reliable with minimal observations, the advantage of method 2 is that it induces larger differences in accuracy, and requires a participant to detect a specific difference.
- Each test was performed in a blocked design in the same order across participants. Each set of stimuli will appear in a contiguous block. Testing was designed to assess the similarity between the target and (1) any of a set of highly-similar Latin character pairs in the same case (2) a set of 3-4 dissimilar Latin character pairs, and (3) any highly-similar comparisons, which may not directly bear on the decision, but may help to calibrate and validate the measures.

## **Participants**

- In this study, we intend to test 20 undergraduate students, primarily students of U.S. origin. Because Greek characters are relatively unfamiliar to them, and because they are experts in Latin orthography which is the orthography where the confusions are most likely to occur, they serve as a reasonable population for evaluating these characters sets to make inference about a general internet population

**Inverse response time: Sans Lowercase**

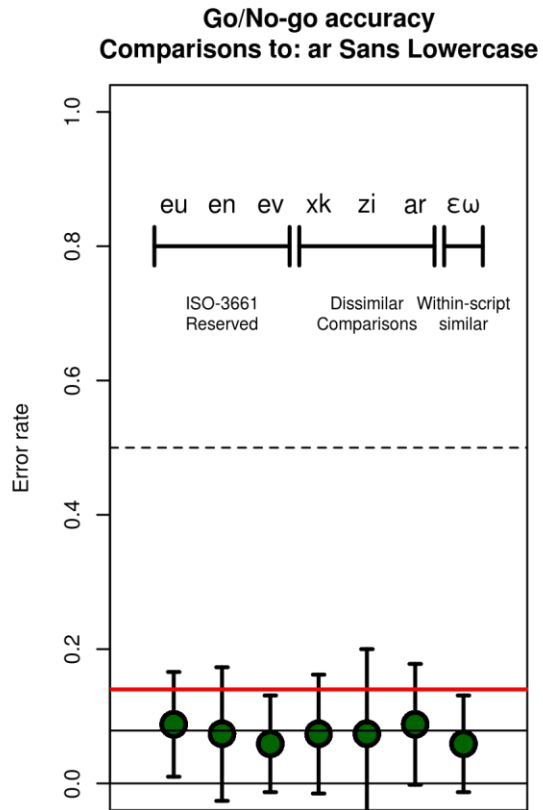
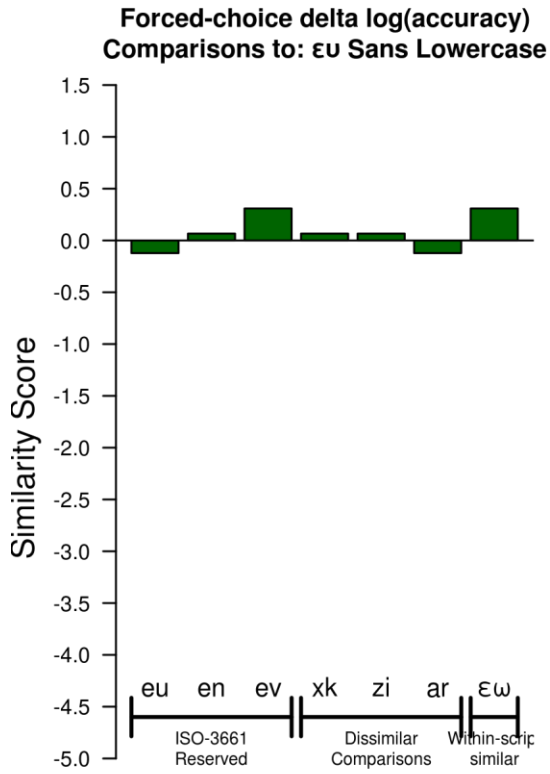
Critical value: 0.9



	mean:	sd:	N:	se:	5%	95%
eu	0.871	0.125	17	0.03	0.807	0.935
en	0.913	0.154	17	0.037	0.833	0.992
ev	0.947	0.133	17	0.032	0.879	1.016
xk	0.998	0.113	17	0.027	0.94	1.056
zi	1.021	0.092	17	0.022	0.973	1.068
ar	0.981	0.084	17	0.02	0.938	1.024
εω	0.806	0.136	17	0.033	0.736	0.876

**Error rate: Sans Lowercase**

Critical value: 0.14

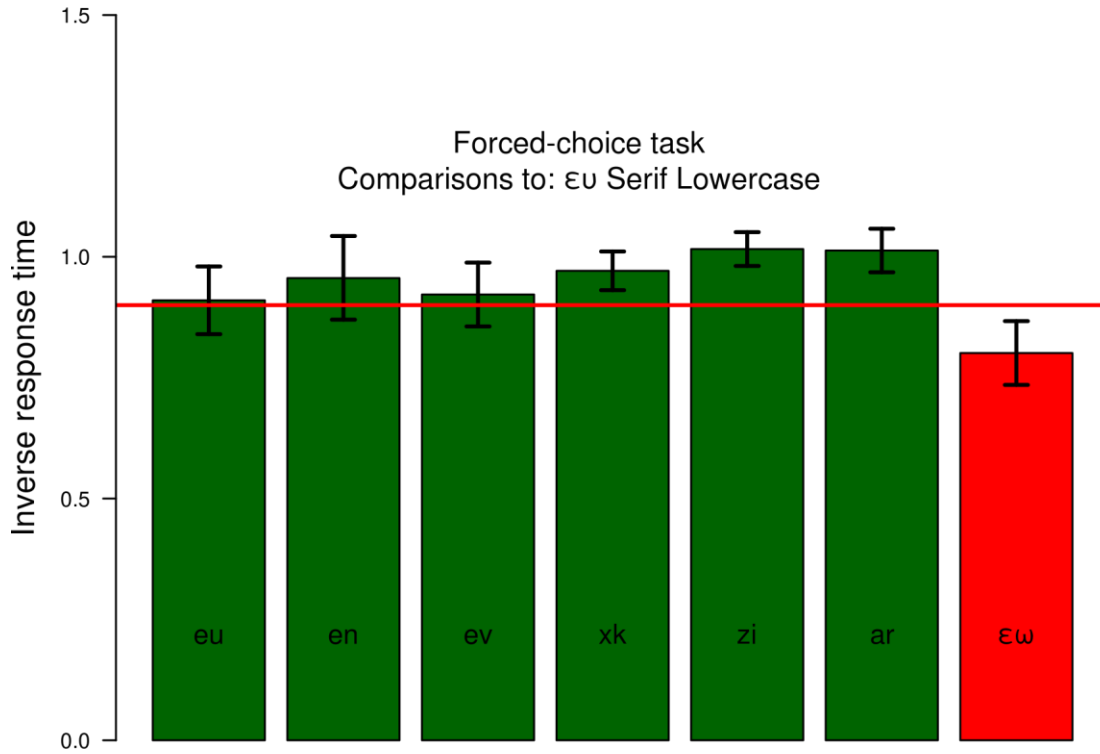


	mean:	sd:	N:	se:	5%	95%
eu	0.088	0.152	17	0.037	0.01	0.166
en	0.074	0.193	17	0.047	-0.026	0.173
ev	0.059	0.141	17	0.034	-0.013	0.131
xk	0.074	0.171	17	0.042	-0.015	0.162
zi	0.074	0.246	17	0.06	-0.053	0.2
ar	0.088	0.175	17	0.043	-0.002	0.178
εω	0.059	0.141	17	0.034	-0.013	0.131

Correlation between error rate and inverse RT: 0.2772

**Inverse response time: Serif Lowercase**

Critical value: 0.9

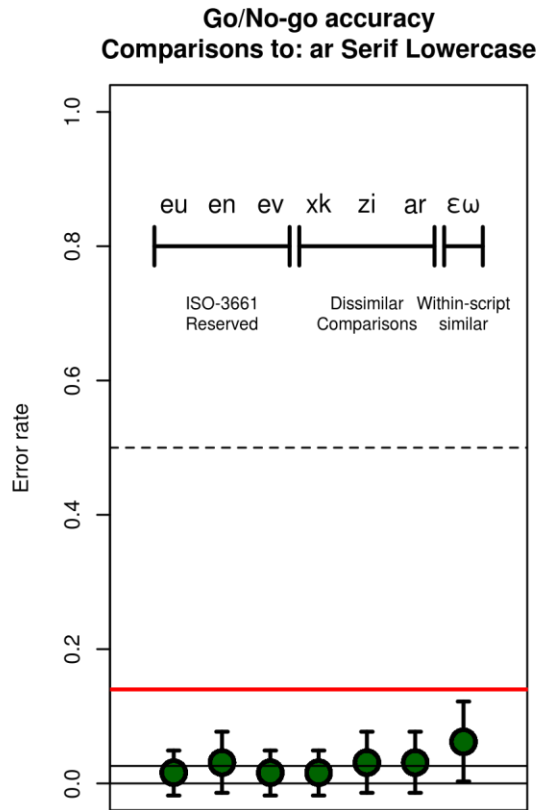
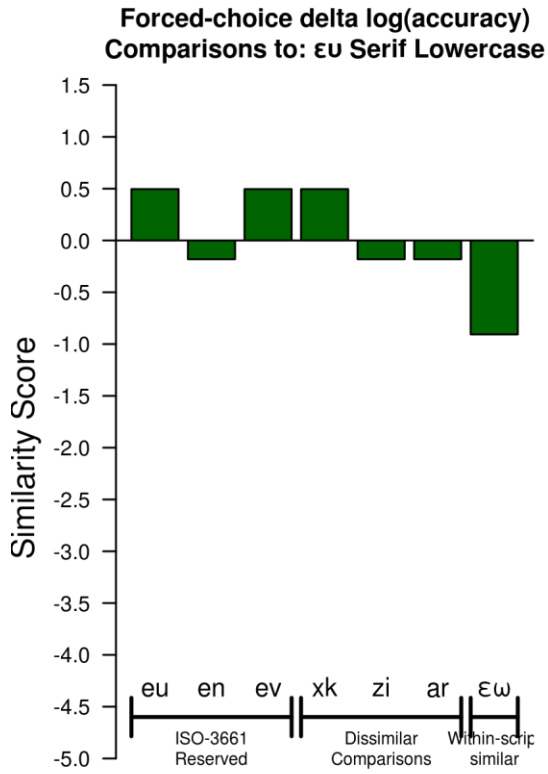


	mean:	sd:	N:	se:	5%	95%
eu	0.91	0.131	16	0.033	0.84	0.98
en	0.956	0.162	16	0.041	0.87	1.043
ev	0.922	0.124	16	0.031	0.856	0.988
xk	0.971	0.075	16	0.019	0.931	1.011
zi	1.016	0.066	16	0.016	0.981	1.051
ar	1.013	0.085	16	0.021	0.968	1.058
εω	0.801	0.124	16	0.031	0.735	0.867



**Error rate: Serif Lowercase**

Critical value: 0.14

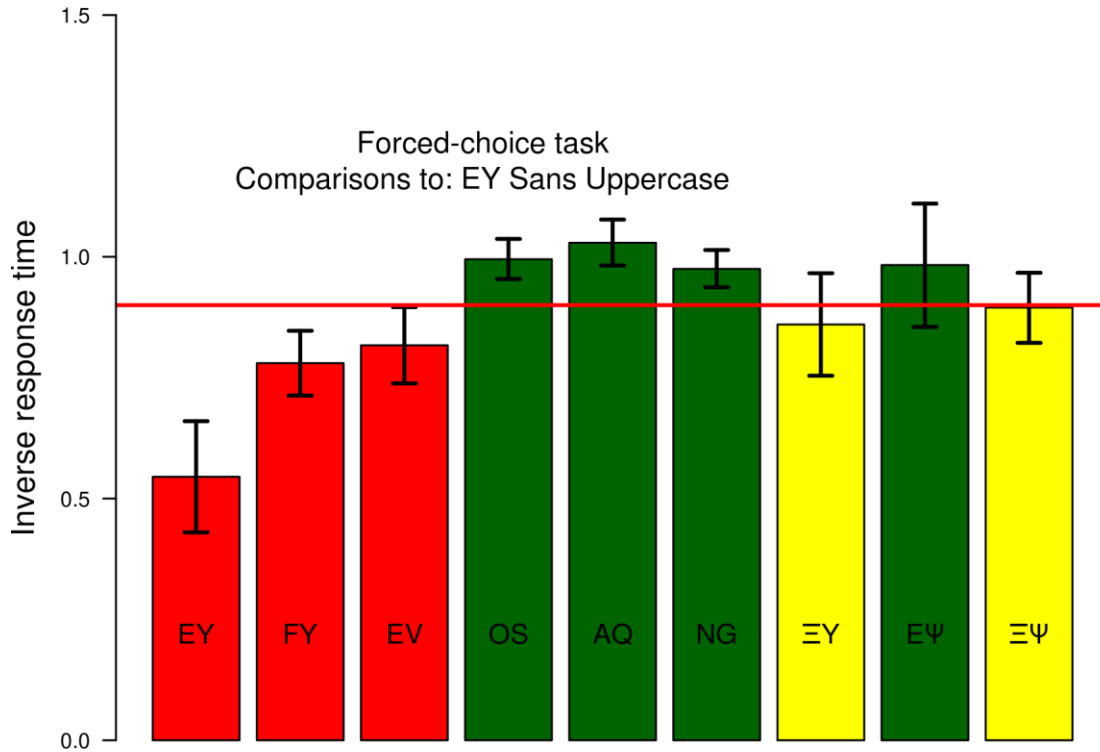


	mean:	sd:	N:	se:	5%	95%
eu	0.016	0.062	16	0.016	-0.018	0.049
en	0.031	0.085	16	0.021	-0.014	0.077
ev	0.016	0.062	16	0.016	-0.018	0.049
xk	0.016	0.062	16	0.016	-0.018	0.049
zi	0.031	0.085	16	0.021	-0.014	0.077
ar	0.031	0.085	16	0.021	-0.014	0.077
εω	0.062	0.112	16	0.028	0.003	0.122

Correlation between error rate and inverse RT: -0.5553

**Inverse response time: Sans Uppercase**

Critical value: 0.9



	mean:	sd:	N:	se:	5%	95%
EY	0.545	0.216	16	0.054	0.43	0.66
FY	0.78	0.125	16	0.031	0.713	0.847
EV	0.817	0.147	16	0.037	0.738	0.896
OS	0.995	0.078	16	0.02	0.954	1.037
AQ	1.029	0.089	16	0.022	0.982	1.077
NG	0.975	0.072	16	0.018	0.937	1.014
EY	0.86	0.198	16	0.05	0.754	0.966

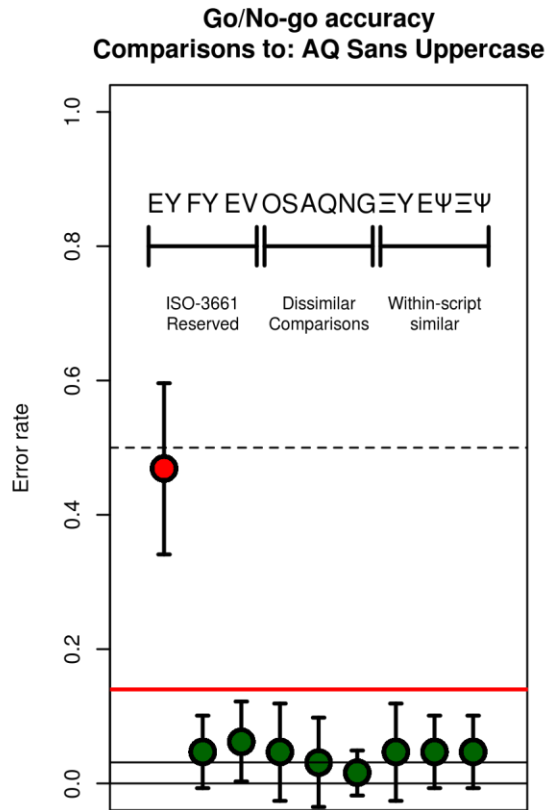
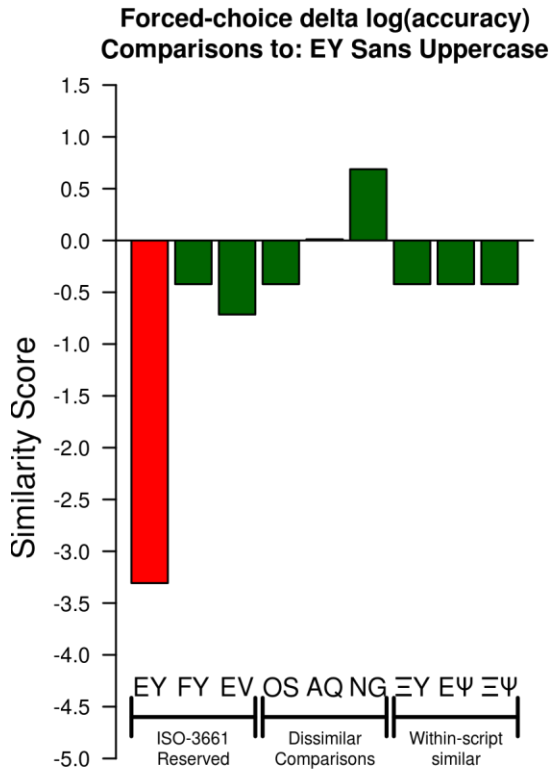
---

	<b>mean:</b>	<b>sd:</b>	<b>N:</b>	<b>se:</b>	<b>5%</b>	<b>95%</b>
E $\Psi$	0.983	0.239	16	0.06	0.855	1.11
$\Xi\Psi$	0.895	0.137	16	0.034	0.822	0.967

---

**Error rate: Sans Uppercase**

Critical value: 0.14



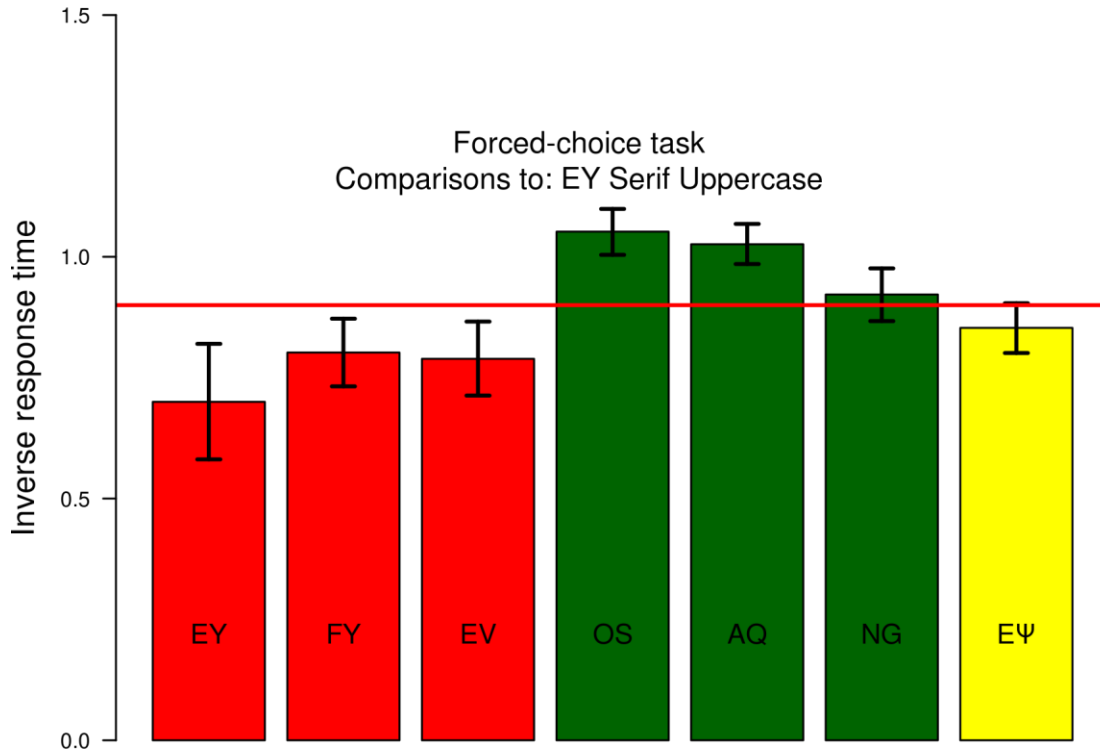
	mean:	sd:	N:	se:	5%	95%
EY	0.469	0.239	16	0.06	0.341	0.596
FY	0.047	0.101	16	0.025	-0.007	0.101
EV	0.062	0.112	16	0.028	0.003	0.122
OS	0.047	0.136	16	0.034	-0.026	0.119
AQ	0.031	0.125	16	0.031	-0.035	0.098
NG	0.016	0.062	16	0.016	-0.018	0.049
EY	0.047	0.136	16	0.034	-0.026	0.119

	<b>mean:</b>	<b>sd:</b>	<b>N:</b>	<b>se:</b>	<b>5%</b>	<b>95%</b>
EΨ	0.047	0.101	16	0.025	-0.007	0.101
ΞΨ	0.047	0.101	16	0.025	-0.007	0.101

Correlation between error rate and inverse RT: -0.8486

**Inverse response time: Serif Uppercase**

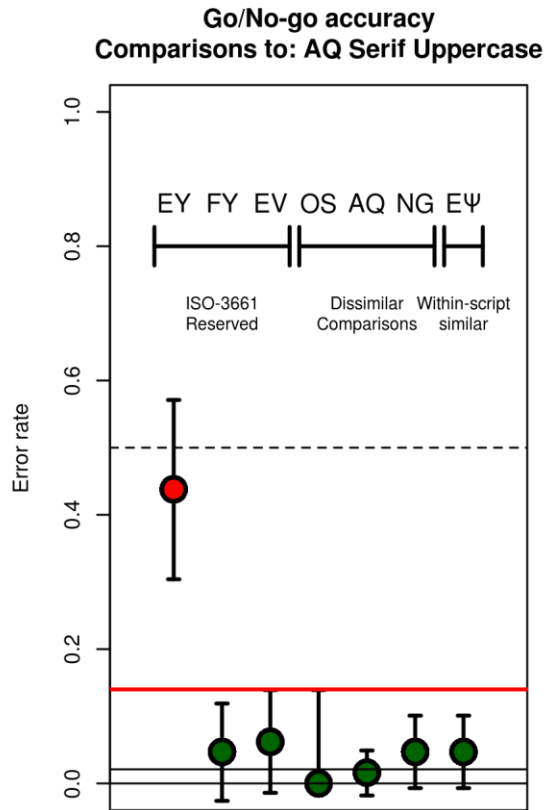
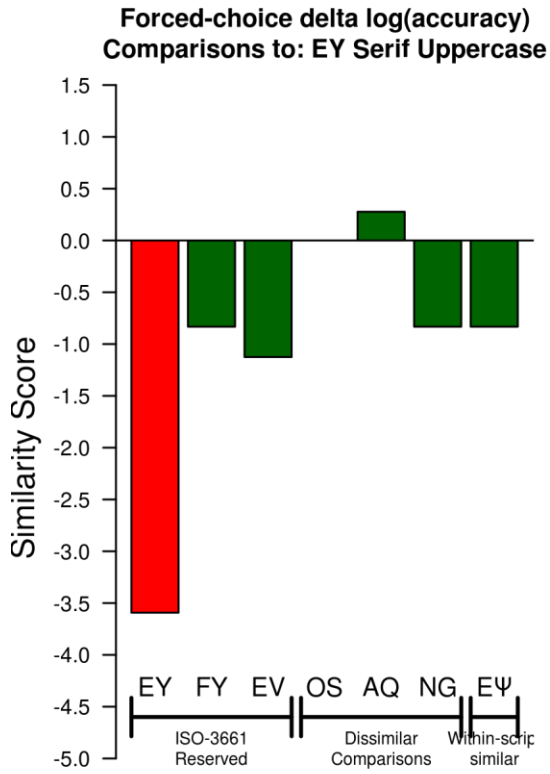
Critical value: 0.9



	mean:	sd:	N:	se:	5%	95%
EY	0.7	0.224	16	0.056	0.581	0.82
FY	0.802	0.131	16	0.033	0.732	0.872
EV	0.789	0.143	16	0.036	0.713	0.866
OS	1.052	0.089	16	0.022	1.004	1.099
AQ	1.026	0.077	16	0.019	0.985	1.068
NG	0.922	0.102	16	0.025	0.867	0.976
EΨ	0.853	0.097	16	0.024	0.801	0.904

**Error rate: Serif Uppercase**

Critical value: 0.14



	mean:	sd:	N:	se:	5%	95%
EY	0.438	0.25	16	0.062	0.304	0.571
FY	0.047	0.136	16	0.034	-0.026	0.119
EV	0.062	0.144	16	0.036	-0.014	0.139
OS	0	0	16	0	-0.014	0.139
AQ	0.016	0.062	16	0.016	-0.018	0.049
NG	0.047	0.101	16	0.025	-0.007	0.101
EΨ	0.047	0.101	16	0.025	-0.007	0.101

Correlation between error rate and inverse RT: -0.7039

### Summary of invRT below threshold

Pair:	Fontface	Mean:	Confidence interval	Threshold < 0.9
<i>eu</i>	<i>Sans Lowercase</i>	<i>0.871</i>	<i>0.935</i>	
<b>EY</b>	<b>Sans Uppercase</b>	<b>0.545</b>	<b>0.66</b>	
<b>FY</b>	<b>Sans Uppercase</b>	<b>0.78</b>	<b>0.847</b>	
<b>EV</b>	<b>Sans Uppercase</b>	<b>0.817</b>	<b>0.896</b>	
<b>EY</b>	<b>Serif Uppercase</b>	<b>0.7</b>	<b>0.82</b>	
<b>FY</b>	<b>Serif Uppercase</b>	<b>0.802</b>	<b>0.872</b>	
<b>EV</b>	<b>Serif Uppercase</b>	<b>0.789</b>	<b>0.866</b>	

Italic indicates mean exceeds threshold. Bold indicates mean significantly exceeds threshold.

### Summary of Error rate above threshold

Pair:	Fontface	Mean:	Confidence interval	Threshold > 0.14
<b>EY</b>	<b>Sans Uppercase</b>	<b>0.469</b>	<b>0.341</b>	
<b>EY</b>	<b>Serif Uppercase</b>	<b>0.438</b>	<b>0.304</b>	

Italic indicates mean exceeds threshold. Bold indicates mean significantly exceeds threshold.



## **Behavioral Evaluation of candidate 2-letter similarity using Same/different go/no-go task**

Candidate: EU in Greek. (epsilon upsilon)

This document evaluates the candidate with respect to its overall discriminability from other pairs, using a Go/No-go same-different task. In this task, participants see two pairs on the screen, left and right of center, outside their central vision. They must respond only when the two differ.

Note: Some non-Latin character pairs were tested but not considered in the final analysis.

### **Presentation**

- Sans serif stimuli were displayed as rendered in the location bar of a popular internet browser running on Microsoft Windows. Serif and italic stimuli were obtained via screenshots from a word processing application using Times New Roman font face to match the size of the sans serif font (Approximately 10-11pt size, non-italic, non-bold with normal spacing).
- Participants were instructed to view the screen from a comfortable distance, to best match their naturalistic screen viewing conditions.

### **Procedures**

- Testing used two procedures: 1. A delayed match-to-sample forced-choice identification task, and 2. A go/no-go response same-different judgment task. The advantage of method 1 is that it tends to produce differences in response time based on confusability that are highly reliable with minimal observations, the advantage of method 2 is that it induces larger differences in accuracy, and requires a participant to detect a specific difference.
- Each test was performed in a blocked design in the same order across participants. Each set of stimuli will appear in a contiguous block. Testing was designed to assess the similarity between the target and (1) any of a set of highly-similar Latin character pairs in the same case (2) a set of 3-4 dissimilar Latin character pairs, and (3) any highly-similar comparisons, which may not directly bear on the decision, but may help to calibrate and validate the measures.

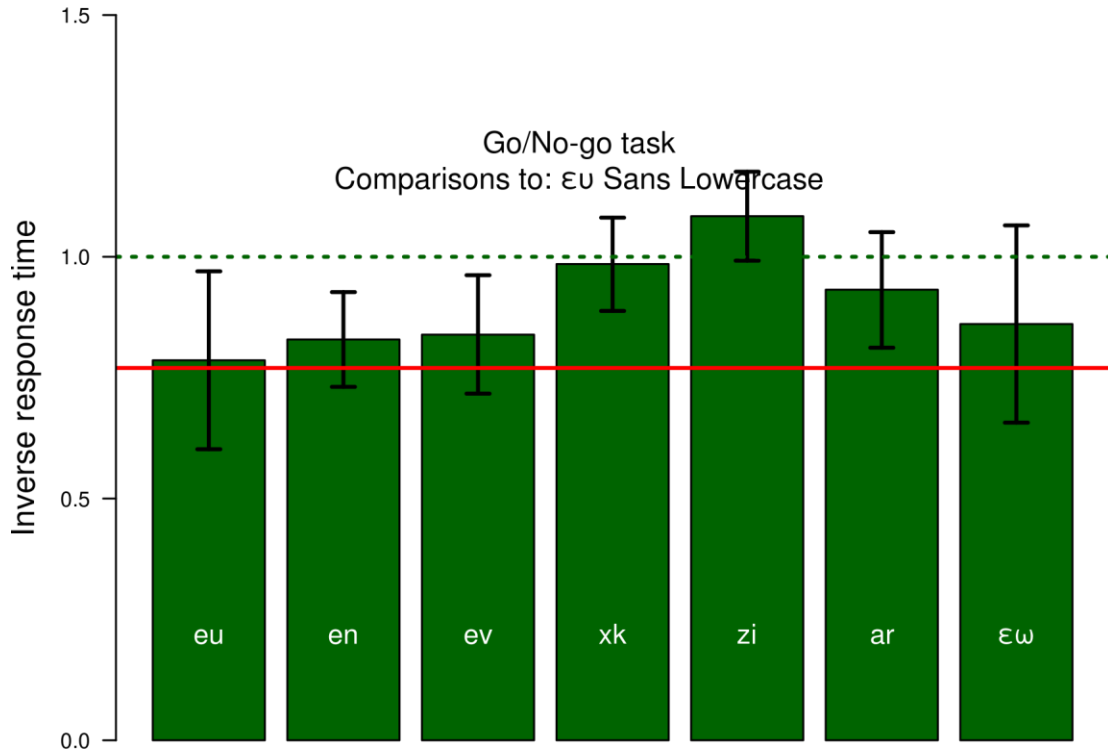
### **Participants**

- In this study, we intend to test 20 undergraduate students, primarily students of U.S. origin. Because Greek characters are relatively unfamiliar to them, and because they are experts in Latin orthography which is the orthography where the confusions are most

likely to occur, they serve as a reasonable population for evaluating these characters sets to make inference about a general internet population

