Personality Psychology

# Reducing the Overlap Between Machiavellianism and Subclinical Psychopathy: The M7 and P7 Scales

Michael P. Grosz[1] [a], Peter D. Harms[2], Michael Dufner[3], Livia Kraft[4], Eunike Wetzel[5]

[1] Department of Psychology, University of Münster, Münster, Germany, [2] Department of Management, University of Alabama, Tuscaloosa, Alabama, US, [3] Department of Psychology, Witten/Herdecke University, Witten, Germany, [4] Department of Psychology, University of Leipzig, Leipzig, Germany, [5] Department of Psychology, Otto-von-Guericke University Magdeburg, Magdeburg, Germany

## Collabra: Psychology

Machiavellianism (Mach) and subclinical psychopathy are two widely studied antagonistic personality traits with distinct theoretical conceptualizations. Mach is conceptualized by strategic deviousness, cynicism, and pragmatic morality, whereas subclinical psychopathy is conceptualized by impulsive antisocial tendencies, callousness, and rule-breaking. However, existing measures of the two traits are typically highly correlated and have very similar nomological networks. Notably, even though psychopathy scales should be more strongly positively associated with antisocial impulsivity and more strongly negatively associated with conscientiousness than Mach scales, existing Mach and psychopathy scales tend to be similarly related to these constructs. We created a new Mach scale, the M7, and a new psychopathy scale, the P7, by selecting items from existing Mach and psychopathy scales on the basis of the correlations of these items with antisocial impulsivity and conscientiousness. Across three studies (combined $N = 4,607$), the M7 and P7 showed acceptable to good psychometric properties in terms of closeness to unidimensionality, measurement precision, temporal stability, measurement invariance across language and gender groups, and convergent and discriminant validity (nomological network, self-other agreement, and interpersonal perceptions in group interactions). Most importantly, the new scales assess clearly distinct latent traits that are more in line with their theoretical conceptualizations than established scales are.

## Reducing the Overlap Between Machiavellianism and Subclinical Psychopathy: The M7 and P7 Scales

Machiavellianism (Mach), subclinical psychopathy, and subclinical narcissism are three interpersonally noxious and socially aversive personality traits that are frequently grouped together as the Dark Triad (Jones & Paulhus, 2014; Miller, Vize, Crowe, & Lynam, 2019; Paulhus & Williams, 2002). Although the theoretical conceptualizations of the three traits are distinct, recent empirical work has indicated that Mach and subclinical psychopathy scales exhibit a lack of discriminant validity. They are highly correlated with each other, have very similar nomological networks, and most of their items load substantially on the same higher order factor (e.g., McHoskey et al., 1998; Miller et al., 2017; Moshagen et al., 2018; Muris et al., 2017; Persson et al., 2019; Vize et al., 2018). Consequently, it has been suggested that established Machiavellianism scales are not sufficiently distinct from established psychopathy scales (e.g., McHoskey et al., 1998; Miller et al., 2017; Vize et al., 2018).

However, it is possible that not all existing Mach/psychopathy items show the same lack of discriminant validity as their scale scores do. In fact, several researchers have suggested that a multidimensional structure underlies Mach (e.g., Christie & Geis, 1970) and psychopathy scales

(e.g., Lilienfeld & Andrews, 1996). This indicates that not all Mach/psychopathy items measure the exact same construct. Some Mach items might show a stronger overlap with the theoretical concept of Mach and a weaker overlap with the theoretical concept of subclinical psychopathy than other Mach items and vice versa for subclinical psychopathy. The goals of the current set of studies were to identify such Mach and psychopathy items and use these items to develop new Mach and subclinical psychopathy scales that assess the two distinctive traits they are supposed to measure.

### Theoretical Conceptualizations of Machiavellianism and Subclinical Psychopathy

Although several clinical and forensic psychologists have considered aspects of Mach to be subordinate fachets of *clinical* psychopathy (e.g., Verschuere & te Kaat, 2019), most researchers interested in Mach and psychopathy as *subclinical* personality traits agree that Mach and psychopathy are theoretically distinct traits. The concept of Mach was developed by Christie and Geis (1970) to describe callous pragmatists without gross psychopathology who become influential through their manipulative tendencies. Accordingly, Mach is theoretically characterized by a cyn-

[a]  Correspondence concerning this manuscript should be addressed to Michael P. Grosz. E-mail: michael.grosz@uni-muenster.de; Phone: +49 (2 51) 83 - 34279; Mail: Department of Psychology, University of Münster, Fliednerstraße 21 (Pavillon 1), 48149 Münster, Germany.

ical world view, pragmatic morality, cold rationality, deviousness, and a belief in the effectiveness of (one's own) manipulative tactics (e.g., Christie & Geis, 1970; Jones & Paulhus, 2014). People high on Mach believe they need to be morally flexible, devious, and rationally cold to be successful and to avoid being exploited (e.g., Christie & Geis, 1970; Láng, 2015).

Subclinical psychopathy is theoretically characterized by thrill-seeking, impulsive antisocial tendencies, irresponsibility, amorality, and callousness (e.g., Jones & Paulhus, 2014; Miller et al., 2017; Paulhus & Williams, 2002). Like Mach, psychopathy is characterized by antagonism, but the reason people high on subclinical psychopathy act in an antagonistic fashion is that they are thrill-seeking and reckless and they rarely experience empathy and moral/self-conscious emotions such as remorse, guilt, and shame. Accordingly, a crucial difference in current theoretical conceptualizations of the two traits is that Mach is characterized by strategic rather than impulsive behavior, whereas subclinical psychopathy is characterized by thrill-seeking and reckless behavior rather than planful behavior (e.g., Jones & Paulhus, 2011, 2014; Miller et al., 2017). For example, Jones and Paulhus (2014, p. 29) stated, "the element of impulsivity is key in distinguishing psychopathy from Machiavellianism."

## Divergence Between Theoretical Concepts and Empirical Measurement

The theoretical difference between Mach and subclinical psychopathy, however, is not appropriately reflected in the Mach and psychopathy scales frequently used in research on the Dark Triad. First, several studies have found that Mach and subclinical psychopathy scale scores or items exhibit high loadings on the same higher-order factor (e.g., Miller et al., 2017; Moshagen et al., 2018; Persson et al., 2019). Second, two recent meta-analyses found average (uncorrected) correlations between the Mach and psychopathy scale scores of .58 (Muris et al., 2017) and .52 (Vize et al., 2018). Third, the same two meta-analyses compared the nomological networks of the Dark Triad traits and found that Mach and psychopathy profiles showed a high degree of convergence. Hence, several authors have suggested that Mach and subclinical psychopathy scales measure largely the same construct (e.g., Miller et al., 2017; Muris et al., 2017; Persson et al., 2019; Vize et al., 2018).

This lack of discriminant validity is usually attributed to Mach scales not accurately reflecting the strategic long-term aspects of Mach. For example, Miller et al. (2017) criticized that Mach scales are negatively correlated not only with agreeableness but also with conscientiousness. Accordingly, Mach was found to be moderately to strongly positively correlated with the impulsivity-related NEO-FFI facets. Furthermore, the empirical Big Five profiles of Mach scales were inconsistent with the prototypical expert-rated Big Five profile of Mach with regard to impulse control traits (Miller et al., 2017; see also, Persson, 2019).

Yet, psychopathy scales and their item content might also be responsible for the lack of discriminant validity. Psychopathy questionnaires used in Dark Triad research often contain items that assess core features of Mach—such as cynicism (Jonason & Webster, 2010) or strategic antisocial behavior (Jones & Paulhus, 2014). Some psychopathy scales that are used in Dark Triad research have even knowingly included subscales with Machiavellian labels and item content (e.g., Lilienfeld & Andrews, 1996). These psychopathy scales were inspired by theories on clinical psychopathy,

and many concepts of clinical psychopathy are much broader than the concept of subclinical psychopathy. For example, Cleckley's (1964) conceptualization of psychopathy includes a large and diverse set of traits such as impulsive antisocial acts, pathological egocentricity, deceitfulness, and social adeptness. For another example, the Triarchic Model of Psychopathy (Patrick et al., 2009) includes a variety of personality facets from three distinct phenotypic components: Disinhibition (i.e., impulsivity and negative affectivity), Boldness (i.e., social dominance, low stress reactivity, and thrill-adventure seeking), Meanness (e.g., callousness, cold-heartedness, and antagonism). Accordingly, many clinical conceptualizations of psychopathy include central features of Mach and narcissism such as strategic and cunning antisocial behavior or grandiose self-concepts (e.g., Verschuere & te Kaat, 2019). By contrast, the concept of *subclinical* psychopathy is narrower and focuses on the unique features of psychopathy (i.e., the features of psychopathy that do not overlap with Mach and narcissism such as thrill-seeking, impulsive antisocial tendencies, irresponsibility; e.g., Jonason & Webster, 2010; Jones & Paulhus, 2014). Taken together, problems exist for both the Mach and psychopathy scales that are used in Dark Triad research in terms of the inclusion of content that is not central to the construct of interest.

