

Supplementary Materials:

Problems with the Big Five assessment in the World Values Survey

Contents

A: Information on variables	2
B: Data preprocessing	3
C: Figures included in main text	4
D: Descriptive statistics	8
E: Interitem correlations	9

A: Information on variables

The **o** items relate to openness, **c** items to conscientiousness, **e** items to extraversion, **a** items to agreeableness, and **s** items to emotional stability. The items with the suffix 2 are reverse coded.

Item	WVS variable	Description	Reverse coded
o1	V160J	... has an active imagination	No
o2	V160E	... has few artistic interests	Yes
c1	V160H	... does a thorough job	No
c2	V160C	... tends to be lazy	Yes
e1	V160F	... is outgoing, sociable	No
e2	V160A	... is reserved	Yes
a1	V160B	... is generally trusting	No
a2	V160G	... tends to find fault with others	Yes
s1	V160D	... is relaxed, handles stress well	No
s2	V160I	... gets nervous easily	Yes

B: Data preprocessing

```
# Load packages
library("rio")
library("ggplot2")
library("reshape2")
library("grid")
library("gridExtra")
library("tidyr")
library("stargazer")

# Load dataset (Stata format)
wvs <- import("WV6_Stata_v_2016_01_01.dta")

# Code missing values
trait.vars <- c("V160A", "V160B", "V160C", "V160D", "V160E",
               "V160F", "V160G", "V160H", "V160I", "V160J")
wvs[trait.vars][wvs[trait.vars] < 0] <- NA

# Reverse code and save variables
wvs$o1 <- wvs$V160J
wvs$o2 <- (wvs$V160E*-1)

wvs$c1 <- wvs$V160H
wvs$c2 <- (wvs$V160C*-1)

wvs$e1 <- wvs$V160F
wvs$e2 <- (wvs$V160A*-1)

wvs$a1 <- wvs$V160B
wvs$a2 <- (wvs$V160G*-1)

wvs$s1 <- wvs$V160D
wvs$s2 <- (wvs$V160I*-1)

wvs$male <- wvs$V240
wvs$male[wvs$male < 0] <- NA
wvs$male[wvs$male == 2] <- 0

wvs$age <- wvs$V242
wvs$age[wvs$age < 0] <- NA
```