**Inverse response time: Sans Lowercase**

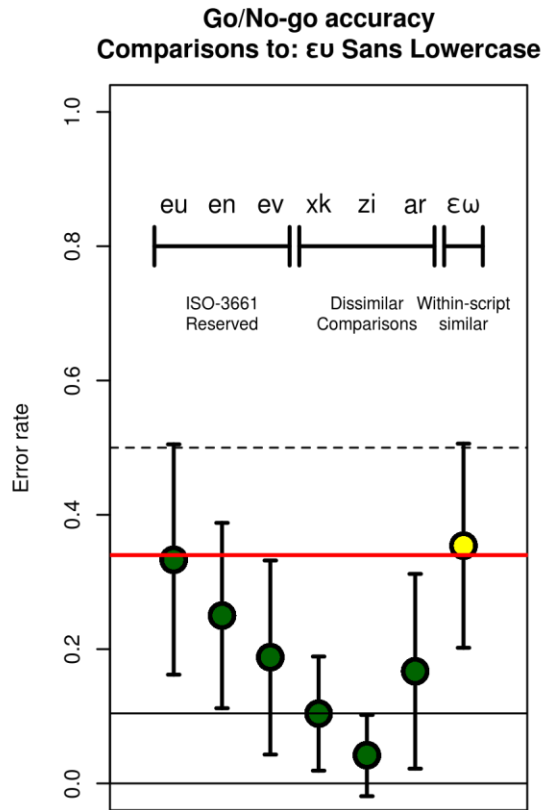
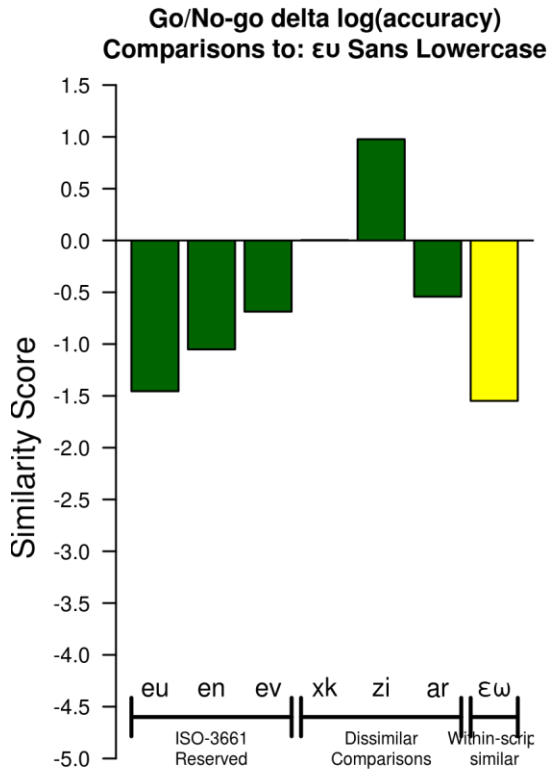
Critical value: 0.77



	mean:	sd:	N:	se:	5%	95%
eu	0.786	0.345	16	0.086	0.602	0.97
en	0.829	0.183	16	0.046	0.731	0.927
ev	0.839	0.23	16	0.057	0.717	0.962
xk	0.985	0.181	16	0.045	0.888	1.081
zi	1.084	0.173	16	0.043	0.992	1.176
ar	0.932	0.224	16	0.056	0.812	1.051
εω	0.861	0.382	16	0.096	0.657	1.065

**Error rate: Sans Lowercase**

Critical value: 0.34

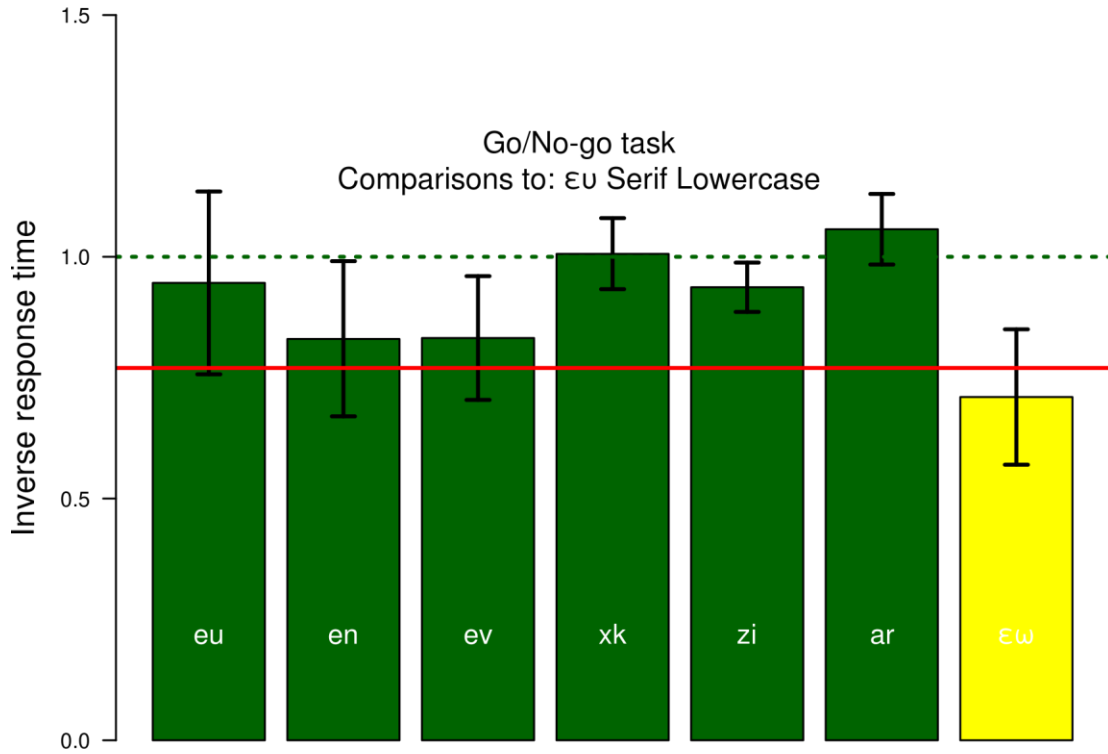


	mean:	sd:	N:	se:	5%	95%
eu	0.333	0.322	16	0.081	0.162	0.505
en	0.25	0.258	16	0.065	0.112	0.388
ev	0.188	0.271	16	0.068	0.043	0.332
xk	0.104	0.16	16	0.04	0.019	0.189
zi	0.042	0.114	16	0.028	-0.019	0.102
ar	0.167	0.272	16	0.068	0.022	0.312
εω	0.354	0.285	16	0.071	0.202	0.506

Correlation between error rate and inverse RT: -0.8673

**Inverse response time: Serif Lowercase**

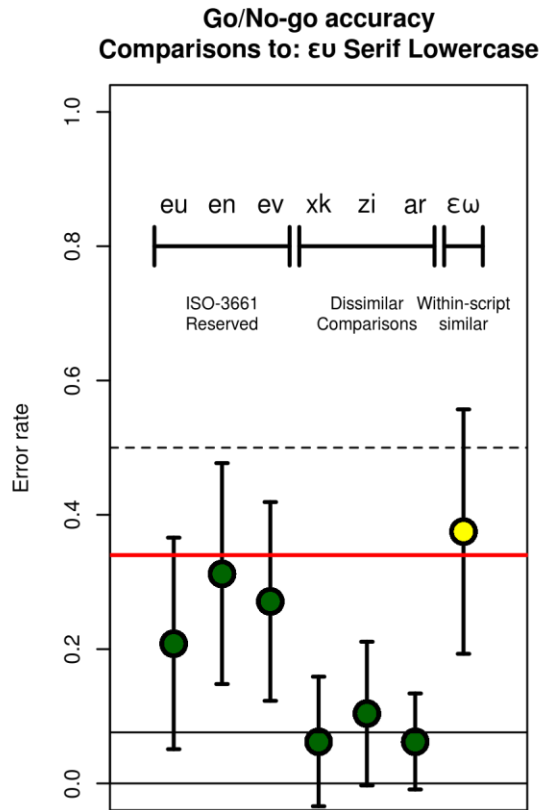
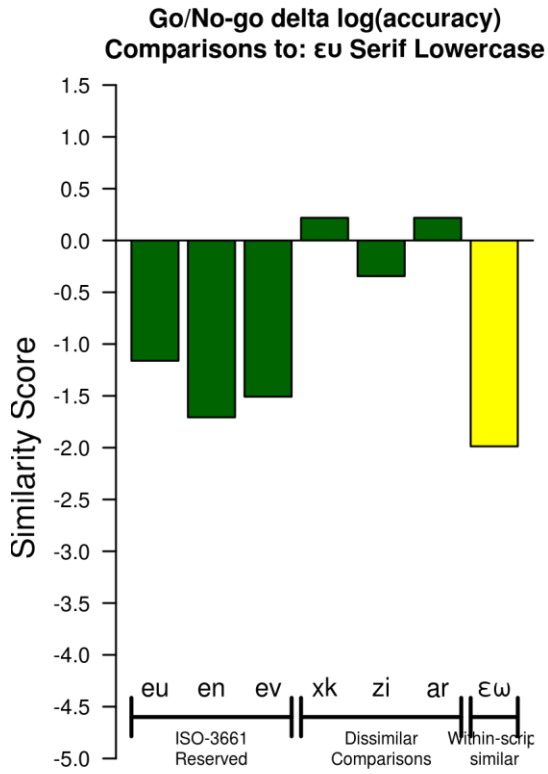
Critical value: 0.77



	mean:	sd:	N:	se:	5%	95%
eu	0.946	0.355	16	0.089	0.757	1.135
en	0.83	0.301	16	0.075	0.67	0.991
ev	0.832	0.24	16	0.06	0.704	0.96
xk	1.006	0.138	16	0.035	0.933	1.08
zi	0.937	0.096	16	0.024	0.886	0.988
ar	1.057	0.137	16	0.034	0.984	1.13
εω	0.71	0.263	16	0.066	0.57	0.85

**Error rate: Serif Lowercase**

Critical value: 0.34

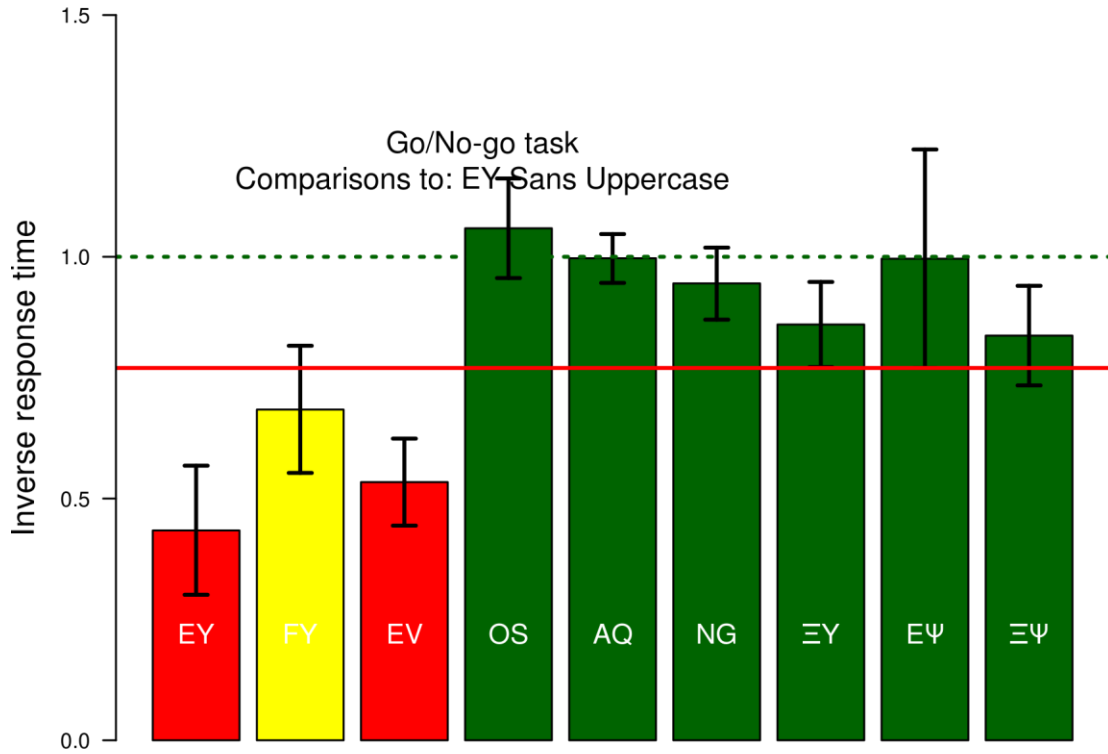


	mean:	sd:	N:	se:	5%	95%
eu	0.208	0.295	16	0.074	0.051	0.366
en	0.312	0.31	16	0.077	0.148	0.477
ev	0.271	0.278	16	0.07	0.123	0.419
xk	0.062	0.181	16	0.045	-0.034	0.159
zi	0.104	0.201	16	0.05	-0.003	0.211
ar	0.062	0.134	16	0.034	-0.009	0.134
εω	0.375	0.342	16	0.085	0.193	0.557

Correlation between error rate and inverse RT: -0.9476

**Inverse response time: Sans Uppercase**

Critical value: 0.77



	mean:	sd:	N:	se:	5%	95%
EY	0.434	0.251	16	0.063	0.301	0.568
FY	0.684	0.247	16	0.062	0.553	0.816
EV	0.534	0.169	16	0.042	0.444	0.624
OS	1.059	0.193	16	0.048	0.956	1.162
AQ	0.997	0.094	16	0.024	0.946	1.047
NG	0.945	0.14	16	0.035	0.87	1.019
EY	0.86	0.165	16	0.041	0.772	0.948

---

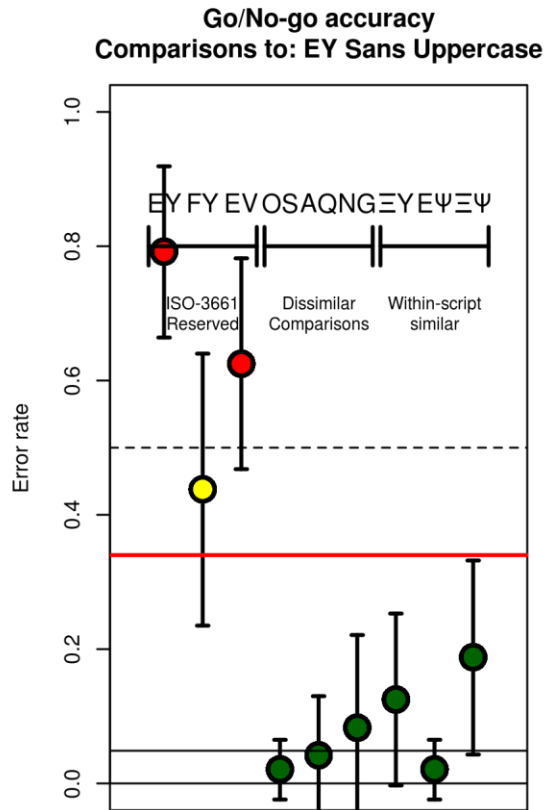
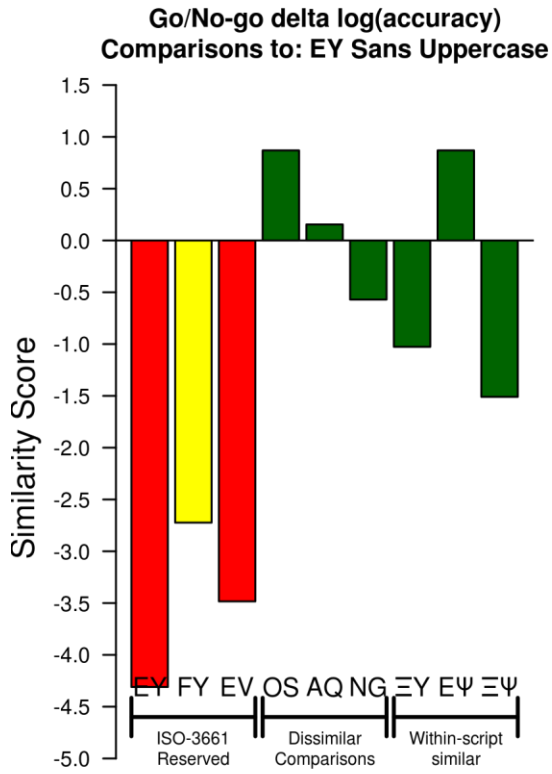
	<b>mean:</b>	<b>sd:</b>	<b>N:</b>	<b>se:</b>	<b>5%</b>	<b>95%</b>
E $\Psi$	0.996	0.424	16	0.106	0.77	1.222
$\Xi\Psi$	0.837	0.194	16	0.048	0.734	0.94