## The Current Research

The current set of studies was aimed at developing and validating a Mach scale and a subclinical psychopathy scale that are more in line with theoretical conceptualizations of Mach and subclinical psychopathy than established scales are. Following the established tradition of Dark Triad research, our intention was to develop brief scales that assess Mach and subclinical psychopathy in a unidimensional way. The aim for brevity was one reason why we aimed for unidimensional scales. It should be possible to assess Mach with a relatively unidimensional scale because Mach is a hierarchical construct that can be assessed at a higher level of abstraction (e.g., Rauthmann & Will, 2011). It should be possible to assess subclinical psychopathy with a relatively unidimensional scale because the construct of subclinical psychopathy is narrower than the construct of clinical psychopathy. That is, subclinical psychopathy focuses on the features that are nonoverlapping with Mach and narcissism. Generally, in the tradition of the Dark Triad research, the aim has been to create relatively unidimensional scales that isolate the unique features of Mach and psychopathy (e.g., Paulhus et al., 2020). Although established psychopathy scales that are used in Dark Triad research might also overlap with narcissism and sadism scales, we did not develop new narcissism and sadism measures because doing so would have diverted the focus and gone beyond the scope of the current research. The target audiences for the new scales consist of personality, social, developmental, and organizational psychologists who intend to assess Mach and subclinical psychopathy in a valid and economic way for research purposes.

In Study 1, we created a scale for Mach, the *M7*, and a scale for subclinical psychopathy, the *P7*. In addition, we investigated the M7 and P7 with respect to their closeness to unidimensionality, measurement precision, temporal stability, measurement invariance across two language and two gender groups, and convergent and discriminant validity. In Studies 2 and 3, we tested the convergent, discriminant, and predictive validity of the M7 and P7 in preregistered studies.[1] Supplemental tables and figures, the R code

for all of the analyses and the Monte Carlo simulation studies, the data used in the main analyses of the five samples, the preregistration for Studies 2 and 3, and other materials are available on the OSF project page: https://osf.io/6udfp/.[2]

## Study 1

## Method

**Samples.** We used data from three samples in Study 1. Samples 1 and 3 comprised 1,240 English-speaking participants (41% women; $M_{age}$ = 29.22, $SD$ = 8.91) and 1,743 English-speaking participants (49% women; $M_{age}$ = 28.50, $SD$ = 9.45), respectively. Both samples filled out an online questionnaire via Amazon's Mechanical Turk website in April 2012 and July 2012, respectively. Only participants from the United States were enabled to enroll in the study. They were paid $0.90. Both data collections were approved by the IRB at a Midwestern university in the US.

Sample 2 comprised 910 German-speaking participants (75% women, 25% men, 0.2% transgender; $M_{age}$ = 22.56, $SD$ = 4.67).[3] It was recruited via an online student research pool and flyers at two German universities between May 2016 and June 2017. Sample 2 was part of a longitudinal study and for the temporal stability analysis we used data from the first two measurement occasions, which were roughly six months apart. Of the 910 respondents at Time 1, 541 (59%) also responded at Time 2. The data collection took place on a computer in the laboratory at Time 1 and online at Time 2. Respondents received monetary compensation (eight to 15 Euros) or research participation credit at Time 1 and research participation credit or an Amazon voucher for eight Euros at Time 2. The data collection for Sample 2 was exempt from IRB approval. The data from Sample 2 were also used in Wetzel and Frick (2020).

**Measures.** The response format and options for the personality measures used in all studies are reported in Table S1. The Cronbach's alpha values for all scale scores can be found on the diagonals of Tables S2 to S4.

**M7 and P7.** We created the M7 scale and the P7 scale by selecting items from an initial item pool consisting of 25 Mach items and 28 psychopathy items: 18 MACH-IV items, 15 SD3 items, and 20 SRP-III items (Tables S5 and S6).[4] In the first step of item selection, we removed all items from the Mach item pool that were either (a) among the 1/3 of the Mach items that were most strongly positively correlated with the Impulsivity subscale from the Mini-Markers of Evil (Harms et al., 2013) in Sample 1 or (b) among the 1/3 of the Mach items that were most strongly negatively correlated with the Conscientiousness subscale from the Big Five Inventory (BFI; Lang et al., 2001) in Sample 2. Vice ver-

sa, we removed all items from the psychopathy item pool that were either among the 1/3 of the psychopathy items that were least strongly positively correlated with impulsivity in Sample 1 or among the 1/3 of the psychopathy items that were least strongly negatively correlated with conscientiousness in Sample 2. Twelve of the 25 Mach items and 14 of the 28 psychopathy items remained in the item pool. To be clear, the aim of this selection process was *not* to create a Mach scale that is positively related with conscientiousness, but rather to reduce the negative correlation with conscientiousness.

In the second step of item selection, we repeatedly fit a one-factor confirmatory factor analysis model to the remaining set of Mach/psychopathy items in both samples and iteratively removed the items with the lowest standardized item loadings until each item loaded ≥ .40 on the first factor in both samples. Seven Mach and 12 psychopathy items remained in the item pool. In the third step of item selection, we tried to balance the number of Mach and psychopathy items and to attain a number of items that optimally balances efficiency with external validity. We removed two psychopathy items for which the meaning of the German translations did not completely match the meaning of the English items. We removed two psychopathy items with contents that might partly reflect culture-specific circumstances (items: "I enjoy drinking and doing wild things" and "I have shoplifted"). Finally, we removed the psychopathy item with the lowest loading. The remaining two seven-item scales stood as our new Mach and subclinical psychopathy scales, the *M7* and *P7* (for item content and standardized loadings, see Table 1).

**Machiavellianism.** Mach was assessed in the three samples with three established scales, the MACH-IV scale (Christie & Geis, 1970), the Mach subscale from the Short Dark Triad Scale (SD3; Jones & Paulhus, 2014), and the Mach subscale from the Dirty Dozen (Jonason & Webster, 2010). The Dirty Dozen was only administered in Sample 3.

**Psychopathy.** Psychopathy was assessed in the three samples with three established scales: the Self-Report Psychopathy Scale (SRP-III; Paulhus et al., 2009), the Psychopathy subscale from the SD3 (Jones & Paulhus, 2014), and the Psychopathy subscale from the Dirty Dozen (Jonason & Webster, 2010).

**Antisocial impulsivity.** In Samples 1 and 3, impulsivity was assessed with a four-item subscale from the Mini-Markers of Evil (Harms et al., 2013). The Mini-Markers of Evil is a 24-item adjective scale for assessing aspects of antagonistic and socially aversive personality traits. The four items on the Impulsivity subscale were: "impulsive," "rebellious," "reckless," and "thrill-seeking."

**Narcissism.** Narcissism was assessed in all three samples with the nine-item subscale from the SD3 (Jones & Paulhus,

---

1 In the preregistrations, we called the scales the Mach7 and Psycho7 scales, but we later decided to rename them the M7 and P7. Furthermore, in the preregistration, the three studies of the current manuscript were portrayed as six separate studies. In the course of the streamlining process, we merged the six studies into three studies.

2 The shared data contain only the variables used in the main analyses and they do not include demographic variables to protect participants' anonymity. The shared data coming from Sample 2 contain only the responses to the M7 and P7 items and the scale scores for the other constructs.

3 One hundred thirty-two participants were excluded from the original sample ($N$ = 1,042) because they failed data quality checks.

4 By accident, only 19 of the 20 MACH-IV items were administered to Sample 1. Thus, we did not include the missing item in the initial item pool. Furthermore, we did not include one of the original MACH-IV items because it had no equivalent in the German version of the MACH-IV. We did not include three original SD3 items because they had no equivalent in the preliminary version of the SD3 that was used in Samples 1 and 3. Of the 34 English SRP-III items, only the 20 items that had equivalents in the German version of the SRP-III were included.

**Table 1: Item Content and Standardized Factor Loadings for the Items from the M7 and P7**

| # | Origin | Item content | Standardized factor loadings | | |
|---|---|---|---|---|---|
| | | | Sample 1 (English) | Sample 2 (German) | Sample 3 (English) |
| | | **M7** | | | |
| 1 | MACH-IV | It is wise to flatter important people. | .43 | .55 | .51 |
| 2 | MACH-IV | The best way to handle people is to tell them what they want to hear. | .60 | .62 | .49 |
| 3 | MACH-IV | Anyone who completely trusts anyone else is asking for trouble. | .41 | .47 | .51 |
| 4 | SD3 | Avoid direct conflict with others because they may be useful in the future. | .55 | .56 | .54 |
| 5 | SD3 | Whatever it takes, you must get the important people on your side. | .62 | .64 | .59 |
| 6 | SD3 | It's not wise to tell your secrets. | .50 | .55 | .54 |
| 7 | SD3 | There are things you should hide from other people because they don't need to know.[a] | .54 | .57 | .49 |
| | | **P7** | | | |
| 1 | SRP-III | I would get a "kick" out of scamming someone. | .76 | .45 | .80 |
| 2 | SRP-III | Rules are made to be broken. | .55 | .57 | .59 |
| 3 | SRP-III | It's fun to see how far you can push people before they get upset. | .78 | .60 | .75 |
| 4 | SRP-III | I've often done something dangerous just for the thrill of it. | .67 | .72 | .77 |
| 5 | SRP-III | I enjoy taking risks. | .55 | .70 | .71 |
| 6 | SD3 | I like to get revenge on authorities. | .67 | .47 | .68 |
| 7 | SD3 | People often say I'm out of control. | .69 | .46 | .70 |

*Note.* The standardized factor loadings are from six separate one-factor confirmatory factor analyses using the WLSMV estimator (Study 1). The German versions of the M7 and P7 can be found in Table S7. We have permission from Delroy Paulhus and Daniel Jones to use the SRP-III and SD3 items. The authors of the Mach-IV scales are already deceased. MACH-IV = MACH-IV scale (Christie & Geis, 1970); SD3 = Short Dark Triad (Jones & Paulhus, 2014); SRP-III = Self-report Psychopathy Scale III (Paulhus et al., 2009).

[a] The item content was slightly different in the German sample. The German item was translated from the following English item from the final version of the SD3: "There are things you should hide from other people to preserve your reputation."

2014). In Sample 3, narcissism was additionally assessed with the Dirty Dozen (Jonason & Webster, 2010).

*Big Five.* The Big Five personality traits were assessed with a 42-item German version of the BFI (Lang et al., 2001) in Sample 2 and with the 20-item Mini-IPIP (Donnellan et al., 2006) in Sample 3.

*HEXACO.* A subsample of Sample 2 (N = 446) filled out the German version of the HEXACO-60 (Moshagen et al., 2014), which assesses six general personality traits.

*Empathy.* Empathy was assessed in Samples 1 and 3 with the Toronto Empathy Questionnaire (Spreng et al., 2009). This questionnaire primarily measures the emotional aspects of empathy (i.e., affective insight into the feelings of others).

*Counterproductive work behavior.* Counterproductive work behavior was assessed in Samples 1 and 3 with the 12-item Organizational Deviance Scale and the seven-item Interpersonal Deviance Scale (Bennett & Robinson, 2000).

### Analyses.

#### Closeness to unidimensionality.

*Explained common variance.* We assessed the scales' closeness to unidimensionality by running a minimum rank factor analysis on the polychoric correlation matrix of each set of items and calculating the percentage of explained common variance (i.e., the proportion of explained variance that was explained by the first factor; e.g., Ten Berge & Sočan, 2004). To run this analysis, we used the FACTOR software (version 10.10.01; e.g., Lorenzo-Seva & Ferrando, 2013).

*Global model fit and residual item-pair correlations.* We additionally investigated unidimensionality by fitting a one-factor confirmatory factor analysis model to the M7 and P7 scales in each of the three samples and inspecting the global model fit and the residual item-pair correlations (i.e., differences between the observed and model-implied correlations). Poor model fit and residual item-pair correlations greater than .20 in absolute value can be interpreted as indicators of multidimensionality and local dependence (e.g., Reise et al., 2013). We used the benchmarks for good model fit proposed by Hu and Bentler (1999, p. CFI $\geqslant$ .95, RMSEA $\leqslant$ .06, and SRMR $\leqslant$ .08). To run these analyses, we used the WLSMV estimator from the R package lavaan (version 0.6-6; Rosseel, 2012).

*Measurement precision.* To investigate the measurement precision of the M7 and P7, we fit graded response models (Samejima, 1969) to the two scales and plotted their information curves. Additionally, we transformed the item fit

statistic $S-X^2$ into the effect size $r$ to assess the size of item misfit (Kang & Chen, 2011; see Tables S8 to S13). We ran these analyses with the R package mirt (version 1.29; Chalmers, 2012).

*Temporal stability.* To investigate the temporal consistency and test-retest reliability, we first computed the manifest correlation between the M7 (P7) at Time 1 and Time 2. Second, we fit a latent state model to the M7 (P7) items and correlated the latent trait at Time 1 with the latent trait at Time 2. In the models with latent variables, we assumed strong measurement invariance.[5] Furthermore, the errors of the same item were allowed to be correlated across the two measurement occasions to account for item specificity, which could otherwise inflate the estimates of stability over time (Marsh & Hau, 1996). In the temporal stability analyses, we used full information maximum likelihood estimation.

*Measurement Invariance Across Language and Gender.* To investigate measurement invariance across the two language groups (i.e., German and English), we first investigated configural invariance. That is, we tested whether all unconstrained factor loadings were significant in both language groups. Furthermore, we investigated metric measurement invariance (i.e., setting only factor loadings equal across groups) in global tests separately for the M7 and the P7. We did not test scalar invariance (i.e., factor loadings and intercepts set equal across groups) because some M7 and P7 items were administered with different response formats in German and English (five-point versus six-point rating scale; Table S1). To investigate measurement invariance across two gender groups (i.e., males and females), we again investigated configural invariance by testing whether all unconstrained factor loadings were significant in both gender groups. Furthermore, we tested metric and scalar invariance across the two gender groups in global tests separately for the M7 and the P7 and separately for each of the two language groups. The fit indices would indicate metric noninvariance if $\Delta$CFI $\geqslant$ .01, supplemented by $\Delta$RMSEA $\geqslant$ .015 or $\Delta$SRMR $\geqslant$ .030. The fit indices would indicate scalar noninvariance if $\Delta$CFI $\geqslant$ .01, supplemented by $\Delta$RMSEA $\geqslant$ .015 or $\Delta$SRMR $\geqslant$ .010 (Chen, 2007). To run these analyses, we used the WLSMV estimator from the R package lavaan (version 0.6-6; Rosseel, 2012) and the R package *semTools* (version 0.5-3; Jorgensen et al., 2018).

*Convergent and discriminant validity.* To investigate the convergent and discriminant validity of the M7 and P7, we correlated the M7 and P7 with a range of self-report measures that should or should not be related to Mach and psychopathy. We also compared the nomological networks of the M7 and P7 to the nomological network of established Mach and psychopathy scales (Mach-IV, SD3, Dirty Dozen, and SRP-III).

## Results and Discussion

### *Closeness to unidimensionality.*

*Explained common variance.* The results of the minimum rank factor analysis indicated that the first Mach factor explained 68%, 70%, and 65% of the common variance in
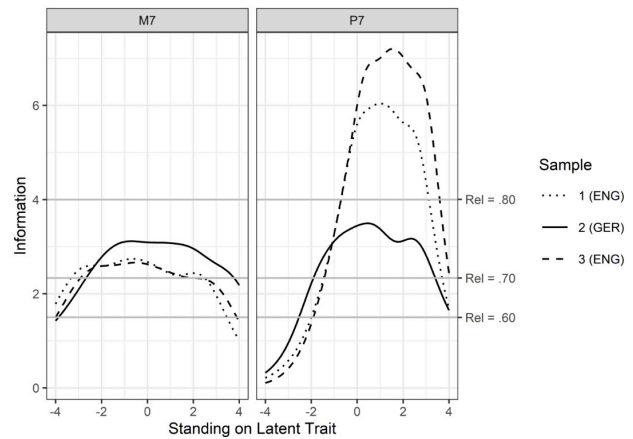


**Figure 1: Test information curves from graded response models fit to the M7 and P7 data from three samples ($N_1$ = 1,240; $N_2$ = 910; $N_3$ = 1,743). The superimposed horizontal grey lines indicate reliabilities of .60, .70, and .80, assuming a standard normally distributed latent trait in the population [for further details, see Grosz et al., 2019, Figure 1].**

the M7 items in Samples 1 to 3. The first psychopathy factor explained 75%, 68%, and 79% of the common variance in the P7 items in Samples 1 to 3. These explained amounts of common variance were all above the 60% benchmark proposed by Reise et al. (2013).

*Global model fit and residual item-pair correlations.* The fit of a unidimensional model to the M7 data of Samples 1 to 3 was not adequate according to the RMSEA (.109, .100, and .122.), acceptable according to the CFI (.925, .948, and .908), and good according to the SRMR (.070, .067, and .077). The fit of a unidimensional model to the P7 data was not adequate according to the RMSEA (.123, .112, and .143) and acceptable to good according to the CFI (.966, .938, and .965) and SRMR (.071, .079, and .085). There were no residual item-pair correlations exceeding .20 in absolute value in any of the three samples for the M7 and only one in one of the three samples for the P7 (Table S11). Taken together, these results suggested that the M7 and P7 were close enough to being unidimensional.

Following the suggestion by an anonymous reviewer, we additionally ran exploratory and confirmatory factor analyses to investigate whether a model with more than one factor would also fit the M7 and P7, respectively. For each of the two scales, a model with two correlated factors fit better than the respective unidimensional model (for details on the models with two correlated factors, see Figures S1 and S2). That said, on the basis on the above-mentioned evidence indicating that the M7 and P7 are close enough to unidimensionality, we treated the M7 and P7 as unidimensional scales.

**Measurement precision.** According to the item fit statistics, most of the items fit the graded response models well. The few items that showed significant misfit did not show

---

5  We tested the assumption of strong measurement invariance for each construct in a separate model by inspecting the differences in the fit indices between models with configural, weak, and strong invariance ($\Delta$CFI < .01; Chen, 2007). The assumption of strong measurement invariance held for both scales.

**Table 2: Measurement Invariance Across Language Groups (German and English)**

| Model | N | df | χ² | CFI | RMSEA | SRMR |
|---|---|---|---|---|---|---|
| M7 configural IV | 3515 | 28 | 626.46 | .926 | .110 | .071 |
| M7 metric IV | 3515 | 34 | 634.31 | .926 | .100 | .071 |
| P7 configural IV | 3523 | 28 | 884.2 | .961 | .132 | .079 |
| P7 metric IV | 3523 | 34 | 1051.65 | .954 | .130 | .085 |

*Note.* CFI = comparative fit index; IV = Invariance; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual.

**Table 3: Measurement Invariance Across Gender Groups (Male and Female)**

| Model | N | df | χ² | CFI | RMSEA | SRMR |
|---|---|---|---|---|---|---|
| German language group | | | | | | |
| M7 configural IV | 894 | 28 | 151.67 | .950 | .100 | .069 |
| M7 metric IV | 894 | 34 | 164.4 | .948 | .093 | .071 |
| M7 scalar IV | 894 | 50[a] | 180.28 | .948 | .076 | .071 |
| P7 configural IV | 898 | 28 | 180.63 | .936 | .110 | .082 |
| P7 metric IV | 898 | 34 | 187.65 | .935 | .100 | .084 |
| P7 scalar IV | 898 | 48 | 199.66 | .936 | .084 | .083 |
| English language group | | | | | | |
| M7 configural IV | 2609 | 28 | 487.17 | .922 | .112 | .071 |
| M7 metric IV | 2609 | 34 | 521.15 | .918 | .105 | .074 |
| M7 scalar IV | 2609 | 48 | 526.67 | .919 | .087 | .073 |
| P7 configural IV | 2614 | 28 | 752.69 | .959 | .141 | .083 |
| P7 metric IV | 2614 | 34 | 758.78 | .959 | .128 | .083 |
| P7 scalar IV | 2614 | 48 | 768.9 | .959 | .107 | .083 |

*Note.* CFI = comparative fit index; IV = Invariance; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual.

[a] The German M7 scalar IV model had 50 (and not 48) degrees of freedom because we needed to collapse two response categories in the German M7 data because nobody responded with the lowest answer category in one of the gender groups.

large misfit ($r < .15$; Tables S8 to S13). The information curves in [Figure 1](#) show that the M7 measured its latent trait with acceptable measurement precision (reliability ⩾ .70) across a broad range of latent trait levels in all three samples (i.e., from around -3 to +2.5). This suggests that the M7 does a reasonably effective job of discriminating between levels of Mach at low, high, and especially moderate levels of Mach. The P7 scale measured its latent trait with acceptable measurement precision (reliability ⩾ .70) for lower levels of the latent trait (-1.5 to 0) in all three samples. At average to high latent trait levels (0 to 3.5), measurement precision was acceptable to high in the German sample (reliability ⩾ . 75) and high to very high (reliability ⩾ . 85) in the two U.S. samples. This suggests that the P7 does a good job of discriminating between different psychopathy levels across a broad latent trait range. It discriminates better at higher ends of its latent trait than at the lower ends, which is typical for undesirable traits (see e.g., Grosz et al., 2019; Webster & Jonason, 2013).

**Temporal Stability**. The temporal consistency (rank-order stability) for the two scales was high. The manifest (latent) correlation between Time 1 and Time 2 was .71 (.84) for the M7 scale and .78 (.92) for the P7 scale. Thus, the M7 and P7 seem to assess relatively stable traits. The high manifest correlations across measurement times also indi-

cate high test-retest reliabilities.

**Measurement Invariance Across Language and Gender Groups**. The results of the measurement invariance analyses indicated that the assumption of configural measurement invariance over language and gender groups was tenable for both the M7 and the P7: All unconstrained factor loadings were significant (all $ps ⩽ .001$). The assumption of metric measurement invariance over language and gender groups was also tenable for both the M7 and the P7 (all ΔCFI < .01, ΔRMSEA < .015, and ΔSRMR < .010; for details, see Tables [2](#) and [3](#)). These results indicate that the constructs measured by the M7 and P7 have the same meaning across the language and gender groups. The assumption of scalar measurement invariance over gender groups was also tenable for both the M7 and the P7 (all ΔCFI < .01 and ΔSRMR < .010; for details, see [Table 3](#)).

Given that scalar measurement invariance held, we used manifest scores to test gender differences. In both German and English, men scored higher on the M7 (manifest means = 3.25 and 3.28, respectively) than women (manifest means = 3.09 and 3.21, respectively; Cohen's $d$ = 0.24 and 0.11, respectively; both $ps$ = .002). Similarly, in both German and English, men scored higher on the P7 (manifest means = 2.59 and 2.33, respectively) than women (manifest means = 2.25 and 1.94, respectively; Cohen's $d$ = 0.52 and 0.60, re-

**Table 4: Manifest Correlations for the Machiavellianism and Psychopathy Scales with the Other Scales averaged across Samples 1 to 3**

| | Machiavellianism scales | | | | Psychopathy scales | | | |
|---|---|---|---|---|---|---|---|---|
| | M7 | MACH-IV | SD3 | DD | P7 | SRP-III | SD3 | DD |
| Narcissism | | | | | | | | |
|     SD3 | .20 | .18 | .28 | .33 | .34 | .35 | .35 | .16 |
|     DD | .31 | .31 | .35 | .64 | .42 | .44 | .41 | .48 |
| Impulsivity | .17 | .21 | .24 | .37 | .62 | .60 | .48 | .36 |
| Empathy | -.16 | -.30 | -.32 | -.40 | -.43 | -.52 | -.53 | -.57 |
| Counterproductive work behavior | | | | | | | | |
|     Interpersonal deviance | .11 | .23 | .21 | .36 | .53 | .58 | .49 | .44 |
|     Organizational deviance | .14 | .23 | .21 | .41 | .53 | .58 | .45 | .41 |
| Big Five | | | | | | | | |
|     Extraversion | -.09 | -.14 | -.05 | .09 | .15 | .17 | .09 | -.08 |
|     Agreeableness | -.32 | -.49 | -.42 | -.34 | -.33 | -.35 | -.44 | -.52 |
|     Conscientiousness | -.09 | -.24 | -.13 | -.19 | -.29 | -.32 | -.27 | -.19 |
|     Neuroticism | .15 | .18 | .12 | .14 | .02 | -.02 | .09 | .16 |
|     Openness | -.11 | -.13 | -.10 | -.13 | -.04 | -.03 | -.11 | -.15 |
| HEXACO | | | | | | | | |
|     Honesty-Humility | -.58 | -.58 | -.60 | | -.34 | -.43 | -.44 | |
|     Emotionality | -.08 | -.22 | -.16 | | -.33 | -.39 | -.26 | |
|     Extraversion | -.05 | -.17 | -.05 | | .06 | .15 | -.01 | |
|     Agreeableness | -.19 | -.28 | -.30 | | -.19 | -.14 | -.37 | |
|     Conscientiousness | .09 | -.01 | .03 | | -.27 | -.31 | -.22 | |
|     Openness to Experience | -.10 | -.05 | -.11 | | .12 | .16 | .09 | |

*Note.* Zero-order correlations of various Machiavellianism and psychopathy scale scores (i.e., unweighted average scores) with the scale scores for other constructs averaged across Samples 1 to 3 ($N_1$ = 1,240; $N_2$ = 910; $N_3$ = 1,743; the $N$ is 446 for correlations involving HEXACO scales because the HEXACO was only administered to parts of Sample 2). The unaveraged correlations for each sample can be found in Tables S2 to S4. Please note that many scales were not administered in all of the three samples. Thus, the correlation coefficients in the table are sometimes based on only one or two of the three samples. DD = Dirty Dozen (Jonason & Webster, 2010); MACH-IV = MACH-IV scale (Christie & Geis, 1970); SD3 = Short Dark Triad (Jones & Paulhus, 2014); SRP-III = Self-report Psychopathy Scale III (Paulhus et al., 2009).

spectively; both *ps* ≤ .001). These differences are consistent with the gender differences found in previous research (e.g., Jones & Paulhus, 2014).

**Convergent and discriminant validity.** Table 4 displays the zero-order correlations between the M7 and P7 and the other measures averaged across Samples 1 to 3 (for the full correlation matrices, see Tables S2 to S4).

*Convergent validity.* Across the three samples, the M7 was very strongly correlated with the MACH-IV (*rs* = .60, .70, and .60) and the Mach subscale of the SD3 (*rs* = .83, .81, and .82). The M7 correlated only .36 with the Mach subscale of the Dirty Dozen. The P7 was very strongly correlated with the SRP-III (*rs* = .87, .82, and .87), the Psychopathy subscale from the SD3 (*rs* = .78, .74, and .79), and the Psychopathy subscale from the Dirty Dozen (*r* = .55). This strong convergent validity with established Mach and psychopathy measures implies that future research with the M7 and P7 will not need to start from scratch but will be able to build on the rich empirical research on Mach and psychopathy.

*Discriminant validity.* In line with the aim to reduce the empirical overlap between Mach and psychopathy, the M7 and P7 correlated only .23, .28, and .25 with each other in Samples 1 to 3. These correlation coefficients are descrip-

tively substantially lower than the high average intercorrelations found in two recent meta-analyses (Muris et al., 2017; Vize et al., 2018). The correlation coefficients are still positive, which would be expected because the constructs have antagonism in common.

The M7 was weakly to moderately correlated with impulsivity, conscientiousness, and counterproductive work behavior, whereas the P7 was strongly to very strongly correlated with these constructs (Table 4). This pattern of results is in line with the rule-breaking and disinhibition attributed to psychopathy rather than to Mach (e.g., Jones & Paulhus, 2014; Miller et al., 2017). Similarly, the M7 was weakly to moderately negatively associated with empathy, whereas the P7 was strongly negatively associated with empathy (Table 4), fitting with the notion that psychopathy is characterized by callousness (e.g., Jones & Paulhus, 2014). Furthermore, although both the M7 and the P7 were strongly negatively correlated with agreeableness (Table 4), only the M7 was very strongly negatively correlated with honesty-humility: *r* = -.58 (M7) versus *r* = -.34 (P7). The very strong negative association of the M7 with honesty-humility is consistent with the strategic dishonest and exploitative tendencies often attributed to Mach (e.g., Christie & Geis,

1970; Jones & Paulhus, 2014).

*Comparison with established Machiavellianism and psychopathy scales.* The nomological network of the M7 scale was similar to the nomological networks of established Mach scales except that, on a descriptive level, the M7 showed slightly lower positive correlations with impulsivity and counterproductive work behavior and lower negative correlations with conscientiousness and empathy than the other Mach scales (Table 4; see also Tables S2 to S4). The nomological network of the P7 scale was also similar to the nomological networks of established psychopathy scales except that, on a descriptive level, the P7 showed slightly lower negative correlations with empathy and honesty-humility than the other psychopathy scales. Thus, the nomological networks of the new M7 and P7 scales are largely compatible with existing Mach and psychopathy scales. At the same time, the M7 and P7's nomological networks were more consistent with the theoretical constructs of Mach and subclinical psychopathy than the nomological networks of established scales, indicating that we successfully reduced the overlap between Mach and psychopathy measures through the item selection procedure we applied in Study 1.

## Study 2

In Study 1, we demonstrated that the M7 and P7 showed considerably less overlap and less similarity in terms of their nomological networks than established measures. However, the nomological networks of the M7 and P7 were only investigated in an exploratory manner. The aim of Study 2 was to conduct a preregistered, confirmatory test of the differences in the nomological networks of the M7 and P7. To investigate the nomological network and discriminant validity in Study 2, we used an approach similar to Miller et al. (2017): We examined the M7 and P7's relations to facets of the IRT-based IPIP-NEO-120 (Maples et al., 2014). We investigated the similarity of the Five-Factor Model (FFM) profiles associated with the M7 and P7 using double-entry q-correlations ($r_{ICC}$). We hypothesized that the similarity between the M7 and P7 FFM profiles would be lower than the similarity found by Miller et al. for existing Mach and psychopathy measures ($r_{ICC}$ = .97).

Furthermore, we tested several hypotheses about the associations of the M7 and P7 with impulsivity-related NEO facets. We hypothesized that the M7 would be less strongly positively associated with Excitement Seeking from Extraversion and Immoderation (or "Impulsiveness") from Neuroticism than the P7, and we hypothesized that the M7 would be less strongly negatively associated with Self-Discipline and Cautiousness from Conscientiousness than the P7.

Finally, we investigated the discriminant validity of the M7 and P7 by correlating them with self-reported aggressive behavior. Both people high in Mach and people high in psychopathy show antisocial behavior, but people high in Mach do so in a covert way (e.g., lying, cheating), whereas people high in psychopathy even show overtly aggressive behavior (e.g., Jones & Neria, 2015). Thus, we hypothesized that the M7 would be less strongly positively correlated with self-reported overt physically aggressive behavior than the P7.

## Method

**Participants and procedure.** We recruited participants on the crowdsourcing website www.prolific.ac and asked them to fill out an online survey. Because the participants completed the survey anonymously (e.g., IP addresses were not tracked), Study 2 was exempt from IRB approval. People were only able to participate if they indicated on www.prolific.ac that they were between 18 and 80 years old, that their first language was English, and that they were from the US, UK, or Canada. Participants received 2 British pounds.

On the basis of a Monte Carlo simulation study, we specified in the preregistration that the data collection would be terminated after 450 people had clicked on the link on the last page of the survey that brought them back to www.prolific.ac. A total of 460 people answered at least one of the questions in our survey, and thus the sample size for Study 4 ranged from 453 to 460 (63% women, 36% men, 1% transgender; $M_{age}$ = 35.08, $SD$ = 12.75). Sixty-eight percent lived in the UK, 22% in the US, 8% in Canada, and one person in France.

*Data exclusion criteria.* We ran all analyses twice. First, we ran the analysis without excluding any participants on the basis of their response behavior. Second, as a robustness check, we ran the analysis again after excluding 19 people who showed signs of careless responding (see preregistration). Because the two analyses yielded very similar results, we report only the analysis without exclusions (for robustness check results, see Tables S14 to S16).

*Preregistration.* Before we collected the data, we preregistered the rationale, hypotheses, design, exclusion criteria, and analysis plan for Study 2 at https://osf.io/4s6c9.

## Measures.[6]

Mach and psychopathy were assessed with the M7 and P7, respectively ($\alpha$ = .75 and .82, respectively; Table 1). The NEO-FFI dimensions and facets were assessed with the IRT-based IPIP-NEO-120 (Maples et al., 2014). The IPIP-NEO-120 measures each of the five domains with 24 items (all $\alpha$s between .84 and .91) and each facet with four items (all $\alpha$s between .63 and .91; for details on $\alpha$s, see Table 5). Overt (physical) aggression was measured with the nine-item Physical Aggression subscale from the Aggression Questionnaire ($\alpha$ = .88; Buss & Perry, 1992). The response format and options for the personality measures are reported in Table S1.

## Results and Discussion

All hypotheses were tested as specified in the preregistration. The analysis was run with Mplus (version 8; Muthén & Muthén, 2017) and the R package MplusAutomation (version 0.7-2; Hallquist & Wiley, 2018).

**FFM profile similarity.** In line with our hypothesis, the FFM profile similarity between the M7 and P7 ($r_{ICC}$ = .75) was significantly lower than the profile similarity reported by Miller et al. (2017; $r_{ICC}$ = .97, for the difference in $r_{ICC}$s: $z$ = -6.46, $p \leqslant$ .001, for details, see Table 5). This finding suggests that the M7 and P7 do *not* suffer from the same dis-

---

6 We also administered the Narcissism subscale from the Short Dark Triad (Jones & Paulhus, 2014) and the Mach and Psychopathy subscales from the Short Dark Tetrad (Paulhus et al., 2020). Because we did not have any hypotheses for these three scales, their results are reported in the supplemental online material (Table S17).

**Table 5: Cronbach's Alphas and Correlations of NEO Dimensions and Facets with the M7 and P7 (Study 2)**

|  | α | M7 | P7 |
|---|---|---|---|
| Neuroticism | .91 | .20 | .07 |
| Anxiety | .84 | .14 | .01 |
| Angry hostility | .82 | .23 | .26 |
| Depression | .90 | .22 | .14 |
| Self-consciousness | .67 | .06 | -.19 |
| Immoderation | .77 | .16 | .12 |
| Vulnerability | .82 | .02 | -.10 |
| Extraversion | .91 | .04 | .38 |
| Warmth | .85 | -.01 | .19 |
| Gregariousness | .81 | -.13 | .17 |
| Assertiveness | .84 | .11 | .30 |
| Activity | .80 | .09 | .18 |
| Excitement seeking | .79 | .16 | .63 |
| Positive emotions | .77 | -.04 | .20 |
| Openness | .84 | -.18 | -.09 |
| Fantasy | .76 | .16 | .29 |
| Aesthetics | .76 | -.08 | -.11 |
| Feelings | .77 | -.13 | -.21 |
| Actions | .81 | -.22 | .00 |
| Ideas | .84 | -.17 | -.16 |
| Values | .63 | -.19 | -.13 |
| Agreeableness | .86 | -.54 | -.64 |
| Trust | .82 | -.29 | -.12 |
| Straightforwardness | .77 | -.58 | -.71 |
| Altruism | .67 | -.22 | -.20 |
| Compliance | .78 | -.40 | -.66 |
| Modesty | .79 | -.34 | -.52 |
| Tendermindedness | .66 | -.18 | -.16 |
| Conscientiousness | .88 | -.16 | -.37 |
| Competence | .82 | .14 | .12 |
| Order | .77 | -.03 | -.21 |
| Dutifulness | .70 | -.35 | -.49 |
| Achievement striving | .74 | .00 | -.01 |
| Self-discipline | .86 | -.12 | -.20 |
| Cautiousness | .91 | -.25 | -.59 |

*Note.* The sample sizes ranged from 453 to 460. We assessed the NEO-FFI dimensions and facets with the IRT-based IPIP-NEO-120 (Maples et al., 2014).

criminant validity issues as the widely used Mach and psychopathy scales do (e.g., Miller et al., 2017; Muris et al., 2017; Vize et al., 2018).

**Associations with impulsivity-related NEO facets and physical aggressiveness.** In line with our hypotheses, the M7 was less positively associated with excitement seeking ($r$ = .21) and self-reported physical aggression ($r$ = .47) and less negatively associated with cautiousness ($r$ = -.32) than the P7 ($r$s = .78, .76, and -.68, respectively; all $p$s ⩽ .002 for differences between Z-transformed $r$s; see also Table S16). These results are in line with the notion that, in contrast to people high in psychopathy, people high in Mach rarely jump spontaneously into things because of thrill or sensa-

tion seeking (e.g., Jones & Paulhus, 2014) and—in line with their long-term perspective—they are less inclined to show antisocial behavior in an overt way (e.g., Jones & Neria, 2015).

We did not find support for our hypothesis that the M7 would be less positively correlated with immoderation ($r$ = .16) than the P7 ($r$ = .09; $p$ = .856 for difference between Z-transformed $r$s). One reason for this finding might be the item content of the Immoderation scale: The items are concerned with (binge) eating and cravings, not with antisocial forms of impulsiveness or thrill-seeking. Poythress and Hall (2011) suggested in their review that psychopathy is more strongly characterized by sensation seeking than oth-

er forms of impulsivity. Another reason might be that immoderation is a facet of neuroticism, and Mach is more strongly characterized by neuroticism than psychopathy (Tables 4 and 5). The positive association between the M7 and neuroticism is in line with the notion that Mach is characterized by a negative view of the world and other people (e.g., Christie & Geis, 1970; Láng, 2015). We also did not find support for the hypothesis that the M7 would be less negatively correlated with self-discipline ($r = -.20$) than the P7 ($r = -.33$; $p = .018$ for difference between Z-transformed $r$s). That said, the difference was in the predicted direction and almost reached the Bonferroni-corrected significance level of .0083.

## Study 3

In Studies 1 and 2, we used self-report data to investigate the convergent and discriminant validity of the M7 and P7. Yet, the validity of scales needs to be evaluated with more than one assessment method (e.g., Campbell & Fiske, 1959). Hence, in Study 3, we investigated how the self-reported M7 and P7 were correlated with (a) the informant-reported M7 and P7 and (b) interpersonal perceptions *by* previously unacquainted others. Additionally, we investigated how a person's M7 and P7 scores were related to interpersonal perceptions *of* previously unacquainted others. In terms of informant reports, Malesza and Kaczmarek (2018) found correlations between self-ratings and the average of three peer ratings of .49 for Mach and .46 for psychopathy. Similarly, Jones and Paulhus (2014) reported self-other agreements of .42 for Mach and .57 for psychopathy. Hence, we hypothesized that the self-reported M7 (P7) would be positively correlated with the informant-reported M7 (P7). In terms of discriminant validity, we hypothesized that the self-reported M7 would be more positively correlated with the informant-reported M7 than with the informant-reported P7 and that the self-reported P7 would be more positively correlated with the informant-reported P7 than with the informant-reported M7.

People high in psychopathy but not people high in Mach might be initially perceived as impulsive by previously unacquainted others due to their thrill-seeking tendencies. In line with this possibility, Rogers, Le, Buckels, Kim, and Biesanz (2018) found that psychopathy but not Mach was positively associated with being perceived as "aggressive and unrestrained" and negatively associated with being perceived as reasonable by unacquainted others after 3 min of dyadic face-to-face interactions. Hence, we hypothesized that a target's P7 score would be positively related to previously unacquainted perceivers' ratings of the target as impulsive (i.e., the target's P7 score should be positively related to the target effect on impulsivity). Furthermore, due to their overt antisocial and thrill-seeking behavior, people high in psychopathy should be perceived as less likeable and trustworthy than people low in psychopathy. Accordingly, Rogers et al. (2018) found that psychopathy was negatively associated with being perceived as likeable and trustworthy. Hence, we hypothesized that a target's P7 score would be negatively related to perceivers' ratings of the target as likeable and trustworthy.

We did not formulate hypotheses about how people high

in Mach would be perceived by others because they should maintain a low profile and conceal their antisocial tendencies. That said, Mach is theoretically characterized by a mistrust of other people and a fear that other people will exploit them (e.g., Christie & Geis, 1970; Láng, 2015). In line with this, Rogers et al. (2018) found that Mach was negatively associated with perceiving previously unacquainted others as trustworthy. Accordingly, we hypothesized that a perceiver's M7 score would be negatively related to the perceiver's ratings of others as trustworthy (i.e., a perceiver's M7 score should be negatively related to the perceiver effect on trustworthiness).

## Method

**Participants, procedure, and measures**. The data collection was part of a larger study (see https://osf.io/dv2eb/). Target participants were recruited via online advertisements and social media. They filled out an online survey including demographic questions, the M7 ($\alpha = .79$), the P7 ($\alpha = .79$; for response format and options, see Table S1), and several other personality questionnaires that we did not analyze. In the online survey, targets were asked to invite three to five informants (friends, romantic partner, and family) to fill out an online survey including an informant version of the M7 and P7 (for item content, see preregistration). The preregistered plan was to terminate the data collection after 250 targets had completely participated. The final sample size for the self-other agreements was 254 targets (80% women; $M_{age} = 24.60$, $SD = 4.39$) and 748 informants (62% women; $M_{age} = 29.16$, $SD = 11.18$).

The target participants were invited to come to the lab after they had filled out the online survey. This resulted in a sample size of 256 in 50 groups.[7] There were two group sessions in the lab, which were spaced one week apart. In both sessions, participants interacted in the same group of four to six individuals for approximately 60 minutes. In one of the sessions, participants engaged in communal activities, such as playing a getting-acquainted game or designing a group logo together. The other session focused on competition and participants were instructed to persuade their group members to take a certain stance on certain topics during group discussions. The order of the sessions was randomized. Group members provided round-robin ratings before and after each session. We investigated how the M7 and P7 scores relate to the round-robin ratings at the end of the second session because participants had the greatest amount of information about each other at this point. In the round-robin ratings, impulsivity was assessed with one item: "This person is impulsive." Trustworthiness was assessed with three items: "This person seems trustworthy," "I would trust this person," and "I would entrust a secret to this person." Likeability was also assessed with three items: "I like this person," "I get along well with this person," and "I think this person is likeable" (for the response format and the item content in German, see Table S18).

The IRB of the German Society for Psychology (DGPs) approved the study. The data were collected from March 2019 to August 2019. We preregistered the rationale, hypotheses, design, exclusion criteria, power analyses, and analysis plan of Study 3 on April 23, 2019 when approximately

---

7 For the interpersonal perceptions of previously unacquainted others, there were two participants more than for the self-other agreements because two participants did not fill out the online surveys but nevertheless participated in the lab activities.

130 people had at least partially participated in the study: https://osf.io/xthb7.

## Results and Discussion

**Informant-reported M7 and P7.** In line with our hypothesis, the latent trait of the self-reported M7 and the latent trait of the informant version of the M7 (aggregated informant ratings) correlated .35 ($p \leq$ .001) in the preregistered model (for details on the analyses, see the preregistration). In line with our hypothesis, the latent trait of the self-reported P7 and the latent trait of the informant version of the P7 (aggregated informant ratings) correlated .54 ($p \leq$ .001) in the preregistered model. The descriptively lower self-other agreement for the M7 than for the P7 is in line with self-other agreement found for the Mach and psychopathy scales from the SD3 (Jones & Paulhus, 2014) and might be explained by people high in Mach showing less overt (antisocial) behavior than people high in psychopathy.

In line with our hypothesis, we found that self-other agreement for the M7 ($r = $ .35) was significantly larger than the correlation between self-reported M7 and other-reported P7 ($r = $ -.08; for the difference in Z-transformed $r$s: $z = $ 3.69, $p \leq$ .001). Also as hypothesized, we found that self-other agreement for the P7 ($r = $ .54) was significantly larger than the correlation between self-reported P7 and other-reported M7 ($r = $ .17; for the difference in Z-transformed $r$s: $z = $ 3.49, $p \leq$ .001).

**Interpersonal Perceptions by Previously Unacquainted Others.** We tested the hypothesized associations with the restricted maximum likelihood estimation for cross-sectional social relations models (Nestler, 2016; Tables S19 to S30). We deviated from the preregistered analysis plan in one way: In contrast to our expectations in the preregistered plan, incomplete responses could not be included in the analysis with restricted maximum likelihood estimation. Hence, we substituted missing values with the mean of the respective variable. We did so for the 2% of the round-robin ratings and the less than 1% of the M7 and P7 scores that were missing.

In contrast to our hypotheses and the findings by Rogers et al. (2018), the social relations model analyses indicated that a target's P7 score was not positively associated with the impulsivity, the likeability, or trustworthiness ratings of the target (Tables S20, S24, and S28). The results were puzzling because long-term acquaintances seem to have a relatively accurate picture of a target's level of P7 as indicated by the high level of self-other agreement (see above). Perhaps it takes longer than 2 hours of group interaction to notice the impulsivity and antisocial tendencies of people high on the P7 scale. Another potential reason for the lack of an effect of a target's P7 score on the target's likeability and trustworthiness ratings might be the relatively low variance on targets' likeability and trustworthiness (see Tables S24 and S28). Furthermore, the current study had only enough power to detect medium-sized effects. Perhaps the P7 is not moderately but only weakly associated with relatively unacquainted individuals' perceptions of impulsivity, likeability, and trustworthiness. The nonsignificant positive effect of the P7 score (Coef = 0.08, $p = $ .11) and the nonsignificant negative effect of the M7 score (Coef = -0.08, $p = $ .25) on being perceived as impulsive might suggest that people high on the P7 tend to be perceived as more impulsive than people high on the M7.

In line with our hypothesis, a perceiver's M7 score was negatively related to her or his trustworthiness ratings of others (Table S29). This finding is in line with the idea that people high in Mach are cynical about human nature and that they have negative views of other people (e.g., Christie & Geis, 1970). Interestingly, not only did people high on the M7 perceive others as less trustworthy, but they were also perceived by others as less trustworthy and less likeable than people low on the M7 (Tables S23 and S27). Perhaps their negative views of others led to unfavorable interpersonal behavior, which in turn led them to make unfavorable impressions on others. This finding also casts doubt on the idea that people high in Mach can successfully conceal their antagonism, but more research is needed on this question.

## General Discussion

The results of our three studies indicated that Mach and subclinical psychopathy can be distinguished not only theoretically but also empirically. Established Mach and psychopathy scales have frequently come under fire for their very strong correlations with each other and their largely indistinguishable nomological networks (e.g., McHoskey et al., 1998; Miller et al., 2017; Muris et al., 2017; Vize et al., 2018). The results for the newly created M7 and P7 offer evidence that these issues can be mitigated through careful item selection. Compared with established Mach and psychopathy scales, the M7 and P7 showed more moderate correlations with each other, and their nomological networks were more distinct than the nomological networks of established scales.

The nomological networks of the M7 and P7 were not only more distinct but were also more convergent with the theoretical conceptualizations of Mach and psychopathy than the nomological networks of established Mach and psychopathy scales. For example, impulsivity is a central feature of the theoretical concept of psychopathy but not of the theoretical concept of Mach. This theoretical difference has been absent from widely used measures of Mach and psychopathy (e.g., Miller et al., 2017). However, the theoretical difference was manifest in the nomological networks of the M7 and P7. This and the other findings from the current studies indicate that the M7 and P7 are more construct-valid measures than established Mach and psychopathy scales. Despite these differences, the M7 and P7 were strongly correlated with established Mach and psychopathy scales. We intentionally selected items from established measures to ensure that research with the M7 and P7 can benefit from and extend the rich body of existing literature on Mach and psychopathy. Researchers who have collected data on original measures (MACH-IV, SRP-III, and SD3) can even retroactively score the M7 and/or P7 from their existing data sets. Further, researchers who plan on assessing Mach and subclinical psychopathy in future studies now have access to economic, psychometrically sound scales that are able to achieve a significantly better differentiation between the two constructs.

Study 3 indicated that people high on the M7 scale trusted newly acquainted individuals less than people low on the M7 scale did. This lack of trust might drive the antisocial tendencies of Mach (e.g., moral flexibility, cold rationality) as suggested by previous research and theories on Mach (e.g., Christie & Geis, 1970; Láng, 2015). Furthermore, people high on the M7 were perceived as less trustworthy and less likeable than people low on the M7 in Study 3, which is at odds with some previous Mach theories. Future research will be needed to probe whether and how people high on Mach (a) keep a low profile and (b) influence others.

## Limitations and Future Directions

Of course, the M7 and P7 are not perfect measures. For example, some model fit indices suggested that a one-dimensional model did not fit adequately. However, Reise et al. (2013) found that fit indices were less diagnostic of parameter bias in structural equation modeling than a combination of the explained common variance index with the percentage of contaminated correlations. The values of the explained common variance index indicated a more than sufficient degree of unidimensionality for the M7 and P7 across three large samples in the current study. Furthermore, less than adequate model fit indices are quite commonly encountered for personality scales (e.g., Grosz et al., 2019). The reason is probably that personality items often measure not only the broad construct of interest but also facets and nuances (Figures S1 and S2). Hence, we are confident that the psychometric properties of the M7 and P7 are as good as the properties of established personality measures.

We found that the P7 was more strongly associated with self-reported counterproductive work behavior and aggressive behavior and less strongly negatively associated with honesty-humility than the M7. A task for future research is to investigate whether and how the M7 and P7 differentially predict objectively assessed behavioral criteria.

Finally, Moshagen et al. (2018) argued that there is a higher-order factor that accounts for a general tendency to "maximize one's individual utility—disregarding, accepting, or malevolently provoking disutility for others others—, accompanied by beliefs that serve as justifications" (p. 656). If, as we argued, established scales suffer from strongly overlapping content, Moshagen et al.'s conclusion could be due to the fact that the measures of antagonistic traits such as Mach and psychopathy are insufficient rather than that they all measure the same general tendency. The M7 and P7 could be used to revisit this issue.

## Conclusion

Taken together, the newly developed M7 and P7 scales show that Mach and psychopathy can be empirically distinguished. They provide the opportunity for personality researchers to mitigate the vexing problem that the empirical findings in this literature have not reflected the distinct theoretical conceptualizations of Mach and subclinical psychopathy.

## Contributions

Contributed to conception and design: MPG, EW, PDH

Contributed to acquisition of data: MPG, EW, PDH, MD, LK

Contributed to analysis and interpretation of data: MPG, EW, PDH

Drafted and/or revised the article: MPG, EW, PDH, MD, LK

Approved the submitted version for publication: MPG, EW, PDH, MD, LK

## Competing interests

No competing interests exist.

## Supplemental material

Supplemental material for this manuscript can be found on the OSF project page https://osf.io/6udfp/.

## Data accessibility statement

The data used in the main analyses of the five samples can be found at the OSF project page https://osf.io/6udfp/. The shared data contain only the variables used in the main analyses and they do not include demographic variables to protect participants' anonymity. The shared data coming from Sample 2 contain only the responses to the M7 and P7 items and the scale scores for the other constructs.

## Figure titles and legends

*Figure 1*. Test information curves from graded response models fit to the M7 and P7 data from three samples ($N_1$ = 1,240; $N_2$ = 910; $N_3$ = 1,743). The superimposed horizontal gray lines indicate reliabilities of .60, .70, and .80, assuming a standard normally distributed latent trait in the population (for further details, see Grosz et al., 2019, p. Figure 1).

# REFERENCES

Bennett, R. J., & Robinson, S. L. (2000). Development of a measure of workplace deviance. *Journal of Applied Psychology, 85*(3), 349–360. https://doi.org/10.1037/0021-9010.85.3.349

Buss, A. H., & Perry, M. (1992). The Aggression Questionnaire. *Journal of Personality and Social Psychology, 63*(3), 452–459. https://doi.org/10.1037/0022-3514.63.3.452

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*(2), 81–105. https://doi.org/10.1037/h0046016

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software, 48*(6), 1–29. https://doi.org/10.18637/jss.v048.i06

Christie, R., & Geis, F. L. (1970). *Machiavellianism*. Academic Press.

Cleckley, H. (1964). *The mask of sanity: An attempt to reinterpret the so-called psychopathic personality*. Mosby.

Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The mini-IPIP scales: tiny-yet-effective measures of the Big Five factors of personality. *Psychological Assessment, 18*(2), 192–203. https://doi.org/10.1037/1040-3590.18.2.192

Grosz, M. P., Emons, W. H. M., Wetzel, E., Leckelt, M., Chopik, W. J., Rose, N., & Back, M. D. (2019). A Comparison of Unidimensionality and Measurement Precision of the Narcissistic Personality Inventory and the Narcissistic Admiration and Rivalry Questionnaire. *Assessment, 26*(2), 281–293. https://doi.org/10.1177/1073191116686686

Hallquist, M. N., & Wiley, J. F. (2018). MplusAutomation: An R Package for Facilitating Large-Scale Latent Variable Analyses in Mplus. *Structural Equation Modeling: A Multidisciplinary Journal, 25*(4), 1–18. https://doi.org/10.1080/10705511.2017.1402334

Harms, P. D., Jones, D. N., & Brummel, B. J. (2013). *The Mini-Markers of Evil: A new measure for the assessment of dark personality traits*. Association for Research in Personality, Charlotte, NC.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Jonason, P. K., & Webster, G. D. (2010). The dirty dozen: A concise measure of the dark triad. *Psychological Assessment, 22*(2), 420–432. https://doi.org/10.1037/a0019265

Jones, D. N., & Neria, A. L. (2015). The Dark Triad and dispositional aggression. *Personality and Individual Differences, 86*, 360–364. https://doi.org/10.1016/j.paid.2015.06.021

Jones, D. N., & Paulhus, D. L. (2011). The role of impulsivity in the Dark Triad of personality. *Personality and Individual Differences, 51*(5), 679–682. https://doi.org/10.1016/j.paid.2011.04.011

Jones, D. N., & Paulhus, D. L. (2014). Introducing the short dark triad (SD3)a brief measure of dark personality traits. *Assessment, 21*(1), 28–41. https://doi.org/10.1177/1073191113514105

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2018). *semTools: Useful tools for structural equation modeling* (R package version 0.5‐1.). https://CRAN.R-project.org/package=semTools

Kang, T., & Chen, T. T. (2011). Performance of the generalized S-X2 item fit index for the graded response model. *Asia Pacific Education Review, 12*(1), 89–96. https://doi.org/10.1007/s12564-010-9082-4

Kenny, D. A. (2015). *Measuring Model Fit: SEM:Fit*. http://davidakenny.net/cm/fit.htm

Láng, A. (2015). Machiavellianism and early maladaptive schemas in adolescents. *Personality and Individual Differences, 87*, 162–165. https://doi.org/10.1016/j.paid.2015.07.039

Lang, F. R., Lüdtke, O., & Asendorpf, J. B. (2001). Validity and psychometric equivalence of the German version of the Big Five Inventory in young, middle-aged and old adults. *Diagnostica, 47*(3), 111–121. https://doi.org/10.1026//0012-1924.47.3.111

Lilienfeld, S. O., & Andrews, B. P. (1996). Development and preliminary validation of a self-report measure of psychopathic personality traits in noncriminal population. *Journal of Personality Assessment*, *66*(3), 488–524. https://doi.org/10.1207/s15327752jpa6603_3

Lorenzo-Seva, U., & Ferrando, P. J. (2013). FACTOR 9.2: A Comprehensive Program for Fitting Exploratory and Semiconfirmatory Factor Analysis and IRT Models. *Applied Psychological Measurement*, *37*(6), 497–498. https://doi.org/10.1177/0146621613487794

Malesza, M., & Kaczmarek, M. C. (2018). The convergent validity between self- and peer-ratings of the Dark Triad personality. *Current Psychology*, 1–8. https://doi.org/10.1007/s12144-018-9906-7

Maples, J. L., Guan, L., Carter, N. T., & Miller, J. D. (2014). A test of the International Personality Item Pool representation of the Revised NEO Personality Inventory and development of a 120-item IPIP-based measure of the five-factor model. *Psychological Assessment*, *26*(4), 1070–1084. https://doi.org/10.1037/pas0000004

Marsh, H. W., & Hau, K.-T. (1996). Assessing goodness of fit: Is parsimony always desirable? *The Journal of Experimental Education*, *64*(4), 364–390. https://doi.org/10.1080/00220973.1996.10806604

McHoskey, J. W., Worzel, W., & Szyarto, C. (1998). Machiavellianism and psychopathy. *Journal of Personality and Social Psychology*, *74*(1), 192–210. https://doi.org/10.1037/0022-3514.74.1.192

Miller, J. D., Hyatt, C. S., Maples-Keller, J. L., Carter, N. T., & Lynam, D. R. (2017). Psychopathy and Machiavellianism: A Distinction Without a Difference? *Journal of Personality*, *85*(4), 439–453. https://doi.org/10.1111/jopy.12251

Moshagen, M., Hilbig, B. E., & Zettler, I. (2014). Faktorenstruktur, psychometrische Eigenschaften und Messinvarianz der deutschsprachigen Version des 60-item HEXACO Persönlichkeitsinventars. *Diagnostica*, *60*(2), 86–97. https://doi.org/10.1026/0012-1924/a000112

Moshagen, M., Hilbig, B. E., & Zettler, I. (2018). The dark core of personality. *Psychological Review*, *125*(5), 656–688. https://doi.org/10.1037/rev0000111

Muris, P., Merckelbach, H., Otgaar, H., & Meijer, E. (2017). The malevolent side of human nature: A meta-analysis and critical review of the literature on the dark triad (narcissism, Machiavellianism, and psychopathy). *Perspectives on Psychological Science*, *12*(2), 183–204. https://doi.org/10.1177/1745691616666070

Muthén, L. K., & Muthén, B. O. (2017). *Mplus User's Guide* (8th ed.). Muthén & Muthén.

Nestler, S. (2016). Restricted maximum likelihood estimation for parameters of the social relations model. *Psychometrika*, *81*(4), 1098–1117. https://doi.org/10.1007/s11336-015-9474-9

Patrick, C. J., Fowles, D. C., & Krueger, R. F. (2009). Triarchic conceptualization of psychopathy: Developmental origins of disinhibition, boldness, and meanness. *Development and Psychopathology*, *21*(3), 913–938. https://doi.org/10.1017/s0954579409000492

Paulhus, D. L., Buckels, E. E., Trapnell, P. D., & Jones, D. N. (2020). Screening for Dark Personalities: The Short Dark Tetrad (SD4). *European Journal of Psychological Assessment*, 1–15. https://doi.org/10.1027/1015-5759/a000602

Paulhus, D. L., Hemphill, J. F., & Hare, R. D. (2009). *Manual for the self-report psychopathy scale (SRP-III)* [Unpublished Manuscript].

Paulhus, D. L., & Williams, K. M. (2002). The Dark Triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality*, *36*(6), 556–563. https://doi.org/10.1016/s0092-6566(02)00505-6

Persson, B. N. (2019). Searching for Machiavelli but finding psychopathy and narcissism. *Personality Disorders: Theory, Research, and Treatment*, *10*(3), 235–245. https://doi.org/10.1037/per0000323

Persson, B. N., Kajonius, P. J., & Garcia, D. (2019). Revisiting the Structure of the Short Dark Triad. *Assessment*, *26*(1), 3–16. https://doi.org/10.1177/1073191117701192

Poythress, N. G., & Hall, J. R. (2011). Psychopathy and impulsivity reconsidered. *Aggression and Violent Behavior*, *16*(2), 120–134. https://doi.org/10.1016/j.avb.2011.02.003

Rauthmann, J. F., & Will, T. (2011). Proposing a multidimensional Machiavellianism conceptualization. *Social Behavior and Personality: An International Journal*, *39*(3), 391–403. https://doi.org/10.2224/sbp.2011.39.3.391

Reise, S. P., Scheines, R., Widaman, K. F., & