



# How Reliable Are Personality Judgments by Political Experts? The Curious Case of Donald Trump

Jeroen Joly<sup>1</sup> , Joeri Hofmans<sup>2</sup> 

[1] *Centre de Recherche en Science Politique (CReSPo), Université Saint-Louis Brussels, Brussels, Belgium.* [2] *Work and Organizational Psychology, Vrije Universiteit Brussel, Brussels, Belgium.*

---

Personality Science, 2023, Vol. 4, Article e6715, <https://doi.org/10.5964/ps.6715>

**Received:** 2021-05-13 • **Accepted:** 2022-10-10 • **Published (VoR):** 2023-03-08

**Handling Editor:** Ryne Sherman, Hogan Assessment Systems, Tulsa, OK, USA

**Reviewing:** Round 1 - Bert N. Bakker. Open reviews are available. [see [Index of Supplementary Materials](#)]

**Corresponding Author:** Jeroen Joly, Rue du Marais, 119, Brussels, Belgium. E-mail: [Jeroen.joly@usaintlouis.be](mailto:Jeroen.joly@usaintlouis.be)

**Supplementary Materials:** Data [see [Index of Supplementary Materials](#)]



## Abstract

Recent studies have highlighted the importance of personality in electoral politics. With the rise of populist and atypical personalities across stable and established democracies, pundits, journalists and other political experts often rely on their assessments of politicians' personalities to explain their behavior. Additionally, numerous citizens depend on their expertise and assessments to form their own opinion. Given that most political experts have never personally met these politicians, how reliable are their assessments of high-profile politicians' personality? We address this question by analyzing inter-rater reliability of ratings of US President Trumps' personality by seven Belgian political experts. Using the NEO-FFI, our analyses indicate low inter-rater agreement on most of the Big Five personality traits and the facets of Trumps' personality. Therefore, the excessive use of analyses based on third party assessments and interpretations of politicians' personality should be regarded with caution given their potential impact on the wider public.

## Keywords

inter-rater reliability, Big Five Personality, expert ratings, journalism, Krippendorff alpha



### Relevance Statement

Expert analyses often rely, in part, on judgements of character of the politicians involved. Increasingly, academics also rely on third-party raters in the study of politicians' personality. Yet, it is crucial to understand how reliable these are.

### Key Insights

- To what extent do political experts converge on their personality judgments of Donald Trump?
- Substantial disagreement exists between experts for most personality traits and facets.
- Such rater idiosyncrasies are relevant because of the impact of experts on citizens' voting behavior.

Individual characteristics, like personality, play an important role in politics. This is particularly true in electoral politics, where certain personality traits have been shown to relate to popularity or one's position within politics (Joly et al., 2019; Nai, 2019; Scott & Medeiros, 2020). Thanks to traditional media and, more recently, also social media, ordinary citizens have access to a lot of source material that helps them make inferences about politicians' personality. These inferences are highly relevant for voting behavior, with Caprara (2007) arguing that voters' perceptions of politicians are as important in explaining political preferences as traditional socio-demographic characteristics. Media hereby play an essential role in "crafting the images and shaping the messages that are most desirable to voters." (pp. 151–152)

Journalists, political pundits and other types of political experts are continuously called upon in news media to provide explanations on ongoing events. These explanations often involve judgements regarding the decision-makers directly involved. To explain what motivates or drives the specific decision-makers in this process, political experts of all kinds often make inferences based on previous patterns of behavior. In other words, what do we know about a decisionmaker's goals, values and behavior from the past that can explain their current behavior? In short, political experts—like most people—draw a mental map of politicians' personality to explain and predict current and future decisions.

This process is extremely important, as these political experts inform a broader audience on who those specific decision-makers really are and what motivates their behavior, feelings and thinking. Moreover, citizens rely on this information—which often includes normative assessments and other judgements—to make up their own mind. How political experts view a decision-maker's personality, thus, greatly influences their further political assessments. The representations ordinary citizens have of their leaders are often shaped, strengthened or confirmed by so-called experts who color and frame leaders' specific cognitions, behaviors and decisions. The same unconventional behavior of a president, for example, can be positively framed as the 'non-politically correct'

behavior of a maverick, or negatively as being inappropriate and non-presidential. Political experts, therefore, have the power to influence citizens' assessment of their leaders—especially those non-partisan, independent and undecided citizens.

Against this backdrop, the question this study seeks to answer is the extent to which political experts differ in their inferences about the personality of politicians. In case those different experts are highly consistent, selective exposure to one or a few experts has little consequences. Regardless of the experts one is exposed to, one would be informed in a similar way. If such inferences are highly idiosyncratic, however, selective exposure to one or a few experts might make an important difference for voters in the sense that they are then potentially presented with (very) different information on the same politicians, information on which voters might rely when voting.

## Personality and (Electoral) Politics

The direct study of politicians' personality has a relatively recent academic history. During the first decade of this century, a group of researchers led by Barbaranelli, Caprara, Vecchione and Zimbardo set out to examine the link between personality, ideology and partisanship. They studied this link across several countries, among both ordinary citizens (Vecchione et al., 2011) and politicians (Caprara et al., 2003, 2010; see also Dietrich et al., 2012; Nørgaard & Klemmensen, 2019; Schumacher & Zettler, 2019), revealing that certain personality traits, like openness to experience, were associated with left-wing voters and politicians, while higher levels of conscientiousness were found among right-wing voters and politicians.

There has very recently been a renewed interest into the role of personality in politics, and especially into the relationship between personality traits and political success. One strand of research examined the impact of personality on political success by surveying the personality of politicians directly. Looking at Belgian politicians, Joly and colleagues (2019) found lower levels of agreeableness to be associated with different measures of political success reflecting different stages of the political career, including electoral success, parliamentary longevity and access to elite political positions. Moreover, research has shown that personality plays a role at even earlier stages of the political selection process, with Scott and Medeiros (2020) showing that the personality of municipal candidates in Canada differ from those of the general population and that winners differ from losers in terms of personality.

A second approach consists of relying on third-party raters—most often academic experts—to assess political leaders' personality<sup>1</sup>. This approach is particularly popular when evaluating the personality of leaders who are unapproachable or dead (Nai &

---

1) For the sake of completeness, it is worth mentioning that yet another approach to assessing elite personality is Machine Learning. For example, Ramey, Klingler, and Hollibaugh (2019) have measured the personality traits of Congressmembers by applying Machine Learning on their floor speeches.

Maier 2021). Using this approach, several researchers have relied on historians and biographers (Rubenzer et al., 2000) or personality psychologists (Visser et al., 2017) to evaluate the personalities of American presidents, determining which personality traits distinguish successful from unsuccessful presidents (Lilienfeld et al., 2012; Watts et al., 2013). More recently, Nai (2019) collected personality ratings of a large number of political leaders made by numerous scholars, showing that candidates with higher levels of conscientiousness, openness to experience and psychopathy score better at the ballot box.

Whereas studying politicians' personality lies at the very core of both approaches, it is important to realize that they do not study the exact same thing. Whereas self-ratings can be conceived of as a mixture of one's "true" personality and one's self-perceptions, other-ratings are comprised of the subject's "true" personality and interpersonal perceptions (McAbee & Connelly, 2016). In other words, whereas self-ratings are colored by the information that is only accessible to the self and how people perceive themselves, other-ratings are colored by how external observers see the other person, which is influenced, among other things, by stereotypes, impression management by the target, as well as the observer's own personality (McAbee & Connelly, 2016). Important here is that, although self- and other-ratings might diverge, both are relevant. Self-ratings tap into how people think about themselves and are therefore an important determinant of one's thinking, feeling and behavior. The value of observer ratings, instead, lies in the way someone's personality is perceived by observers, and because in the particular case of politicians those observers are political experts that communicate to the outside world, their assessments also impact the assessments of the larger public. This is highly relevant, as research shows that a lot of what we know about others does not stem from a direct observation of the target, but from intermediaries, which may communicate both accurate and inaccurate perceptions of the target (Craik, 2008).

Provided the importance of personality judgments in electoral politics (Joly et al., 2019; Scott & Medeiros, 2020) and given the understanding that political experts play an important role in influencing such personality judgments, a crucial question pertains to the extent to which idiosyncrasies of these political experts trickle down in their personality assessments of politicians. Research on this topic shows that, when people assess the personality of politicians, their ratings tend to be biased by their own political orientation (Nai & Maier, 2021; Rice et al., 2020; Wright & Tomlinson, 2018). Whereas this bias has been shown repeatedly, there is still quite some disagreement about the extent to which political experts are susceptible to it. While some studies found that political orientation biases not only the ratings of the general public but also those of experts (Wright & Tomlinson, 2018), others showed that expert ratings are less driven by their ideological preferences than ratings by laypeople (Nai & Maier, 2021), highlighting the need for further research on this topic. Apart from these mixed findings, previous research has studied the issue by comparing the average ratings by a group

of experts with the average ratings by a group of laypeople (Nai & Maier, 2021; Wright & Tomlinson, 2018). Although this approach allows testing whether experts *on average* provide different ratings than laypeople do, it fails to tap into the level of (in)consistency among different experts.