---



**Error rate: Sans Uppercase**

Critical value: 0.34



	mean:	sd:	N:	se:	5%	95%
EY	0.792	0.24	16	0.06	0.664	0.919
FY	0.438	0.379	16	0.095	0.235	0.64
EV	0.625	0.295	16	0.074	0.468	0.782
OS	0.021	0.083	16	0.021	-0.024	0.065
AQ	0.042	0.167	16	0.042	-0.047	0.13
NG	0.083	0.258	16	0.065	-0.054	0.221
EY	0.125	0.24	16	0.06	-0.003	0.253

---

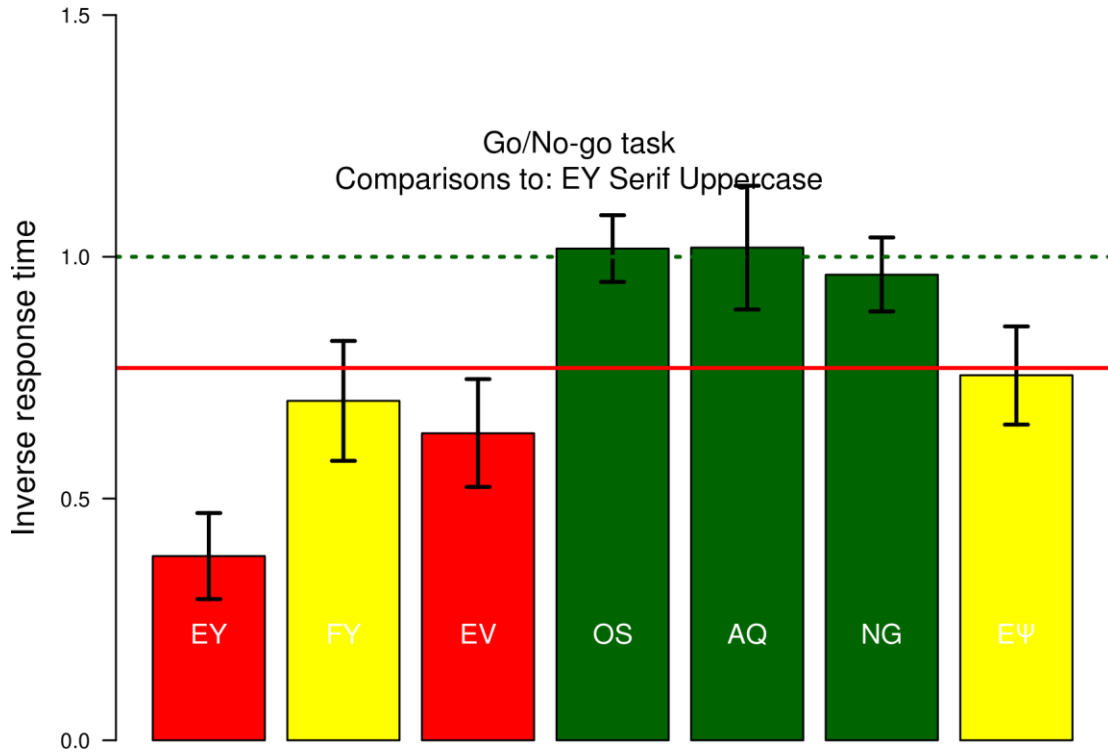
	<b>mean:</b>	<b>sd:</b>	<b>N:</b>	<b>se:</b>	<b>5%</b>	<b>95%</b>
E $\Psi$	0.021	0.083	16	0.021	-0.024	0.065
$\Xi\Psi$	0.188	0.271	16	0.068	0.043	0.332

---

Correlation between error rate and inverse RT: -0.9881

**Inverse response time: Serif Uppercase**

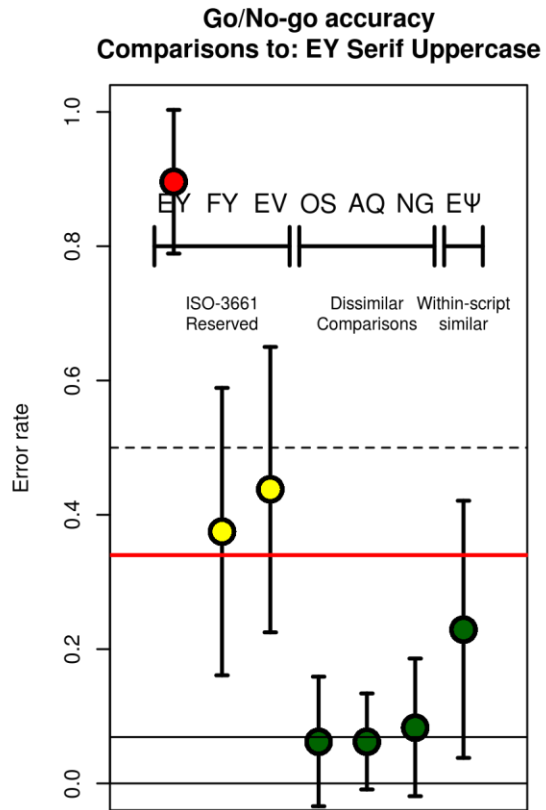
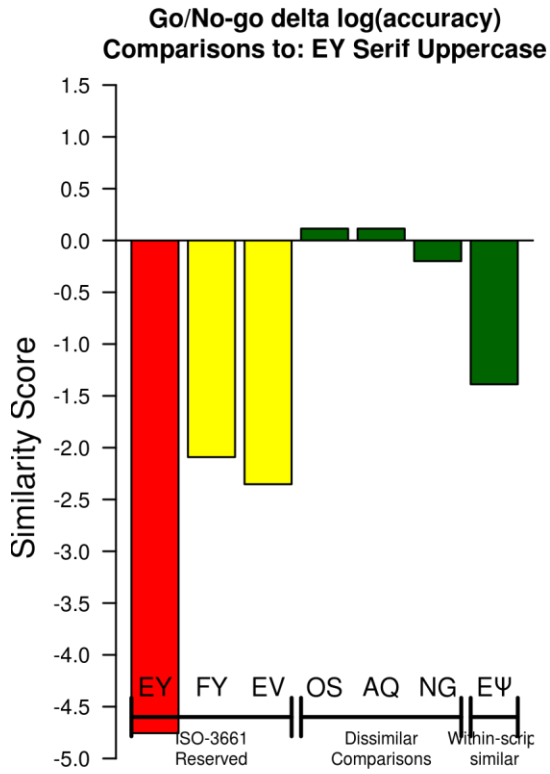
Critical value: 0.77



	mean:	sd:	N:	se:	5%	95%
EY	0.381	0.167	16	0.042	0.292	0.47
FY	0.702	0.233	16	0.058	0.578	0.826
EV	0.635	0.209	16	0.052	0.524	0.747
OS	1.017	0.129	16	0.032	0.948	1.086
AQ	1.019	0.24	16	0.06	0.891	1.147
NG	0.963	0.144	16	0.036	0.887	1.04
EΨ	0.755	0.19	16	0.047	0.653	0.856

**Error rate: Serif Uppercase**

Critical value: 0.34



	mean:	sd:	N:	se:	5%	95%
EY	0.896	0.201	16	0.05	0.789	1.003
FY	0.375	0.401	16	0.1	0.161	0.589
EV	0.438	0.398	16	0.1	0.225	0.65
OS	0.062	0.181	16	0.045	-0.034	0.159
AQ	0.062	0.134	16	0.034	-0.009	0.134
NG	0.083	0.192	16	0.048	-0.019	0.186
EΨ	0.229	0.359	16	0.09	0.038	0.421

Correlation between error rate and inverse RT: -0.973

### Summary of invRT below threshold

Pair:	Fontface	Mean:	Confidence interval	Threshold < 0.77
<b>EY</b>	<b>Sans Uppercase</b>	<b>0.434</b>	<b>0.568</b>	
<i>FY</i>	<i>Sans Uppercase</i>	<i>0.684</i>	<i>0.816</i>	
<b>EV</b>	<b>Sans Uppercase</b>	<b>0.534</b>	<b>0.624</b>	
<b>EY</b>	<b>Serif Uppercase</b>	<b>0.381</b>	<b>0.47</b>	
<i>FY</i>	<i>Serif Uppercase</i>	<i>0.702</i>	<i>0.826</i>	
<b>EV</b>	<b>Serif Uppercase</b>	<b>0.635</b>	<b>0.747</b>	

Italic indicates mean exceeds threshold. Bold indicates mean significantly exceeds threshold.

### Summary of Error rate above threshold

Pair:	Fontface	Mean:	Confidence interval	Threshold > 0.34
<b>EY</b>	<b>Sans Uppercase</b>	<b>0.792</b>	<b>0.664</b>	
<b>EV</b>	<b>Sans Uppercase</b>	<b>0.625</b>	<b>0.468</b>	
<b>EY</b>	<b>Serif Uppercase</b>	<b>0.896</b>	<b>0.789</b>	

Italic indicates mean exceeds threshold. Bold indicates mean significantly exceeds threshold.

Annex B - Final Report of the EPSRP for the application for EU in  
Greek

## **Final Report of the EPSRP for the application for EU in Greek**

1. We are using two tasks: Delayed Matching to Sample (DMTS) and Go/NoGo (GNG).
2. From each task we want to derive two measures of similarity, making sure that one of these measures pays attention to response speed and the other pays attention to response accuracy. Jonathan suggested a simple solution:  $1/RT$  (taking the inverse makes RT distributions much closer to normal; raw RT distributions typically have considerable positive skew) and percent correct. The advantages of these two measures is that they are simple to explain and that they do, taken together, capture both speed and accuracy. We agreed on 5 June that we would use  $1/RT$  i.e.  $inv(RT)$  and percent correct as our two measures.
3. The proposed new DNs to evaluate (in several fonts, in both uppercase and lowercase) are  $\epsilon\nu$ /EY in Greek
4. The data against which we will evaluate any proposed new DN combination are similarity measures from a set of DNs that are already being used or reserved for future use. Let's call these sets *reference sets*. A specific reference set was chosen for each candidate DN; these sets are listed in Appendix A. Our basic approach is this: if in an experiment involving the reference set plus the new proposed DN, the average similarity of the new DN to any member of its reference set is higher than the set of average similarities of the reference set to all the other members of the reference set, that is a negative result for the new proposed DN. This is done in three steps:

*Step (a):* We measure the similarity of the candidate DN to all members of its reference set (Appendix A). This provides us with a mean and one-sided 95% confidence interval for every comparison of the DN with each member of the reference set.

*Step (b):* We measure the similarity of pairs of existing DNs (the anchor set - Appendix B) and use the highest observed similarity as the criterion against which the similarities measured in Step (a) will be evaluated. These criteria are selected to be levels consistent across several different studies.

*Step (c):* To be rejected, there must be evidence that the candidate is highly similar to potentially-confusing IDNs for both behavioral tasks. The DMTS task assesses memory confusion after brief delays, whereas the GNG task assesses the potential confusion of simultaneous glyphs, and so our proposal is that confusability should be demonstrated in both tasks.

For a given task, highly-similar could refer to one or to both measures (Inv RT and error rate) passing the established threshold criterion. If only one of these two measures passes threshold, we treat this as sufficient evidence for rejection provided that the result cannot be due to a speed-accuracy tradeoff. We recommend that this pattern does not need to hold for any

single fontface/IDN combination, but for at least one IDN/fontface in each task.

5. To compare the similarity of the new proposed DN to the set of similarities of the reference set we calculated the average similarity value for each subject across all the items in the reference set and construct a one-sided 95% confidence interval from that set of subject means. This produced a critical value for each of our four measures i.e. a value at the end of the one-sided 95% confidence interval. The resulting cutoff critical values were:

DMTS inv(RT): <0.9

DMTS error rate: >0.14

GNG inv(RT): <.77

GNG error rate: >.34

If the similarity of any new proposed DN to the members of the reference set is outside this 95% confidence interval for both tasks, that is a negative result for the new proposed DN.

The procedures by which we arrived at these values is summarized in Appendix B and described in detail in the documents dmts-anchors.pdf and gonogo-anchors.pdf.

## 6. Results

### DMTS

**Summary of invRT below threshold (if both are below 0.9 then the result is a fail - bold)**

<u>Pair:</u>	<u>Fontface</u>	<u>Mean</u>	<u>Confidence interval</u>
<i>eu</i>	<i>Sans Lowercase</i>	<i>0.871</i>	<i>0.935</i>
<b>EY</b>	<b>Sans Uppercase</b>	<b>0.545</b>	<b>0.66</b>
<b>FY</b>	<b>Sans Uppercase</b>	<b>0.78</b>	<b>0.847</b>
<b>EV</b>	<b>Sans Uppercase</b>	<b>0.817</b>	<b>0.896</b>
<b>EY</b>	<b>Serif Uppercase</b>	<b>0.7</b>	<b>0.82</b>
<b>FY</b>	<b>Serif Uppercase</b>	<b>0.802</b>	<b>0.872</b>
<b>EV</b>	<b>Serif Uppercase</b>	<b>0.789</b>	<b>0.866</b>

Italic indicates mean exceeds threshold. Bold indicates mean significantly exceeds threshold.

**Summary of Error rate above threshold (if both are greater than 0.14 then the**



result is a fail - bold)

<u>Pair:</u>	<u>Fontface</u>	<u>Mean</u>	<u>Confidence interval</u>
EY	Sans Uppercase	<b>0.469</b>	<b>0.341</b>
EY	Serif Uppercase	<b>0.438</b>	<b>0.304</b>

Italic indicates mean exceeds threshold. Bold indicates mean significantly exceeds threshold.

### Same/different go/no-go task

Summary of invRT below threshold (if both are below 0.77 then the result is a fail - bold)

<u>Pair:</u>	<u>Fontface</u>	<u>Mean:</u>	<u>Confidence interval</u>
EY	Sans Uppercase	<b>0.434</b>	<b>0.568</b>
<i>FY</i>	<i>Sans Uppercase</i>	<i>0.684</i>	<i>0.816</i>
EV	Sans Uppercase	<b>0.534</b>	<b>0.624</b>
EY	Serif Uppercase	<b>0.381</b>	<b>0.47</b>
<i>FY</i>	<i>Serif Uppercase</i>	<i>0.702</i>	<i>0.826</i>
EV	Serif Uppercase	<b>0.635</b>	<b>0.747</b>

Italic indicates mean exceeds threshold. Bold indicates mean significantly exceeds threshold.

Summary of Error rate above threshold (if both are above 0.34 then the result is a fail - bold)

<u>Pair:</u>	<u>Fontface</u>	<u>Mean:</u>	<u>Confidence interval</u>
EY	Sans Uppercase	<b>0.792</b>	<b>0.664</b>
EV	Sans Uppercase	<b>0.625</b>	<b>0.468</b>
EY	Serif Uppercase	<b>0.896</b>	<b>0.789</b>

Italic indicates mean exceeds threshold. Bold indicates mean significantly exceeds threshold.

## 7. Conclusion

In the case of EU, EY (Greek upper case)-EV and EY (Latin upper case) comparisons exceeded the threshold criterion in all cases, and so the decision to reject is clear.



APPENDIX A: Reference sets and testing plans for each candidate DN.

Testing plan for : Candidate: ευ/ EY in Greek

Stimuli:

	<b>Serif lowercase Times New Roman</b>	<b>Sans serif lowercase Segoe UI</b>
<b>Evaluation target</b>	<b>ευ</b>	<b>ευ</b>
Similar Latin	eu en ev	eu en ev
Dissimilar Latin comparisons:	xk zi ar	xk zi ar
Highly similar comparisons	εω	εω

<b>Evaluation Target</b>	<b>Serif uppercase Times new roman</b>	<b>Sans serif uppercase Segoe UI Uppercase</b>
	EY	EY
Similar Latin	EY FY EV	EY FY EV
Dissimilar Latin comparisons:	OS,AQ ,NG	OS,AQ,NG
Highly similar comparisons	ΞΥ,ΕΨ,ΞΨ	ΞΥ,ΕΨ,ΞΨ

## APPENDIX B:

General procedures for using the anchor sets to establish the critical values for the DMTS and GNG 1/RT and error measures. For full details of these procedures please consult the research results.

Candidate: Latin Comparison anchor sets

The purpose of these is to establish a set of high-similarity pairs that have an acceptable level of confusability/similarity. Nine pairs were selected from the highly-confusable pairings of the following letter sets, and measures compared to those same candidates with respect to dissimilar letter combinations. Each study and task contained two blocks of these trials. A single set of criteria was chosen based on all three studies.

Stimuli:

- it and lt
- fi and fj
- ai, al, at
- cx and ex

## **Presentation**

- Sans serif stimuli were displayed as rendered in the location bar of a popular internet browser running on Microsoft Windows. Serif and italic stimuli were obtained via screenshots from a word processing application using Times New Roman font face to match the size of the sans serif font (Approximately 10-11pt size, non-italic, non-bold with normal spacing).
- Participants were instructed to view the screen from a comfortable distance, to best match their naturalistic screen viewing conditions.

## **Procedures**

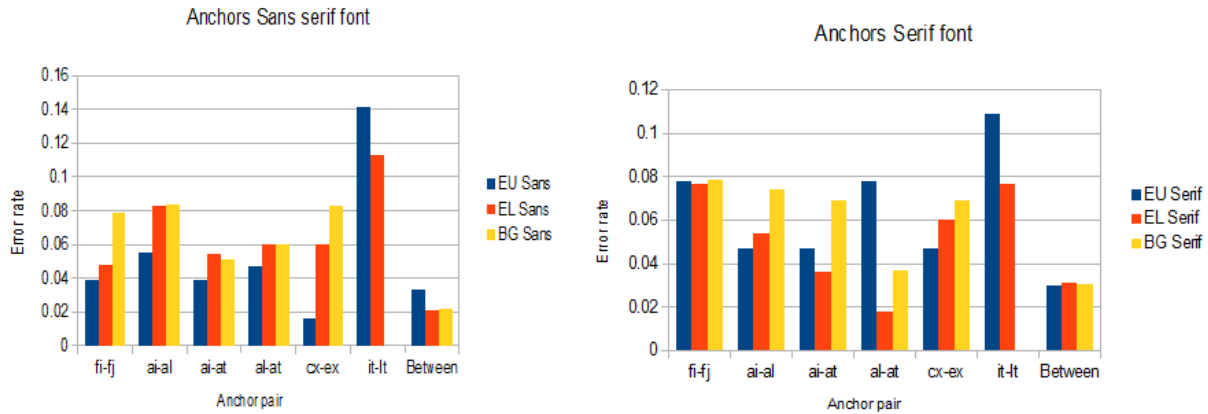
Testing used two procedures: 1. A delayed match-to-sample forced-choice identification task, and 2. A go/no-go response same-different judgment task. The advantage of method 1 is that it tends to produce differences in response time based on confusability that are highly reliable with minimal observations, the advantage of method 2 is that it induces larger differences in accuracy, and requires a participant to detect a specific difference.

Each test was performed in a blocked design in the same order across participants. Each set of stimuli will appear in a contiguous block. Testing was designed to assess the similarity between the target and (1) any of a set of highly-similar Latin character pairs in the same case (2) a set of 3-4 dissimilar Latin character pairs, and (3) any highly-similar comparisons, which may not directly bear on the decision, but may help to calibrate and validate the measures.

## **Participants**

In this study, we intend to test 20 undergraduate students, primarily students of U.S. origin. Because they are experts in Latin orthography, which is the orthography where the confusions are most likely to occur, they serve as a reasonable population for evaluating these characters sets to make inference about a general internet population

## DMTS Anchor Summary



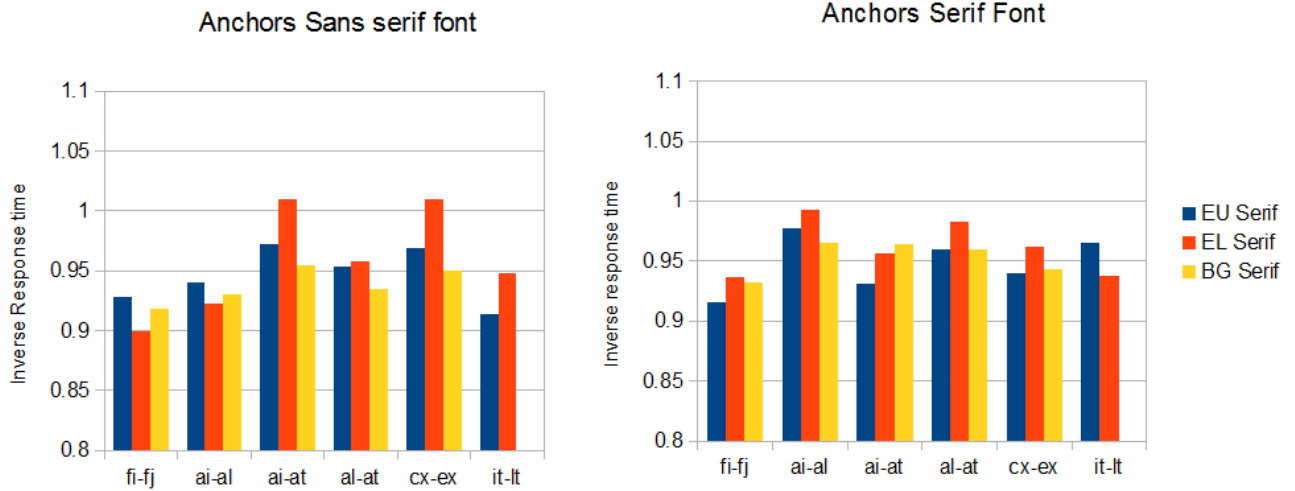
### Error Rate

Option	EU Sans	EL Sans	BG Sans
fi-fj	0.039	0.048	0.0787
ai-al	0.055	0.083	0.0833
ai-at	0.039	0.054	0.0509
al-at	0.047	0.06	0.0602
cx-ex	0.016	0.06	0.0827
it-lt	0.141	0.113	
<b>Between</b>	<b>0.033</b>	<b>0.021</b>	<b>0.0217</b>

Option	EU Serif	EL Serif	BG Serif
fi-fj	0.078	0.077	0.0787
ai-al	0.047	0.054	0.0741
ai-at	0.047	0.036	0.0694
al-at	0.078	0.018	0.037
cx-ex	0.047	0.06	0.0694
it-lt	0.109	0.077	
<b>Between</b>	<b>0.03</b>	<b>0.031</b>	<b>0.0306</b>

- In the tables and figures, EU/EL/BG indicate the study in which the data were collected, the stimuli were not visually different and design differed minimally.
- it-lt has the highest error rate (average .127; max .14). Overall dissimilar error rate is 2-3%, but this tends to be a bit higher for it-lt. This is 3-4 times the baseline error rate.
- Test-retest reliability for Sans is .90 ; serif is .36
- Adjusting accuracy (by subtracting or dividing by baseline) reduces test-retest reliability.
- **Recommendation: use .14 as criterion.**

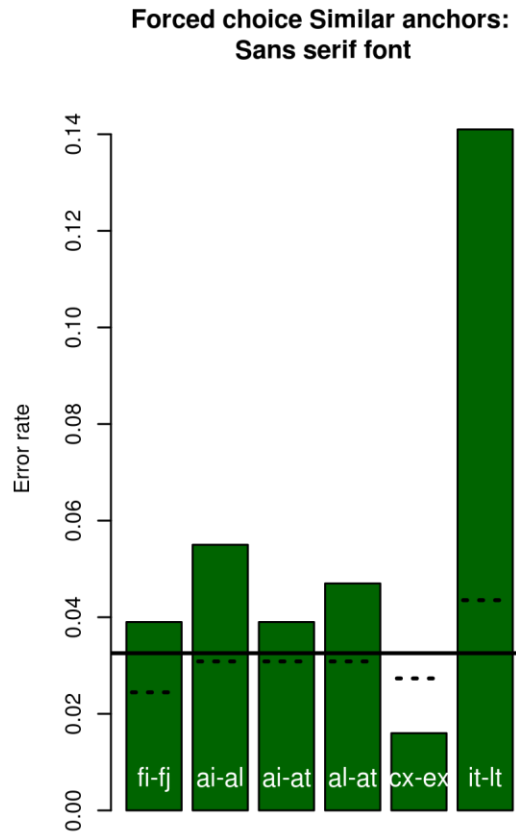
## Inverse Response Time



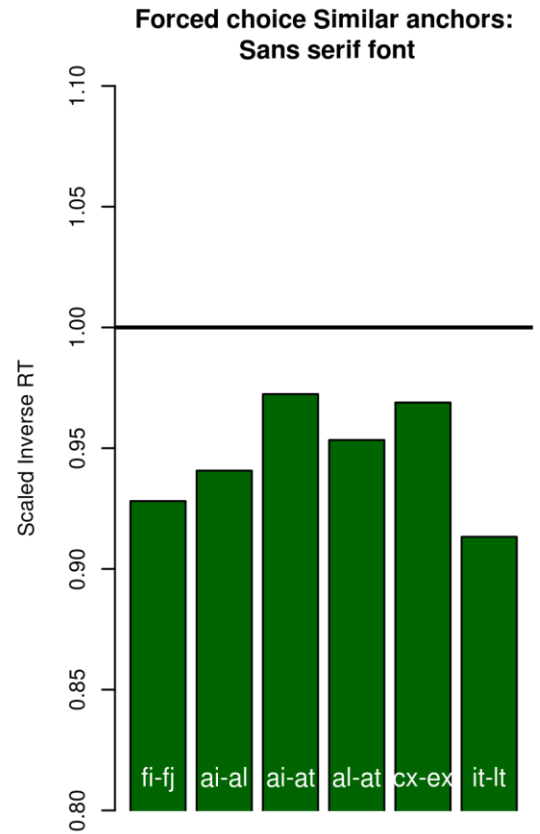
Option	EU Sans	EL Sans	BG Sans
fi-fj	0.9281	0.8995	0.918
ai-al	0.9407	0.9225	0.93
ai-at	0.9724	1.0096	0.955
al-at	0.9534	0.9584	0.935
cx-ex	0.9689	1.01	0.95
it-lt	0.9133	0.9483	-

Option	EU Serif	EL Serif	BG Serif
fi-fj	0.9155	0.9371	0.932
ai-al	0.9773	0.9925	0.965
ai-at	0.9316	0.9561	0.964
al-at	0.9596	0.9826	0.96
cx-ex	0.9401	0.962	0.943
it-lt	0.9648	0.9382	-

- Overall lowest Inverse RT (worst performance) is fi-fj Sans, averaging .915, with lowest of .8995.
- For sans, test-retest reliability was {.78, .98,.99}; for serif, {.63,.76,.72}.
- **Recommendation: Use 0.9 as criterion.**



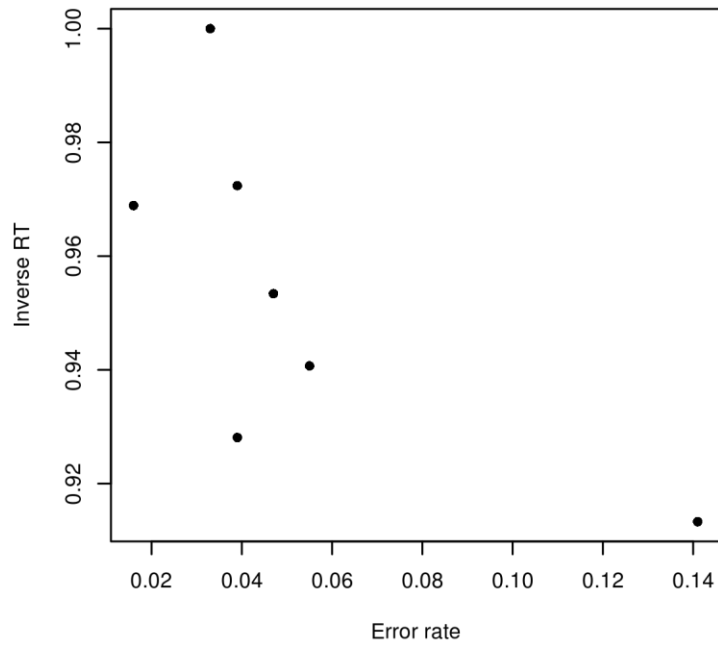
Candidate: EU in Greek. (epsilon upsilon)



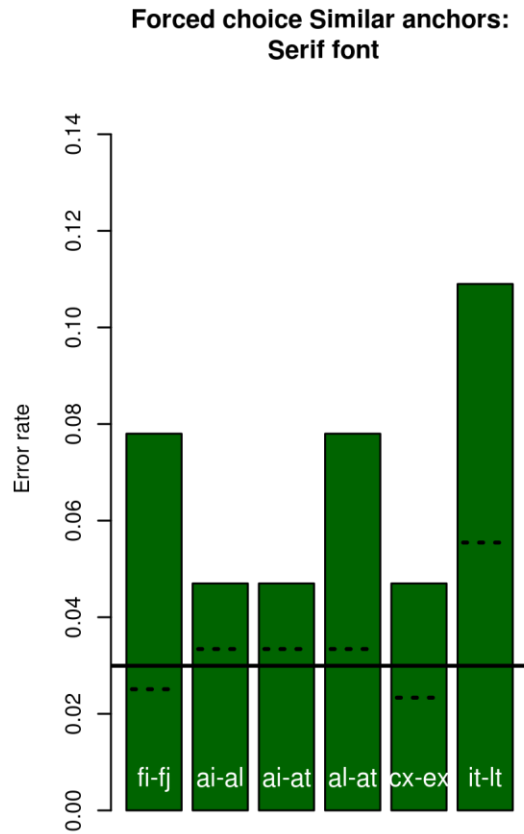
Candidate: EU in Greek. (epsilon upsilon)

Option	Error rate	Between error rate	Inverse RT	Log-odds delta accuracy
fi-fj	0.039	0.024	0.9281	-0.484
ai-al	0.055	0.031	0.9407	-0.597
ai-at	0.039	0.031	0.9724	-0.244
al-at	0.047	0.031	0.9534	-0.597
cx-ex	0.016	0.027	0.9689	0.571
it-lt	0.141	0.044	0.9133	-1.28
Between	0.033	0.033	1	0

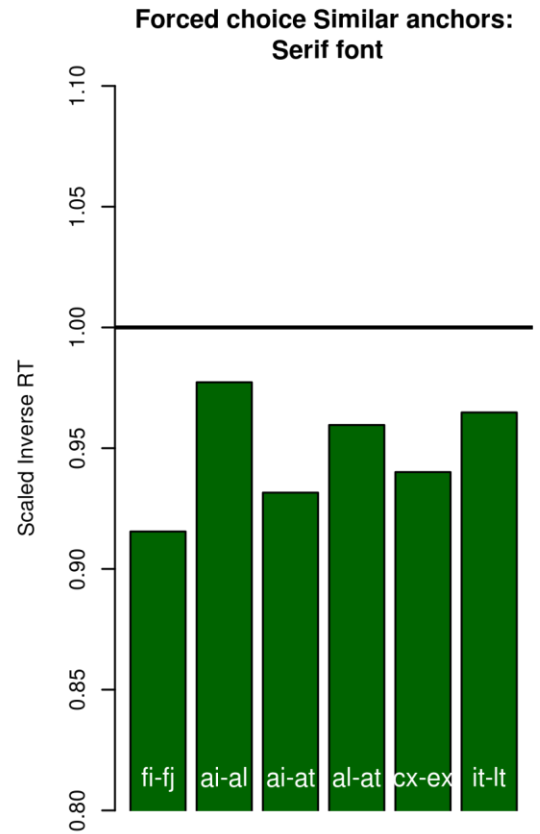




Correlation between error rate and inverse RT: -0.6925

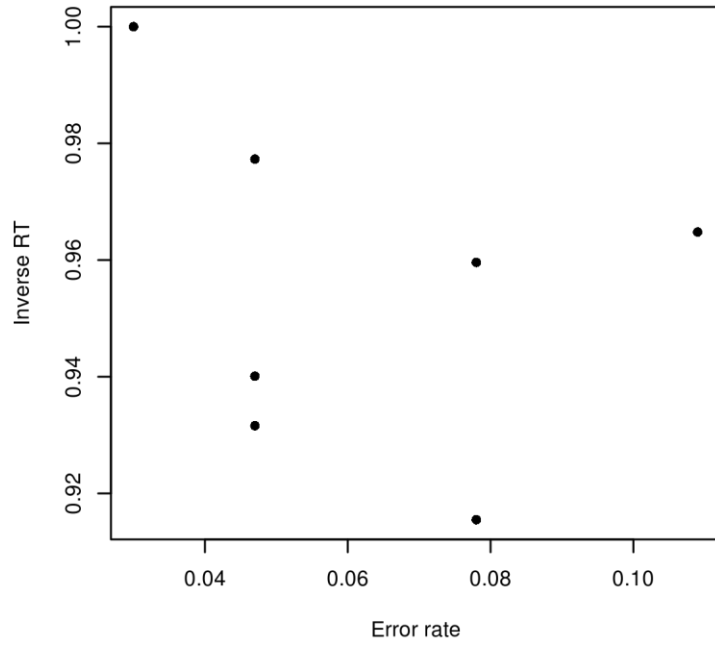


Candidate: EU in Greek. (epsilon upsilon)

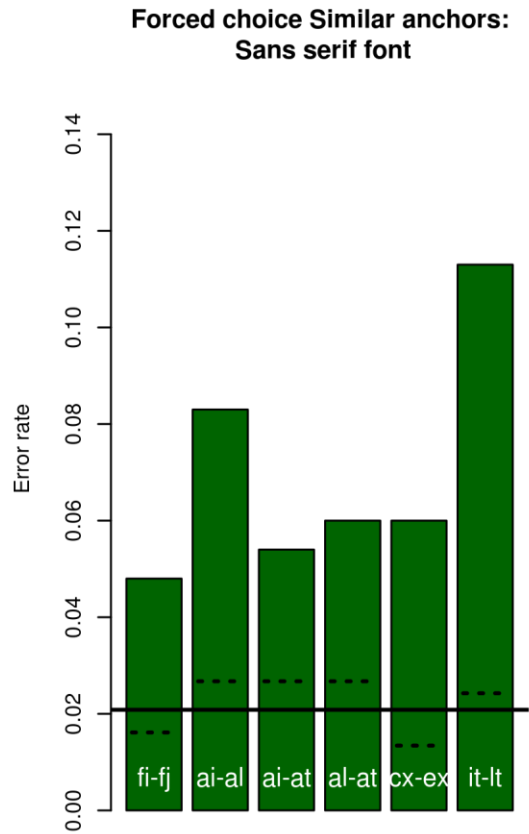


Candidate: EU in Greek. (epsilon upsilon)

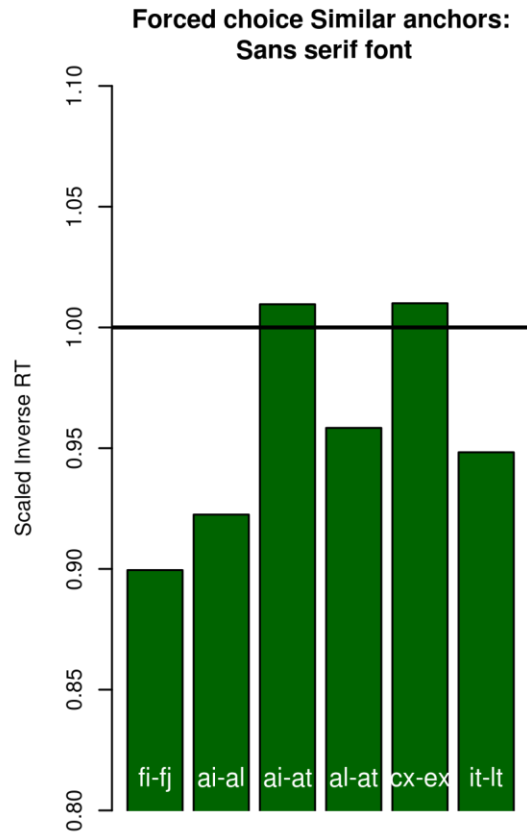
Option	Error rate	Between error rate	Inverse RT	Log-odds delta accuracy
fi-fj	0.078	0.025	0.9155	-1.192
ai-al	0.047	0.033	0.9773	-0.352
ai-at	0.047	0.033	0.9316	-0.352
al-at	0.078	0.033	0.9596	-0.352
cx-ex	0.047	0.023	0.9401	-0.721
it-lt	0.109	0.055	0.9648	-0.738
Between	0.03	0.03	1	0



Correlation between error rate and inverse RT: -0.2772

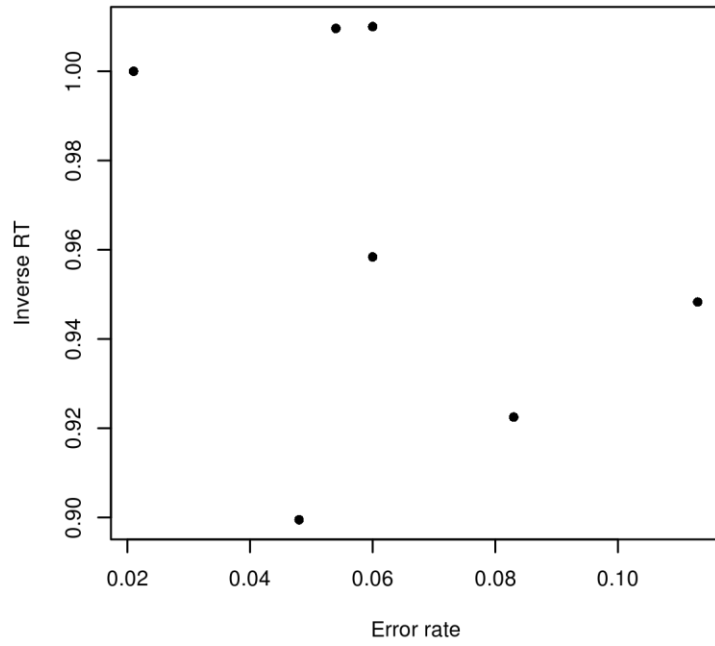


Candidate: EL in Greek. (epsilon lambda)

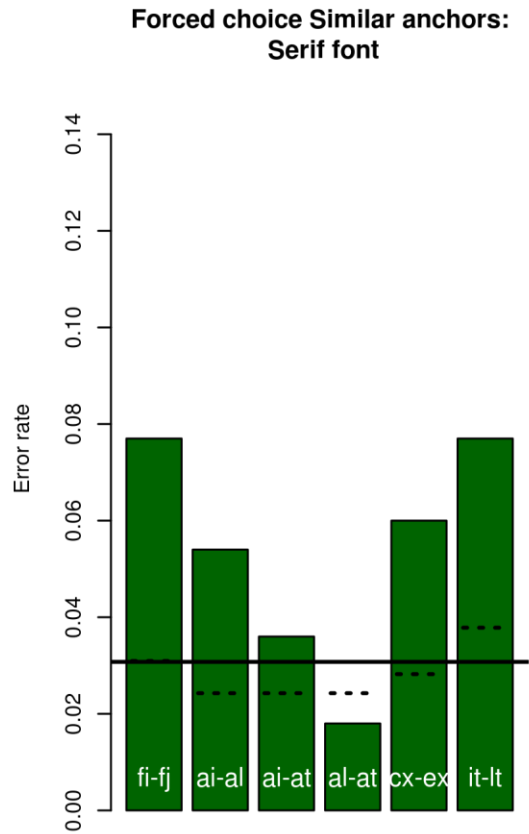


Candidate: EL in Greek. (epsilon lambda)

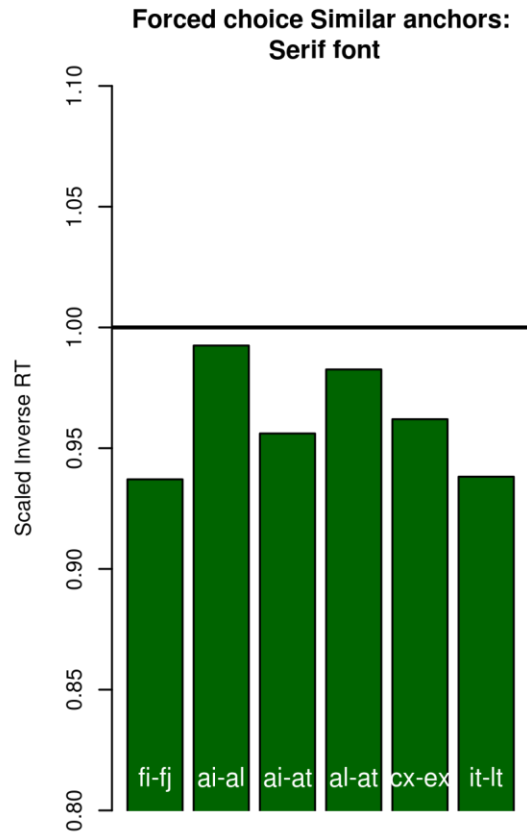
Option	Error rate	Between error rate	Inverse RT	Log-odds delta accuracy
fi-fj	0.048	0.016	0.8995	-1.114
ai-al	0.083	0.027	0.9225	-1.197
ai-at	0.054	0.027	1.0096	-0.723
al-at	0.06	0.027	0.9584	-1.197
cx-ex	0.06	0.013	1.01	-1.537
it-lt	0.113	0.024	0.9483	-1.635
Between	0.021	0.021	1	0



Correlation between error rate and inverse RT: -0.353

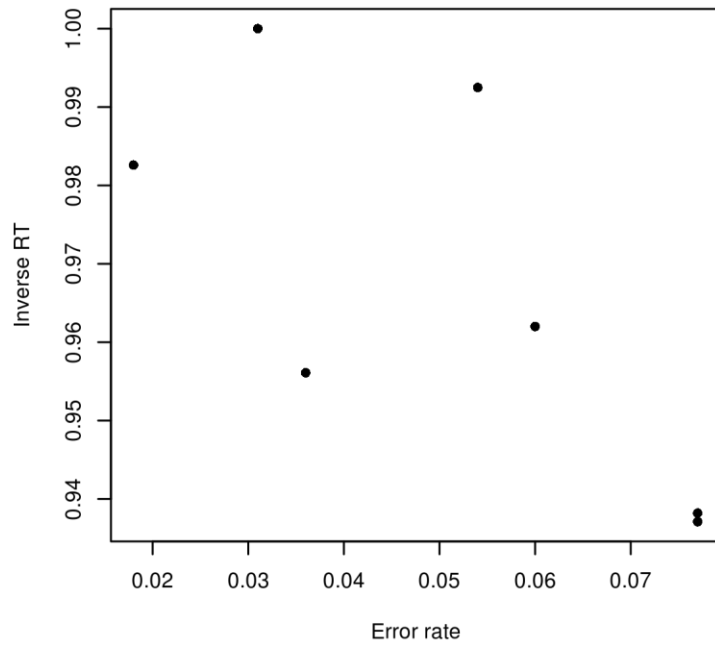


Candidate: EL in Greek. (epsilon lambda)



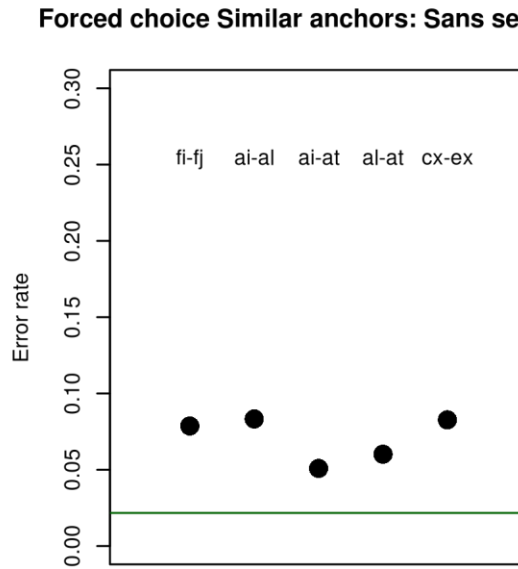
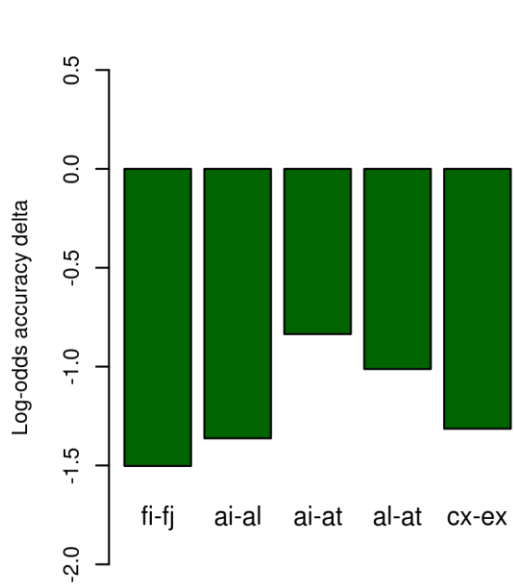
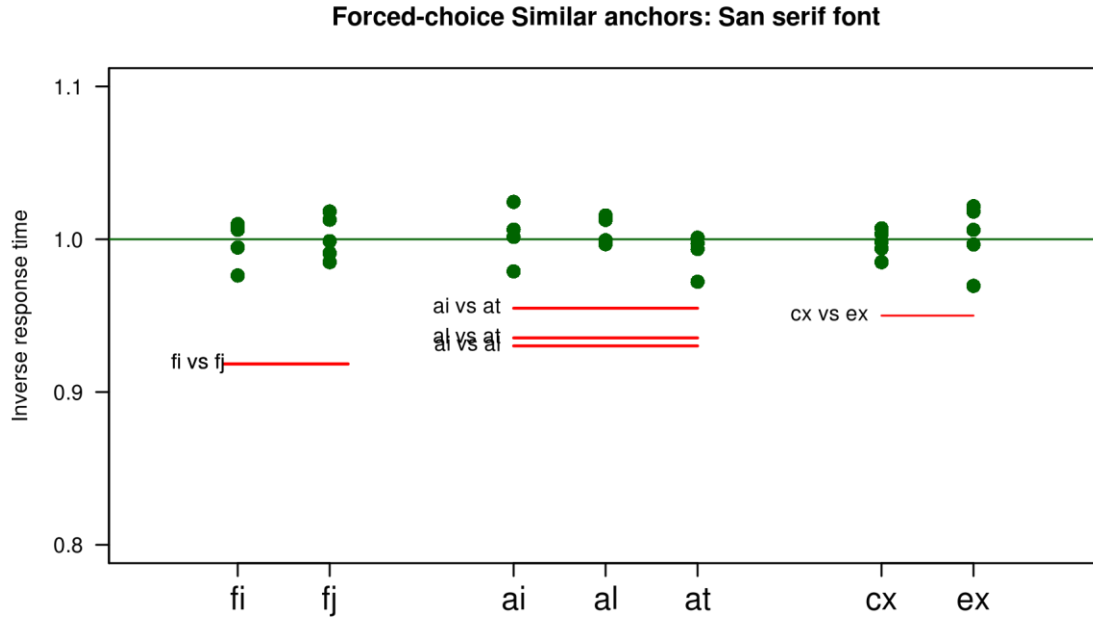
Candidate: EL in Greek. (epsilon lambda)

Option	Error rate	Between error rate	Inverse RT	Log-odds delta accuracy
fi-fj	0.077	0.031	0.9371	-0.966
ai-al	0.054	0.024	0.9925	-0.822
ai-at	0.036	0.024	0.9561	-0.398
al-at	0.018	0.024	0.9826	-0.822
cx-ex	0.06	0.028	0.962	-0.779
it-lt	0.077	0.038	0.9382	-0.757
Between	0.031	0.031	1	0



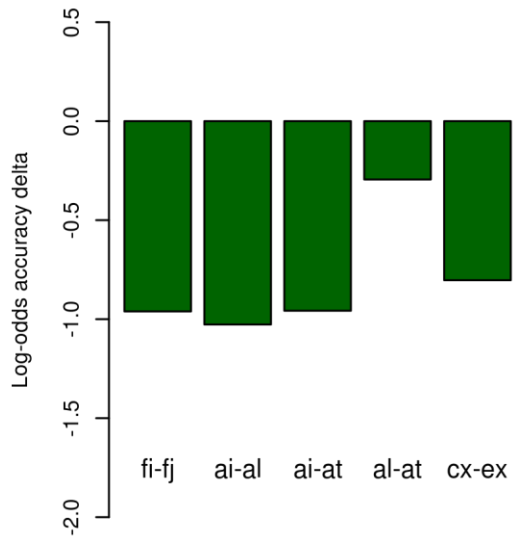
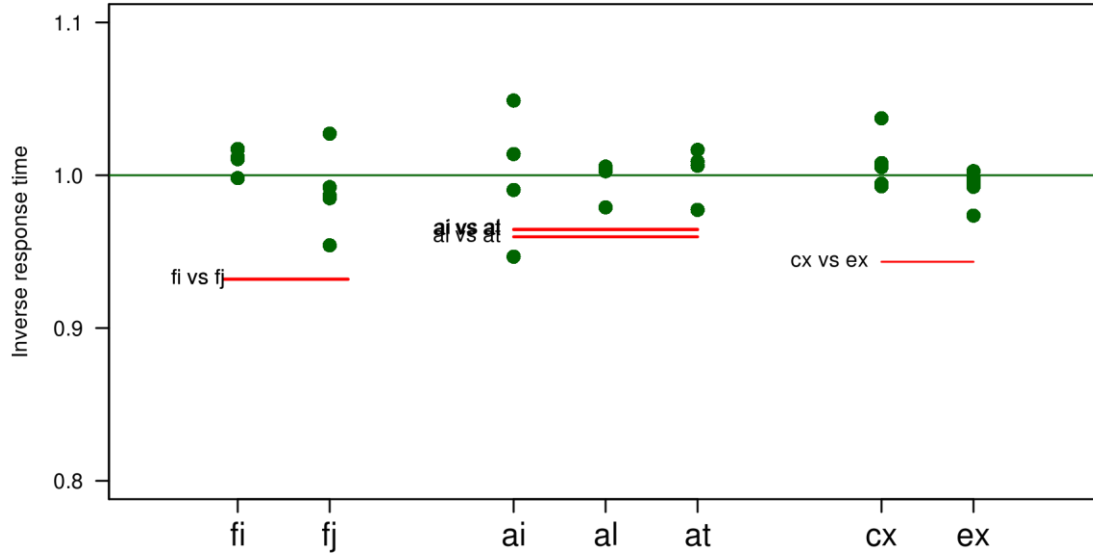
Correlation between error rate and inverse RT: -0.7193

The next figure shows comparisons of similar latin pairs. These serve as a comparison set, with the logic that any new pair evaluated to be less similar than these anchors is justifiably allowable.

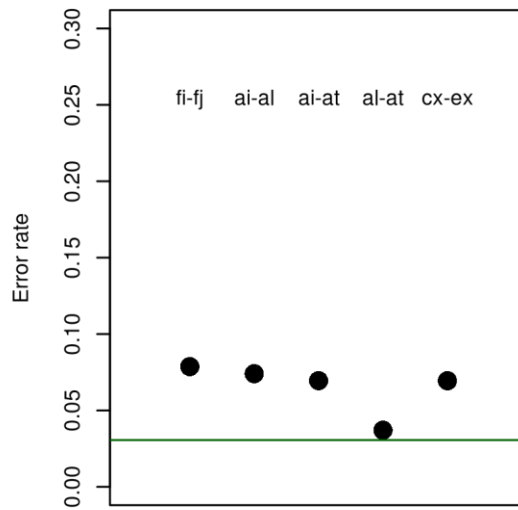




Forced-choice Similar anchors: Serif font



Forced choice Similar anchors: Serif font



**Inverse response time**

---

	<b>fi-fj</b>	<b>ai-al</b>	<b>ai-at</b>	<b>al-at</b>	<b>cx-ex</b>
Sans serif	0.918	0.93	0.955	0.935	0.95
Serif	0.932	0.965	0.964	0.96	0.943

---

**Log-odds difference in accuracy**

---

	<b>fi-fj</b>	<b>ai-al</b>	<b>ai-at</b>	<b>al-at</b>	<b>cx-ex</b>
Sans serif	-1.5025	-1.3627	-0.8355	-1.0124	-1.3141
Serif	-0.961	-1.027	-0.9575	-0.2946	-0.8034

---

**Error rate**

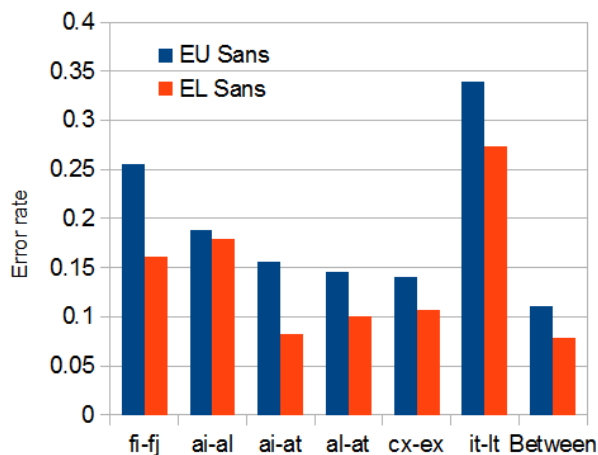
---

	<b>Between</b>	<b>fi-fj</b>	<b>ai-al</b>	<b>ai-at</b>	<b>al-at</b>	<b>cx-ex</b>
Sans serif	0.0217	0.0787	0.0833	0.0509	0.0602	0.0827
Serif	0.0306	0.0787	0.0741	0.0694	0.037	0.0694

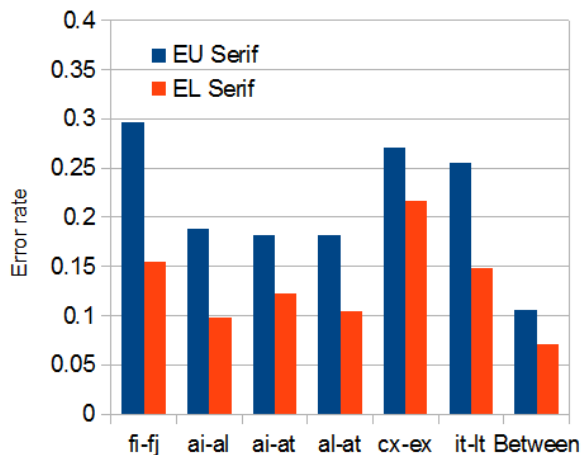
---

## Go/No-Go Task: Accuracy Metric

Go/No-go Anchors: Sans serif font



Go/No-go Anchors: Serif font



Option	EU Sans	EL Sans
fi-fj	0.255	0.161
ai-al	0.188	0.179
ai-at	0.156	0.083
al-at	0.146	0.101
cx-ex	0.141	0.107
it-lt	0.339	0.274
Between	0.111	0.079

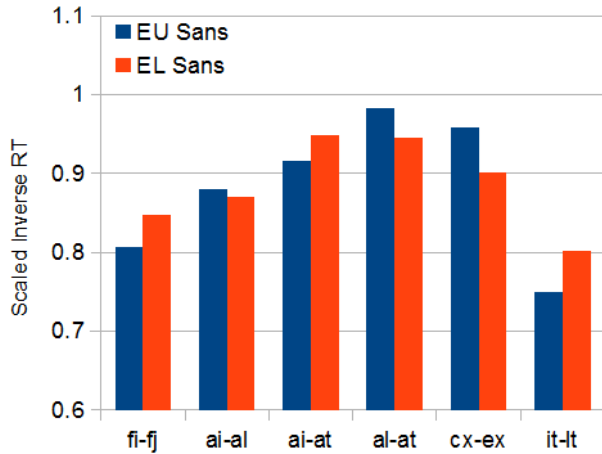
Option	EU Serif	EL Serif
fi-fj	0.297	0.155
ai-al	0.188	0.098
ai-at	0.182	0.122
al-at	0.182	0.104
cx-ex	0.271	0.217
it-lt	0.255	0.149
Between	0.106	0.071

- Test-retest reliability is .922 for Sans and .77 for serif.
- EL study produced overall lower error rates; possibly because these anchors were tested at the end of the study and
- Adjusting accuracy by subtracting error rate obtained for each pair changes these to (.91, .91), and by dividing to (.88, .98).
- Adjusting by dividing seems to make highest values most consistent across experiments, but this adjustment cannot be done reliably on an individual basis (because of error rates of 0, relatively small numbers of observations for the comparison cases, and wide binomial error variability)
- Correlations of adjusted to non-adjusted accuracy scores are all above .95, but it seems likely that the increase in reliability is mostly accidental and might not be replicated in future studies (and was did not occur for DMTS task).
- Worst-case is .339 for it-lt; Average of it-lt sans is .306, consistent with fi-fj serif of .297.
- **Recommendation: use error rate of 0.34 as a conservative criterion**

**Note: Error rate and Inverse RT were correlated  $\{-.937, -.979, -.965, -.89\}$ , suggesting that the overall decision should agree highly between these two measures and both may not be necessary.**

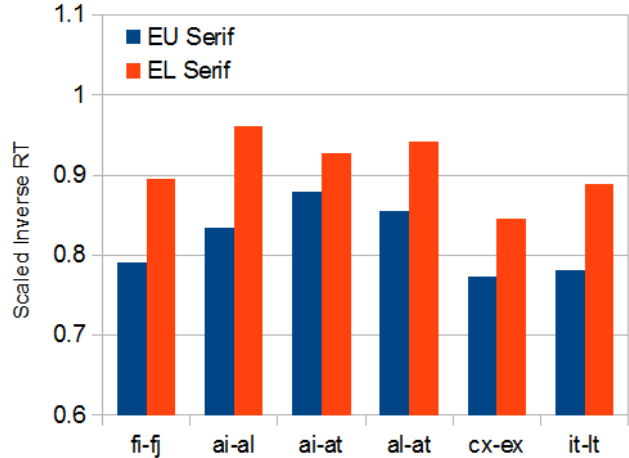
## Go/No-Go Task: Inverse RT Metric

Go/No-go anchors: Sans serif font



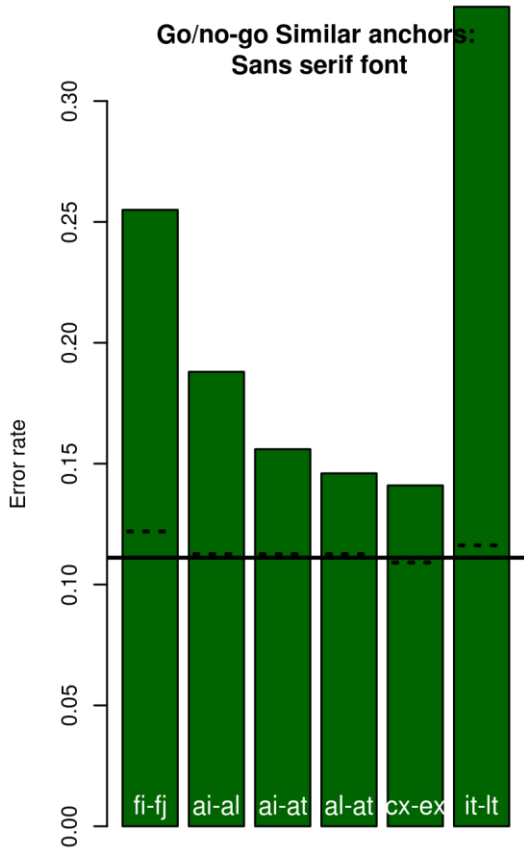
Option	EU Sans	EL Sans
fi-fj	0.8068	0.8472
ai-al	0.8798	0.8704
ai-at	0.9161	0.9486
al-at	0.983	0.9455
cx-ex	0.9585	0.9014
it-lt	0.7493	0.802

Go/No-go anchors: Serif font

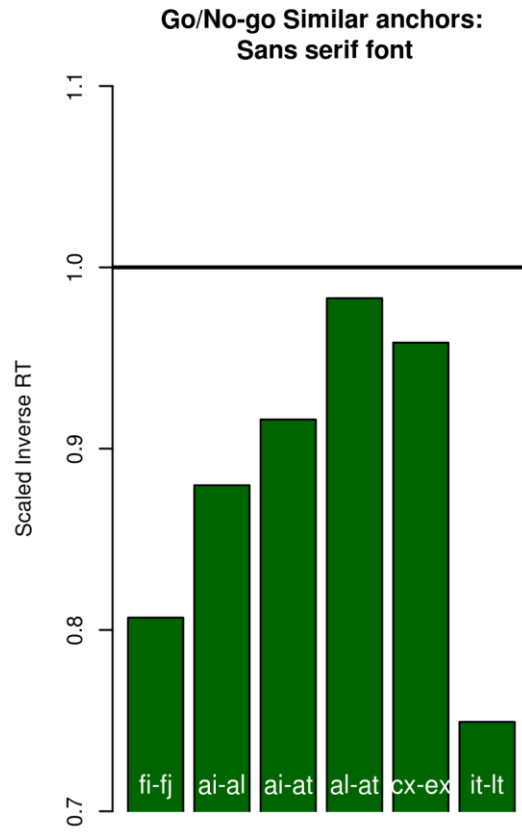


Option	EU Serif	EL Serif
fi-fj	0.7907	0.8953
ai-al	0.8344	0.9606
ai-at	0.8796	0.9281
al-at	0.8552	0.9414
cx-ex	0.7723	0.8454
it-lt	0.781	0.8886

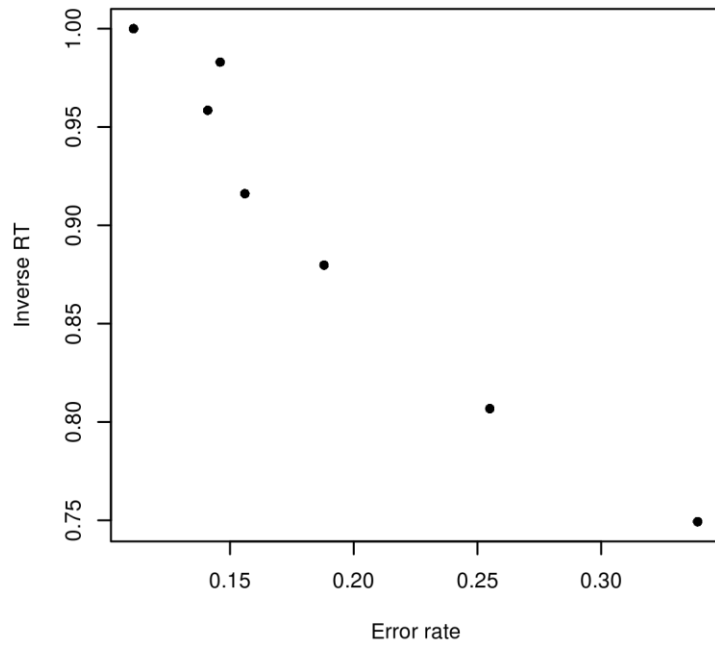
- Test-retest reliability was .906 for sans and .79 for serif. These values are already scaled, so that 1.0 is the average 'different' value.
- EL study produced higher values in the serif font. This is consistent with the overall higher accuracy, and is not a speed-accuracy tradeoff.
- Several cases in each font and each experiment produce scaled RT below 0.8; lowest is 0.77.
- **Recommendation: use 0.77 as criterion.**



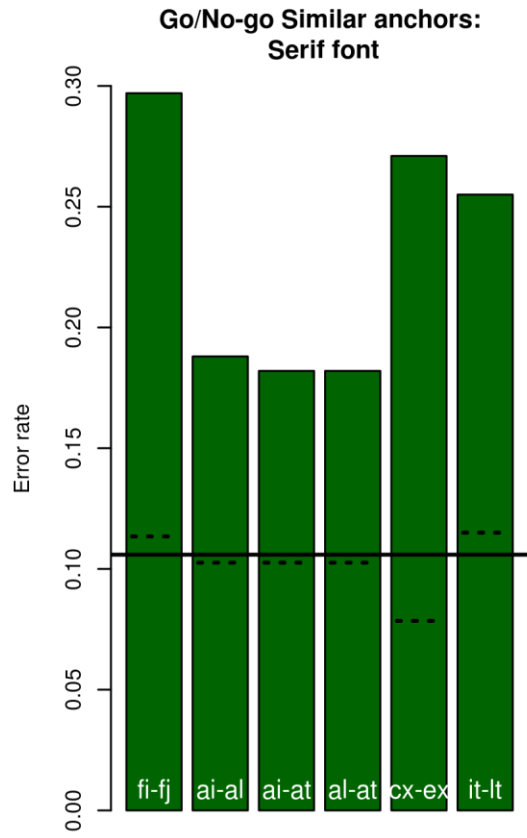
Candidate: EU in Greek. (epsilon upsilon)



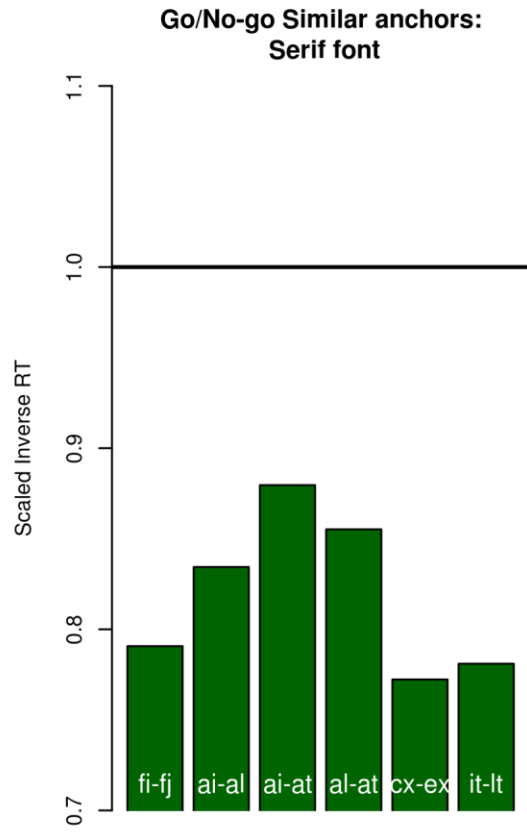
Candidate: EU in Greek. (epsilon upsilon)



Correlation between error rate and inverse RT: -0.9716



Candidate: EU in Greek. (epsilon upsilon)



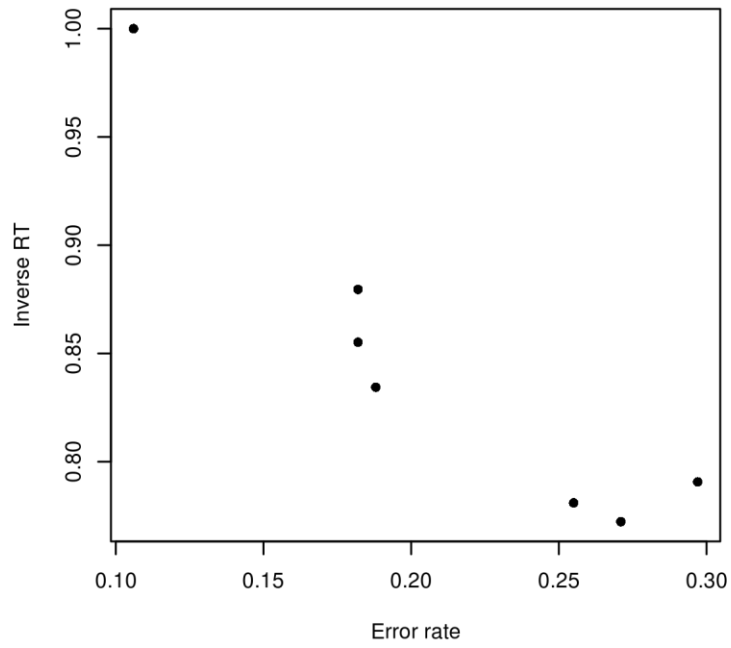
Candidate: EU in Greek. (epsilon upsilon)

|

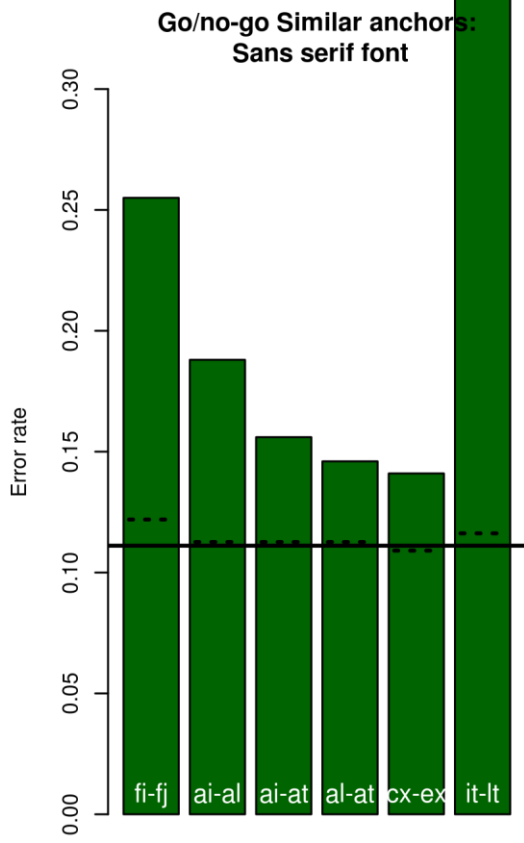
|

|

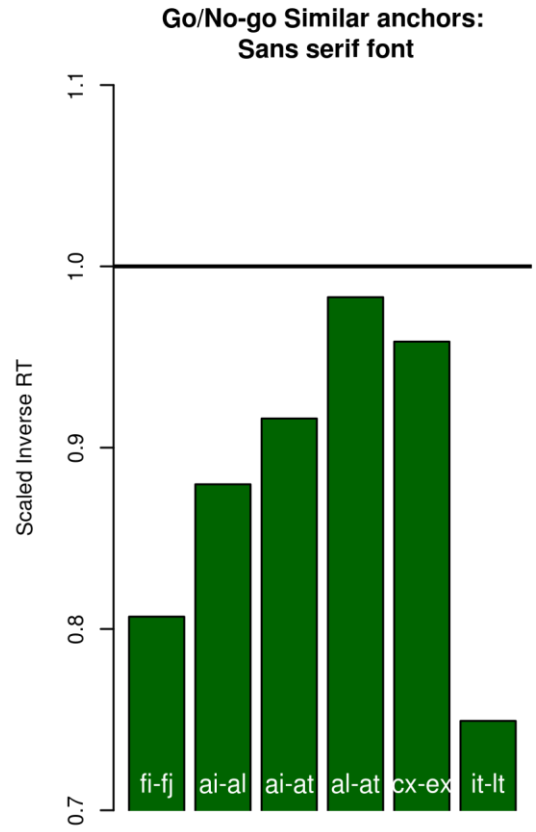




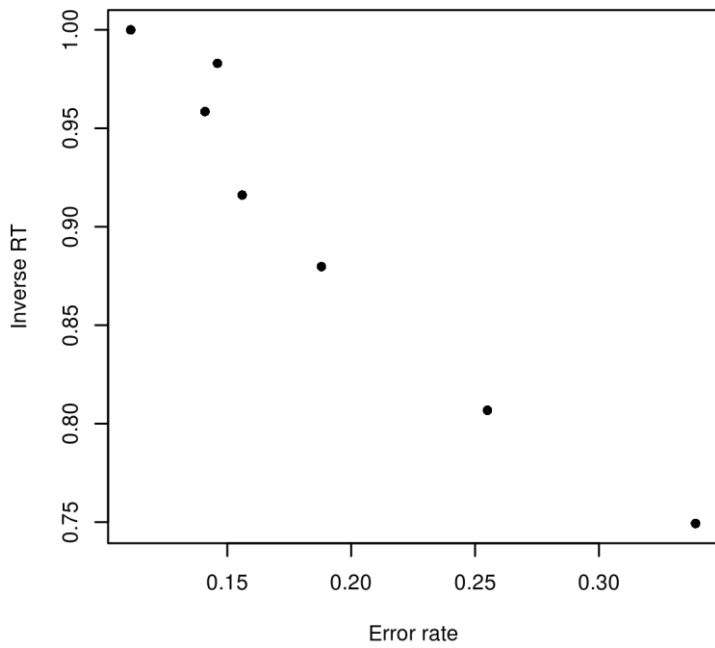
Correlation between error rate and inverse RT: -0.9281



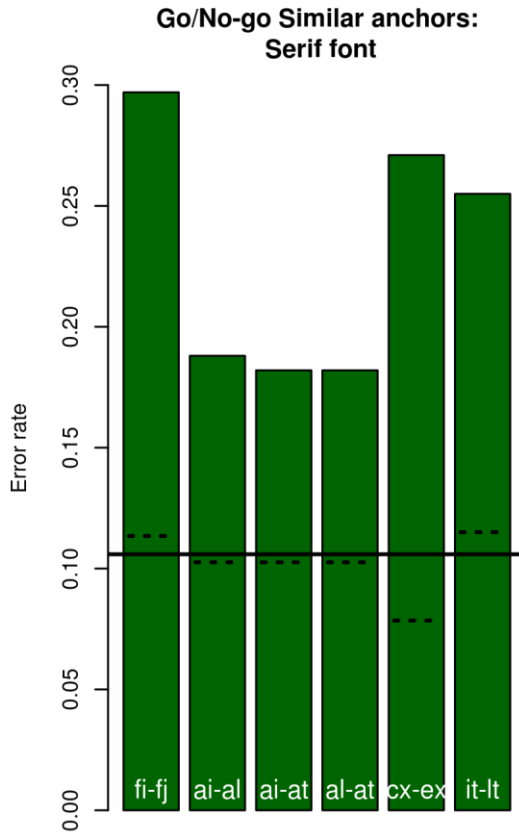
Candidate: EU in Greek. (epsilon upsilon)



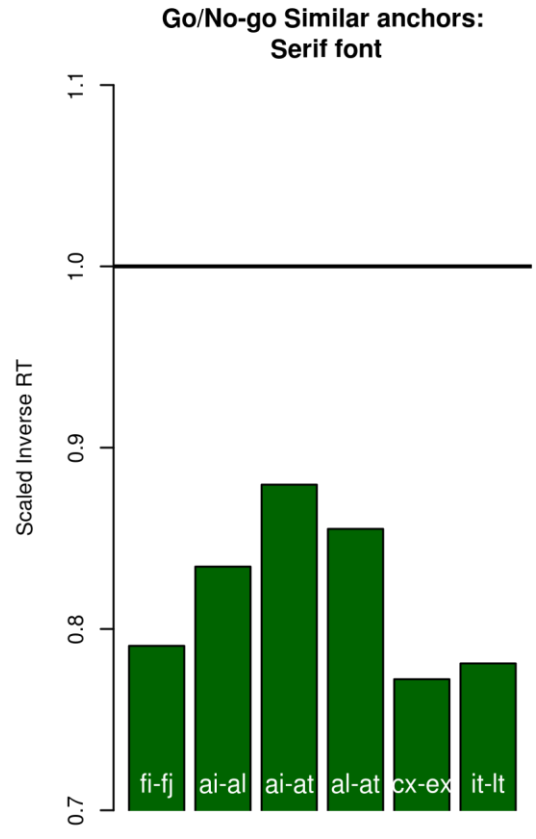
Candidate: EU in Greek. (epsilon upsilon)



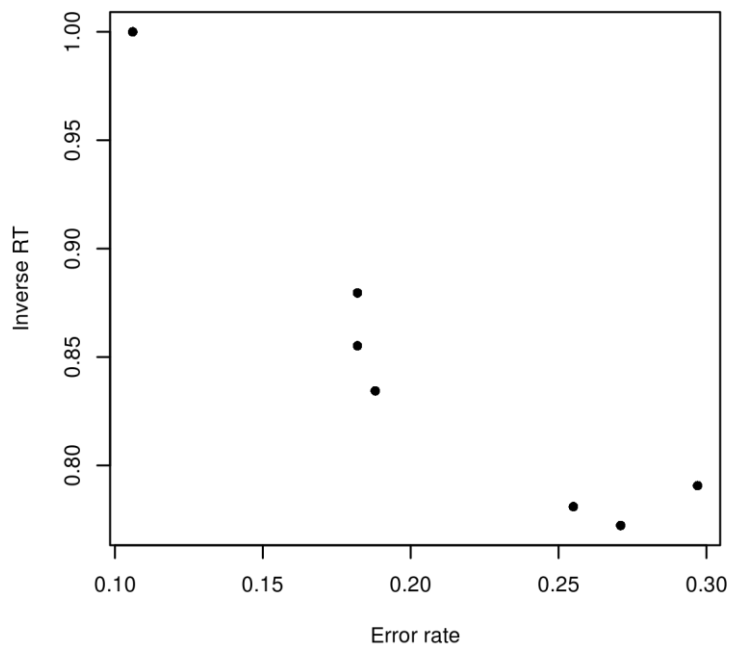
Correlation between error rate and inverse RT: -0.9716



Candidate: EU in Greek. (epsilon upsilon)



Candidate: EU in Greek. (epsilon upsilon)



Correlation between error rate and inverse RT: -0.9281