C: Figures included in main text

```
# Make data frame with the BFI-10 items
b5 <- wvs[c("V2", "male", "age", "o1", "o2", "c1", "c2", "e1", "e2", "a1", "a2", "s1", "s2")]
b5 <- na.omit(b5)

b5.cor <- data.frame(country = unique(b5$V2),
                     n = NA,
                     cor.o = NA,
                     se.o = NA,
                     cor.c = NA,
                     se.c = NA,
                     cor.e = NA,
                     se.e = NA,
                     cor.a = NA,
                     se.a = NA,
                     cor.s = NA,
                     se.s = NA
                     )

for (i in unique(b5$V2)){
  b5.cor$n[b5.cor$country == i] <- NROW(b5[b5$V2 == i,])
  b5.cor$cor.o[b5.cor$country == i] <- cor.test(b5[b5$V2 == i,]$o1,
                                                b5[b5$V2 == i,]$o2)$estimate
  b5.cor$se.o[b5.cor$country == i] <- cor.test(b5[b5$V2 == i,]$o1,
                                                b5[b5$V2 == i,]$o2)$estimate /
    cor.test(b5[b5$V2 == i,]$o1, b5[b5$V2 == i,]$o2)$statistic
  b5.cor$cor.c[b5.cor$country == i] <- cor.test(b5[b5$V2 == i,]$c1,
                                                b5[b5$V2 == i,]$c2)$estimate
  b5.cor$se.c[b5.cor$country == i] <- cor.test(b5[b5$V2 == i,]$c1,
                                                b5[b5$V2 == i,]$c2)$estimate /
    cor.test(b5[b5$V2 == i,]$c1, b5[b5$V2 == i,]$c2)$statistic
  b5.cor$cor.e[b5.cor$country == i] <- cor.test(b5[b5$V2 == i,]$e1,
                                                b5[b5$V2 == i,]$e2)$estimate
  b5.cor$se.e[b5.cor$country == i] <- cor.test(b5[b5$V2 == i,]$e1,
                                                b5[b5$V2 == i,]$e2)$estimate /
    cor.test(b5[b5$V2 == i,]$e1, b5[b5$V2 == i,]$e2)$statistic
  b5.cor$cor.a[b5.cor$country == i] <- cor.test(b5[b5$V2 == i,]$a1,
                                                b5[b5$V2 == i,]$a2)$estimate
  b5.cor$se.a[b5.cor$country == i] <- cor.test(b5[b5$V2 == i,]$a1,
                                                b5[b5$V2 == i,]$a2)$estimate /
    cor.test(b5[b5$V2 == i,]$a1, b5[b5$V2 == i,]$a2)$statistic
  b5.cor$cor.s[b5.cor$country == i] <- cor.test(b5[b5$V2 == i,]$s1,
                                                b5[b5$V2 == i,]$s2)$estimate
  b5.cor$se.s[b5.cor$country == i] <- cor.test(b5[b5$V2 == i,]$s1,
                                                b5[b5$V2 == i,]$s2)$estimate /
    cor.test(b5[b5$V2 == i,]$s1, b5[b5$V2 == i,]$s2)$statistic
}
```

```

b5.cor$name <- NA
b5.cor[b5.cor$country == 12,]$name <- "Algeria"
b5.cor[b5.cor$country == 48,]$name <- "Bahrain"
b5.cor[b5.cor$country == 76,]$name <- "Brazil"
b5.cor[b5.cor$country == 156,]$name <- "China"
b5.cor[b5.cor$country == 400,]$name <- "Jordan"
b5.cor[b5.cor$country == 414,]$name <- "Kuwait"
b5.cor[b5.cor$country == 702,]$name <- "Singapore"
b5.cor[b5.cor$country == 170,]$name <- "Colombia"
b5.cor[b5.cor$country == 218,]$name <- "Ecuador"
b5.cor[b5.cor$country == 818,]$name <- "Egypt"
b5.cor[b5.cor$country == 268,]$name <- "Georgia"
b5.cor[b5.cor$country == 276,]$name <- "Germany"
b5.cor[b5.cor$country == 344,]$name <- "Hong Kong"
b5.cor[b5.cor$country == 356,]$name <- "India"
b5.cor[b5.cor$country == 368,]$name <- "Iraq"
b5.cor[b5.cor$country == 422,]$name <- "Lebanon"
b5.cor[b5.cor$country == 434,]$name <- "Libya"
b5.cor[b5.cor$country == 528,]$name <- "Netherlands"
b5.cor[b5.cor$country == 586,]$name <- "Pakistan"
b5.cor[b5.cor$country == 275,]$name <- "Palestine"
b5.cor[b5.cor$country == 646,]$name <- "Rwanda"
b5.cor[b5.cor$country == 710,]$name <- "South Africa"
b5.cor[b5.cor$country == 764,]$name <- "Thailand"
b5.cor[b5.cor$country == 788,]$name <- "Tunisia"
b5.cor[b5.cor$country == 887,]$name <- "Yemen"

fig.ii.o <- ggplot(b5.cor, aes(x = name, y=cor.o, ymin=cor.o-1.96*se.o,
                               ymax=cor.o+1.96*se.o)) +
  geom_hline(yintercept = 0, size=0.5, linetype="dashed", colour="#999999") +
  geom_pointrange() +
  coord_flip() +
  ylab("r (Imagination, Few artistic interests)") +
  theme_minimal() +
  scale_y_continuous(breaks=c(-.5,0,.3), labels=c("-.5","0",".3")) +
  xlab("")

fig.ii.c <- ggplot(b5.cor, aes(x = name, y=cor.c, ymin=cor.c-1.96*se.c,
                               ymax=cor.c+1.96*se.c)) +
  geom_hline(yintercept = 0, size=0.5, linetype="dashed", colour="#999999") +
  geom_pointrange() +
  coord_flip() +
  ylab("r (Not lazy, Thorough job)") +
  theme_minimal() +
  scale_y_continuous(breaks=c(-.5,0,.5), labels=c("-.5","0",".5")) +
  xlab("") +
  theme(axis.text.y = element_blank())

fig.ii.e <- ggplot(b5.cor, aes(x = name, y=cor.e, ymin=cor.e-1.96*se.e,
                               ymax=cor.e+1.96*se.e)) +
  geom_hline(yintercept = 0, size=0.5, linetype="dashed", colour="#999999") +

```

```

geom_pointrange() +
coord_flip() +
ylab("r (Not reserved, Outgoing)") +
theme_minimal() +
scale_y_continuous(breaks=c(-.5,0,.5), labels=c("-.5","0",".5")) +
xlab("") +
theme(axis.text.y = element_blank())

fig.ii.a <- ggplot(b5.cor, aes(x = name, y=cor.a, ymin=cor.a-1.96*se.a,
                             ymax=cor.a+1.96*se.a)) +
  geom_hline(yintercept = 0, size=0.5, linetype="dashed", colour="#999999") +
  geom_pointrange() +
  coord_flip() +
  ylab("r (Trusting, Does not find faults)") +
  theme_minimal() +
  xlab("")

fig.ii.s <- ggplot(b5.cor, aes(x = name, y=cor.s, ymin=cor.s-1.96*se.s,
                             ymax=cor.s+1.96*se.s)) +
  geom_hline(yintercept = 0, size=0.5, linetype="dashed", colour="#999999") +
  geom_pointrange() +
  coord_flip() +
  ylab("r (Relaxed, not nervous)") +
  scale_y_continuous(breaks=c(-.5,0,.5), labels=c("-.5","0",".5")) +
  theme_minimal() +
  xlab("") +
  theme(axis.text.y = element_blank())

png('figure1.png', height=8, width=8, units="in",res=700)
grid.arrange(fig.ii.o, fig.ii.c, fig.ii.e, fig.ii.a, fig.ii.s,
              widths=c(5, 4, 4), ncol=3)
dev.off()

## pdf
## 2

b5.long <- gather(b5.cor, trait, value, c(cor.o,cor.c,cor.e,cor.a,cor.s),
                  factor_key=TRUE)

png('figure2.png', height=6, width=8, units="in",res=700)
ggplot(b5.long, aes(x=value, fill=trait)) +
  geom_vline(xintercept=0, linetype="dashed") +
  geom_vline(xintercept=-0.3, colour="gray", linetype="dashed") +
  geom_vline(xintercept=0.3, colour="gray", linetype="dashed") +
  scale_y_continuous(breaks=c(0,.25,.50,.75,1), labels=c("", "", "", "", "")) +
  scale_x_continuous(breaks=c(-.5,0,.5), labels=c("-.5","0",".5")) +
  geom_dotplot(stackgroups = TRUE, stackratio = 1.2, binwidth=0.07,
               dotsize = 0.7, binpositions = "all") +
  scale_fill_manual("", labels = c("Openness","Conscientiousness","Extraversion",
                                   "Agreeableness","Emotional Stability",
                                   "Neuroticism","Psychoticism"),
                    values = c("#69D2E7", "#81AD99", "#C02942",
                               "#F38630", "#ECD078")) +

```

```
  xlab("Item-item correlation") +  
  ylab("") +  
  annotate("text", x = -0.8, y = 0.27, label = "Bahrain") +  
  theme_minimal()  
dev.off()
```

```
## pdf  
## 2
```

D: Descriptive statistics

```
# Get country with minimum number of observations
min(b5.cor$n)
```

```
[1] 653
```

```
b5.cor$name[b5.cor$n == min(b5.cor$n)]
```

```
[1] "Yemen"
```

```
# Get country with maximum number of observations
max(b5.cor$n)
```

```
[1] 3317
```

```
b5.cor$name[b5.cor$n == max(b5.cor$n)]
```

```
[1] "South Africa"
```

```
# Create summary statistics table
stargazer(b5[c("male", "age", "o1", "o2", "c1", "c2",
               "e1", "e2", "a1", "a2", "s1", "s2")],
           title = "Summary statistics",
           covariate.labels = c("Male", "Age"),
           summary = TRUE)
```

% Table created by stargazer v.5.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu % Date and time: Tir, Feb 28, 2017 - 13:14:38

Table 2: Summary statistics

Statistic	N	Mean	St. Dev.	Min	Max
Male	32,880	0.494	0.500	0	1
Age	32,880	40.540	15.643	16	99
o1	32,880	3.242	1.239	1	5
o2	32,880	3.027	1.273	1	5
c1	32,880	3.617	1.248	1	5
c2	32,880	3.528	1.297	1	5
e1	32,880	3.536	1.224	1	5
e2	32,880	2.734	1.322	1	5
a1	32,880	3.374	1.295	1	5
a2	32,880	3.203	1.297	1	5
s1	32,880	3.326	1.194	1	5
s2	32,880	2.997	1.287	1	5

E: Interitem correlations

```
get_upper_tri <- function(cormat){
  cormat[lower.tri(cormat)]<- NA
  return(cormat)
}

for(i in unique(b5$V2)) {
  cormat <- round(cor(b5[b5$V2 == i,
                        c("o1","o2","c1","c2","e1","e2","a1","a2","s1","s2")]),
                  2)
  upper_tri <- get_upper_tri(cormat)

  melted_cormat <- melt(upper_tri, na.rm = TRUE)

  p <- ggplot(data = melted_cormat, aes(Var2, Var1, fill = value))+
    geom_tile(color = "white")+
    scale_fill_gradient2(low = "blue", high = "red", mid = "white",
                        midpoint = 0, limit = c(-1,1), space = "Lab",
                        name="Pearson\nCorrelation") +

    theme_minimal() +
    theme(axis.text.x = element_text(angle = 45, vjust = 1,
                                      size = 12, hjust = 1))+

    coord_fixed() +
    ggtitle(b5.cor[b5.cor$country == i,]$name) +
    geom_text(aes(Var2, Var1, label = value), color = "black", size = 2) +
    theme(
      axis.title.x = element_blank(),
      axis.title.y = element_blank(),
      panel.grid.major = element_blank(),
      panel.border = element_blank(),
      panel.background = element_blank(),
      axis.ticks = element_blank(),
      plot.title = element_text(size = 12),
      legend.justification = c(1, 0),
      legend.position = c(0.6, 0.7),
      legend.direction = "horizontal")+
    guides(fill = guide_colorbar(barwidth = 7, barheight = 1,
                                title.position = "top", title.hjust = 0.5))

  print(p)
}
```