(In)consistencies among experts might potentially be very impactful, since research has shown that people tend to select information sources that are consistent with their attitudes, beliefs and social identity (Knobloch-Westerwick & Meng, 2009), which affects and limits their exposure to a specific subset of political experts. For example, Iyengar and Hahn (2009) have shown that conservatives and republicans prefer news reports from Fox News, while Democrats and liberals prefer CNN and NPR. Consequently, because people are exposed to a select group of experts only, idiosyncratic differences between experts in the way they explain someone else's thinking and behavior are critical and potentially very impactful. In the present study, we set out to study such differences, examining the extent to which different political experts converge on their personality judgments of politicians.

To this end, we conducted a case study of probably the most mediatized president, and perhaps the most mediatized politician ever at any given time: Donald Trump. This is a highly interesting case because there is a richness of spontaneous, unscripted, observational material, from press briefings to interviews, on which experts can and have based their personality assessments. Moreover, Donald Trump often goes off script, providing us with an insight into his natural behavior, opinion and attitudes. Finally, Donald Trump has a long and rich history of media appearances, from long before being a politician, in which he has displayed a rather consistent image and personality for external raters to observe.

## Method

### Procedure and Participants

To study rater idiosyncrasies in the evaluation of Trump's personality, we tested the extent to which different experts (dis)agree in their personality ratings of Donald Trump. To this end, we asked 11 Belgian journalists and academics, all experts on American politics with a marked interest in the American Presidency, to rate the personality of US President Donald Trump. These experts (including both Flemish and Francophones) regularly comment on ongoing events and all share their analyses with wider audiences—either through journalistic contributions or op-eds. The 11 experts were contacted in early April of 2018 and received a reminder one week after initial contact; the survey was closed by the end of April. Participating experts did not receive any monetary or other type of reward. Of the 11 contacted political experts, seven participated in the study. While two experts declined because of their busy schedule, another one declined because

they did not feel competent to rate someone's personality as a non-psychologist.<sup>2</sup> Final, one expert never responded.

While the disadvantage of relying on Belgian experts is that they are possibly less exposed to source material, it has the obvious advantage that they are not partisan or personally involved in the sense that they have never voted or worked for either/any American party. While this, of course, does not mean these experts do not have their own ideological preferences, it is safe to assume that Western experts—in this case Belgian experts—are less polarized than American ones. This is also suggested by [Nai and Maier \(2021\)](#), who found Dutch experts to draw a more consistent profile of Trump, regardless of their own preferences.

Although the general aim of our study (i.e., studying the extent of (dis)agreement about Trump's personality) is similar to that of [Nai and Maier \(2021\)](#), there are some notable differences. First, paralleling most research on personality in politics, [Nai and Maier \(2021\)](#) made use of short personality measures, which in their case was the Ten-Item Personality inventory (TiPi; [Gosling et al., 2003](#)). While the use of short forms is beneficial from a practical point of view, it comes with important limitations. First, the conciseness of those measures comes at the cost of reduced richness in the sense that short forms only measure broad traits and not their underlying facets. The consequence is that a lot of nuance is lost. Second, short forms are renowned for their poor factor structures and low internal consistency reliability ([Gosling et al., 2003](#)). Whereas this is not a problem per se, the fact that minor changes in the item scores are directly reflected in the trait scores make short scales very vulnerable for rater mistakes or errors. Finally, [Bakker and Lelkes \(2018\)](#) demonstrated that the associations between personality and political dimensions might differ when using short forms and longer personality questionnaires, which is logical because short forms do not capture all aspects of a trait in the same way longer measures do ([Credé et al., 2012](#)). Because of these reasons, we collected expert ratings using the NEO-FFI ([Costa & McCrae, 1989](#)), a 60-item personality inventory that allows studying not only the broad Big Five personality traits (i.e., neuroticism, extraversion, openness, agreeableness, and conscientiousness), but also several narrow facets underlying those broad traits ([Saucier, 1998](#)). Moreover, a marked advantage of the NEO-FFI is that, unlike the short measures, it has acceptable levels of internal consistency, both in self- and other-ratings ([Foltz et al., 1997](#)). No other data were collected during this study.

## Analysis

Based on the seven expert ratings of Trump's personality, we examined inter-rater reliability for the full NEO-FFI questionnaire, for each of the Big Five traits, and for the

---

2) The full, anonymized, dataset, can be obtained in the [Supplementary Materials](#)

facets as defined by [Saucier \(1998\)](#). These analyses were performed on item  $\times$  expert rater matrices, which means that these analyses tell us something about the extent to which the expert raters gave identical scores on (a subset of) the NEO-FFI items. Apart from these analyses, we also examined inter-rater reliability based on the facet and trait scores. In this case the (facet or trait) scale scores served as the basis of analysis, telling us to what extent the expert raters (dis)agree about Trump's personality at the higher, more abstract facet or trait level.

Inter-rater reliability was assessed using Krippendorff's Alpha, which provides an index of absolute agreement. Krippendorff's Alpha has a number of desirable features in the context of our particular study. First of all, Krippendorff's Alpha it is not affected by the number of observers. Second, it is not impacted by the number of scale points used by the observers. Third, it has a straightforward interpretation with 0 indicating the total absence of reliability (i.e., agreement at chance level) and 1 indexing perfect reliability. It is important to note that Krippendorff's Alpha can be negative, which simply implies that agreement among the observers is worse than the agreement one would observe in random data. Fourth and finally, one can empirically generate a sampling distribution using bootstrapping, which allows producing 95% confidence intervals for the Krippendorff's Alpha values ([Hayes & Krippendorff, 2007](#)). In the present study, we computed Krippendorff's Alphas, as well as 95% bootstrapped confidence intervals (95% CI) using the *icr* package in R. To construct the 95% CIs, we used the bootstrap procedure described in [Hayes & Krippendorff \(2007\)](#), relying on 20,000 bootstrap samples.

Because personality and social psychology scholars are more familiar with the Intra-Class Correlation Coefficient (ICC) than with Krippendorff's alpha, we also computed the *ICC(2,1)*. The *ICC(2,1)* measures absolute agreement, focuses on single raters (rather than the average across raters), and is based on a two-way random-effects model, which assumes that our group of experts is selected from a larger population of experts. The ICCs were computed using the *irr* package in R.

## Results

### Analyses Based on the Item Scores

A first set of analyses was performed on item  $\times$  expert rater matrices. This implies that the Krippendorff alpha's in these analyses pertain to the extent to which the expert raters gave identical scores to the individual NEO-FFI items. [Table 1](#) shows the mean scores, standard deviations (SDs), Intra-Class Correlation coefficients (ICCs), and Krippendorff's alpha coefficients along with their 95% bootstrapped confidence intervals for the full NEO-FFI, the Big Five traits and for each of [Saucier's \(1998\)](#) Big Five trait facets. As can be seen from [Table 1](#), the alpha inter-rater agreement coefficients vary quite a lot. When looking at the full NEO-FFI (i.e., inter-rater agreement across all 60

items), Krippendorff's alpha equals .56. When doing a trait-by-trait analysis, inter-rater agreement appears to be highest for the agreeableness items ( $\alpha = .71$ ), while for openness ( $\alpha = .47$ ), conscientiousness ( $\alpha = .31$ ), extraversion ( $\alpha = .53$ ), and neuroticism ( $\alpha = .47$ ) inter-rater agreement is substantially lower. Provided that Krippendorff suggested that "[I]t is customary to require  $\alpha \geq .800$ . Where tentative conclusions are still acceptable,  $\alpha \geq .667$  is the lowest conceivable limit" (Krippendorff, 2004, p. 241), these  $\alpha$ 's suggest that for four out of the five Big Five traits, inter-rater agreement is problematic.<sup>3</sup>

**Table 1**

*Mean Scores, Standard Deviations (SDs), Intra-Class Correlation Coefficients ICC(2,1), and Krippendorff's Alpha Coefficients Along with the 95% Bootstrapped Confidence Intervals (95% CI) for the full NEO-FFI Scale, the Separate Big Five Traits and for Each of Saucier's (1998) Big Five Trait Facets*

Big Five traits and facets	<i>M</i>	<i>SD</i>	<i>ICC(2,1)</i>	$\alpha$	95% CI
Big Five (all 60 items)	2.89	1.44	.57	.56	[.52, .60]
Openness (12 items)	2.19	1.30	.49	.47	[.37, .57]
Aesthetic Interests (3 items)	1.76	.77	.77	.71	[.63, .78]
Intellectual Interests (3 items)	1.33	.48	.94	.91	[.89, .94]
Unconventionality (4 items)	3.14	1.43	.40	.36	[.17, .53]
Conscientiousness (12 items)	2.73	1.38	.32	.31	[.20, .42]
Orderliness (5 items)	2.49	1.31	.26	.23	[.01, .43]
Goal-Striving (3 items)	3.76	1.14	.21	.14	[-.14, .39]
Dependability (4 items)	2.25	1.27	.45	.42	[.25, .58]
Extraversion (12 items)	3.40	1.24	.53	.53	[.44, .60]
Positive Affect (4 items)	2.61	1.07	.03	.04	[-.24, .29]
Sociability (4 items)	3.57	1.40	.76	.71	[.62, .79]
Activity (4 items)	4.04	.74	.09	.03	[-.22, .27]
Agreeableness (12 items)	2.81	1.68	.72	.71	[.63, .78]
Nonantagonistic Orientation (8 items)	2.64	1.77	.83	.82	[.72, .90]
Prosocial Orientation (4 items)	3.14	1.46	.47	.40	[.22, .57]
Neuroticism (12 items)	3.13	1.34	.50	.47	[.37, .57]
Negative Affect (5 items)	2.83	1.22	.18	.16	[-.07, .36]
Self- Reproach (7 items)	3.35	1.39	.66	.62	[.52, .71]

Looking at the trait facets, the story becomes even more polarized, with some facets having high inter-rater agreement (i.e., intellectual interests, nonantagonistic orientation),

3) A similar conclusion can be made when looking at the ICCs. As values less than .5 indicate poor reliability, the inter-rater reliability of several Big Five traits is problematic.



others for which inter-rater agreement is moderate to low (e.g., unconventionality, dependability, prosocial orientation), and still other facets for which inter-rater agreement does not exceed chance level provided that the 95% CI includes zero (e.g., goal-striving, positive affect, and negative affect)<sup>4</sup>. In sum, our findings suggest that inter-rater agreement is moderate to low, which means that the experts tend to differ quite a lot when rating US President Donald Trump's personality.

## Analyses Based on Scale Scores

In a second set of analyses, we computed scale scores for each facet and trait and used these scale scores as the basis of the analysis. In these analyses the Krippendorff alpha's measure the extent to which the scale scores were identical across experts. When performing this analysis at the facet level (i.e., on a facet scale score  $\times$  expert rater matrix), Krippendorff's alpha equals .65 (95% CI [.58, .70];  $ICC(2,1) = .66$ ), while at the trait-level (i.e., analysis performed on a trait score  $\times$  expert rater matrix), the Krippendorff's alpha is .60 (95% CI [.49, .71] );  $ICC(2,1) = .65$ ).

## Discussion

The goal of this study was to examine the extent to which political experts converge on their statements regarding the personality of political actors. Since political experts are constantly put on display by news media to comment on ongoing political events, their assessments have important consequences on viewers and, therefore, on potential voters. To examine this issue, we performed a case study on the personality of Donald Trump, with seven experts rating Trump's personality according to the Big Five personality traits, as well as their different underlying personality facets. Our analyses indicate that there was substantial disagreement between the political experts, and this was true for most traits and trait facets. In particular, Krippendorff's alpha was low when looking at the overall personality ratings, as well as for four of the five Big Five traits. At the facet level, of the 13 facets, only four obtained an acceptable Krippendorff alpha. Moreover, we found that, when focusing on facet or trait scores rather than item scores, Krippendorff's alpha was still low to moderate.

Looking at the traits and facets for which we observed high inter-rater agreement, we see that traits such as agreeableness, sociability and nonantagonistic orientation are performing well. This fits with the idea that interrater agreement tends to be high when the behavior can be observed and when the trait is evaluative in nature (John & Robins, 1993). Other traits, such as openness, neuroticism and conscientiousness have substantially lower interrater agreement, which is in line with the notion that such traits

---

4) Once again, the ICCs show a very similar pattern of findings.

are less clearly observable. However, we also observed some exceptions to this rule. For example, aesthetic and intellectual interests are generally low on observability, and yet inter-rater agreement is high. Conversely, activity is typically highly observable, yet the experts differed tremendously in their ratings<sup>5</sup>. These findings can possibly be explained by the fact that political experts, but also normal citizens, often do not have direct access to Trump's everyday behavior. Instead, they draw on information they receive via the media, and it has been shown that their coverage can be highly selective (e.g., [Bhatia et al., 2018](#); [Haselmayer et al., 2017](#)). In that sense the high interrater agreement of agreeableness is telling, provided that this trait is predictive of political success ([Joly et al., 2019](#)). Therefore, this trait probably gets a lot of attention when politicians are covered by the media, which implies that the trait is highly observable for experts and citizens. Of course, one should be cautious to generalize from this particular case to every politician, since the peculiarities of Trump might make that some aspects of his character get more media attention than other aspects.

Another important issue to keep in mind when reading our paper is that inter-rater consistency can be studied and operationalized in a number of ways. In the present paper, we looked at absolute agreement, or the extent to which different experts agreed on the absolute levels they used to describe Trump's personality. Another approach to qualifying agreement is to look at patterns of agreement, rather than absolute agreement. Using this approach, one might for example look at profile similarity, or the extent to which the different experts agree on Trump's personality profile (across traits, trait facets, or even items). In this case one tests whether the profile of peaks and valleys, rather than the absolute levels of those peaks and valleys, in Trump's personality is similar across raters ([Furr, 2008](#)). Yet another approach would be to ask experts to rate a number of politicians and to look at profile similarity across—rather than within—politicians. In this case, one would look at whether experts are consistent in their evaluation of Trump relative to other politicians. Finally, one can look at the reliability of single raters or the reliability of the average scores across the different raters. Because we were interested in the possible consequences of selective exposure to one or a few experts, our focus is on the reliability of single experts. Most studies relying upon expert ratings, however, are generally more interested in averaging across raters, and those ICCs are by definition higher than single measures ICCs. Because of the multitude of ways to look at inter-rater consistency, our conclusion is not that humans fail to reliably agree on the reputations of others, or that studies relying on expert ratings are per definition problematic. On the contrary, previous research has shown that reputation matters ([Hogan, 2005](#)) and that reputation can be reliably captured (e.g. [Rubenzer et al., 2000](#)).

---

5) Although we argue that activity is highly observable, the NEO-FFI item "I often feel as if I'm bursting with energy" refers to an internal feeling state that is not directly observable to people beyond the self.

What our study does show is that, despite those findings, there is substantial disagreement between individual experts when it comes to the absolute ratings of Trump's personality. In other words, the overall pattern of low to moderate interrater agreement, suggest the existence of substantial differences in how political experts rate the personality of politicians. Since research shows that a lot of what we know about others does not stem from a direct observation of the target but from intermediaries (Craig, 2008), such idiosyncrasies are highly relevant because they might potentially trickle down into the assessments and eventually the voting behavior of the large public.

---

**Funding:** The authors have no funding to report.

---

**Acknowledgments:** The authors have no additional (i.e., non-financial) support to report.

---

**Competing Interests:** The authors have declared that no competing interests exist.

---

**Author Contributions:** *Jeroen Joly*—Idea, conceptualization | Design planning | Resource provision (materials, participants, etc.) | Research implementation (software, hardware, etc.) | Data collection | Data management (storage, curation, processing, etc.) | Validation, reproduction, checking | Writing | Feedback, revisions | Supervision, mentoring | Project coordination, administration. *Joeri Hofmans*—Idea, conceptualization | Data management (storage, curation, processing, etc.) | Visualization (data presentation, figures, etc.) | Data analysis | Validation, reproduction, checking | Writing | Feedback, revisions | Supervision, mentoring.

---

**Data Availability:** Data and syntax are publicly available (see Joly & Hofmans, 2021).

---

## Supplementary Materials

For this article, the following Supplementary Materials are available (for access see [Index of Supplementary Materials](#) below):

- raw and original anonymized data at individual item level
- R syntax aggregating items at trait and facet levels and recreating analyses
- open peer-review

### Index of Supplementary Materials

Joly, J., & Hofmans, J. (2021). *Supplementary materials to "How reliable are personality judgments by political experts? The curious case of Donald Trump"* [Data, syntax]. OSF. <https://osf.io/mpwby>

Personality Science. (Ed.). (2023). *Supplementary materials to "How reliable are personality judgments by political experts? The curious case of Donald Trump"* [Open peer-review]. PsychOpen GOLD. <https://doi.org/10.23668/psycharchives.12535>

## References

- Bakker, B. N., & Lelkes, Y. (2018). Selling ourselves short? How abbreviated measures of personality change the way we think about personality and politics. *The Journal of Politics*, *80*(4), 1311–1325. <https://doi.org/10.1086/698928>
- Bhatia, S., Goodwin, G. P., & Walasek, L. (2018). Trait associations for Hillary Clinton and Donald Trump in news media: A computational analysis. *Social Psychological & Personality Science*, *9*(2), 123–130. <https://doi.org/10.1177/1948550617751584>
- Caprara, G. V. (2007). The personalization of modern politics. *European Review*, *15*(2), 151–164. <https://doi.org/10.1017/S1062798707000178>
- Caprara, G. V., Barbaranelli, C., Consiglio, C., Picconi, L., & Zimbardo, P. G. (2003). Personalities of politicians and voters: Unique and synergistic relationships. *Journal of Personality and Social Psychology*, *84*(4), 849–856. <https://doi.org/10.1037/0022-3514.84.4.849>
- Caprara, G. V., Francescato, D., Mebane, M., Sorace, R., & Vecchione, M. (2010). Personality foundations of ideological divide: A comparison of women members of parliament and women voters in Italy. *Political Psychology*, *31*(5), 739–762. <https://doi.org/10.1111/j.1467-9221.2010.00780.x>
- Costa, P. T., & McCrae, R. R. (1989). *The NEO-PI/NEO-FFI manual supplement*. Psychological Assessment Resources.
- Craik, K. H. (2008). *Reputation: A network interpretation*. Oxford University Press.
- Crédé, M., Harms, P., Niehorster, S., & Gaye-Valentine, A. (2012). An evaluation of the consequences of using short measures of the Big Five personality traits. *Journal of Personality and Social Psychology*, *102*(4), 874–888. <https://doi.org/10.1037/a0027403>
- Dietrich, B. J., Lasley, S., Mondak, J. J., Rempel, M. L., & Turner, J. (2012). Personality and legislative politics: The Big Five trait dimensions among U.S. state legislators. *Political Psychology*, *33*(2), 195–210. <https://doi.org/10.1111/j.1467-9221.2012.00870.x>
- Foltz, C., Morse, J. Q., Calvo, N., & Barber, J. P. (1997). Self- and observer ratings on the NEO-FFI in couples: Initial evidence of the psychometric properties of an observer form. *Assessment*, *4*(3), 287–295. <https://doi.org/10.1177/107319119700400308>
- Furr, R. M. (2008). A framework for profile similarity: Integrating similarity, normativeness, and distinctiveness. *Journal of Personality*, *76*(5), 1267–1316. <https://doi.org/10.1111/j.1467-6494.2008.00521.x>
- Gosling, S. D., Rentfrow, P. J., & Swann, W. B., Jr. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, *37*(6), 504–528. [https://doi.org/10.1016/S0092-6566\(03\)00046-1](https://doi.org/10.1016/S0092-6566(03)00046-1)
- Haselmayer, M., Wagner, M., & Meyer, T. M. (2017). Partisan bias in message selection: Media gatekeeping of party press releases. *Political Communication*, *34*(3), 367–384. <https://doi.org/10.1080/10584609.2016.1265619>
- Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, *1*(1), 77–89. <https://doi.org/10.1080/19312450709336664>

- Hogan, R. (2005). In defense of personality measurement: New wine for old whiners. *Human Performance*, 18(4), 331–341. [https://doi.org/10.1207/s15327043hup1804\\_1](https://doi.org/10.1207/s15327043hup1804_1)
- Iyengar, S., & Hahn, K. S. (2009). Red media, blue media: Evidence of ideological selectivity in media use. *Journal of Communication*, 59(1), 19–39. <https://doi.org/10.1111/j.1460-2466.2008.01402.x>
- John, O. P., & Robins, R. W. (1993). Determinants of interjudge agreement on personality traits: The Big Five domains, observability, evaluativeness, and the unique perspective of the self. *Journal of Personality*, 61(4), 521–551. <https://doi.org/10.1111/j.1467-6494.1993.tb00781.x>
- Joly, J., Soroka, S., & Loewen, P. (2019). Nice guys finish last: Personality and political success. *Acta Politica*, 54, 667–683. <https://doi.org/10.1057/s41269-018-0095-z>
- Knobloch-Westerwick, S., & Meng, J. (2009). Looking the other way: Selective exposure to attitude-consistent and counterattitudinal political information. *Communication Research*, 36(3), 426–448. <https://doi.org/10.1177/0093650209333030>
- Krippendorff, K. (2004). *Content analysis: An introduction to its methodology* (2nd ed.). Sage.
- Lilienfeld, S. O., Waldman, I. D., Landfield, K., Watts, A. L., Rubenzer, S., & Faschingbauer, T. R. (2012). Fearless dominance and the US presidency: Implications of psychopathic personality traits for successful and unsuccessful political leadership. *Journal of Personality and Social Psychology*, 103(3), 489–505. <https://doi.org/10.1037/a0029392>
- McAbee, S. T., & Connelly, B. S. (2016). A multi-rater framework for studying personality: The trait-reputation-identity model. *Psychological Review*, 123(5), 569–591. <https://doi.org/10.1037/rev0000035>
- Nai, A. (2019). The electoral success of angels and demons: Big Five, Dark Triad, and performance at the ballot box. *Journal of Social and Political Psychology*, 7(2), 830–862. <https://doi.org/10.5964/jspp.v7i2.918>
- Nai, A., & Maier, J. (2021). Can anyone be objective about Donald Trump? Assessing the personality of political figures. *Journal of Elections, Public Opinion, and Parties*, 31(3), 283–308. <https://doi.org/10.1080/17457289.2019.1632318>
- Nørgaard, A. S., & Klemmensen, R. (2019). The personalities of Danish MPs: Trait- and aspect-level differences. *Journal of Personality*, 87(2), 267–275. <https://doi.org/10.1111/jopy.12388>
- Ramey, A. J., Klingler, J. D., & Hollibaugh, G. E. Jr. (2019). Measuring elite personality using speech. *Political Science Research and Methods*, 7(1), 163–184. <https://doi.org/10.1017/psrm.2016.12>
- Rice, M. G., Rempel, M. L., & Mondak, J. J. (2020). Personality on the hill: Expert evaluations of U.S. senators' psychological traits. *Political Research Quarterly*, 74(3), 674–687. <https://doi.org/10.1177/1065912920928587>
- Rubenzer, S. J., Faschingbauer, T. R., & Ones, D. S. (2000). Assessing the U.S. presidents using the revised NEO personality inventory. *Assessment*, 7(4), 403–419. <https://doi.org/10.1177/107319110000700408>
- Saucier, G. (1998). Replicable item-cluster subcomponents in the NEO five-factor inventory. *Journal of Personality Assessment*, 70(2), 263–276. [https://doi.org/10.1207/s15327752jpa7002\\_6](https://doi.org/10.1207/s15327752jpa7002_6)

- Schumacher, G., & Zettler, I. (2019). House of Cards or West Wing? Self-reported HEXACO Traits of Danish politicians. *Personality and Individual Differences, 141*, 173–181. <https://doi.org/10.1016/j.paid.2018.12.028>
- Scott, C., & Medeiros, M. (2020). Personality and political careers: What personality types are likely to run for office and get elected? *Personality and Individual Differences, 152*, Article 109600. <https://doi.org/10.1016/j.paid.2019.109600>
- Vecchione, M., Schoen, H., González Castro, J. L., Ciecuch, J., Pavlopoulos, V., & Caprara, G. V. (2011). Personality correlates of party preference: The Big Five in five big European countries. *Personality and Individual Differences, 51*(6), 737–742. <https://doi.org/10.1016/j.paid.2011.06.015>
- Visser, B. A., Book, A. S., & Volk, A. A. (2017). Is Hillary dishonest and Donald narcissistic? A HEXACO analysis of the presidential candidates' public personas. *Personality and Individual Differences, 106*, 281–286. <https://doi.org/10.1016/j.paid.2016.10.053>
- Watts, A. L., Lilienfeld, S. O., Smith, S. F., Miller, J. D., Campbell, W. K., Waldman, I. D., Rubenzer, S. J., & Faschingbauer, T. J. (2013). The double-edged sword of grandiose narcissism: Implications for successful and unsuccessful leadership among U.S. presidents. *Psychological Science, 24*(12), 2379–2389. <https://doi.org/10.1177/0956797613491970>
- Wright, J. D., & Tomlinson, M. F. (2018). Personality profiles of Hillary Clinton and Donald Trump: Fooled by your own politics. *Personality and Individual Differences, 128*, 21–24. <https://doi.org/10.1016/j.paid.2018.02.019>



*Personality Science* (PS) is an official journal of the European Association of Personality Psychology (EAPP).



leibniz-psychology.org

PsychOpen GOLD is a publishing service by Leibniz Institute for Psychology (ZPID), Germany.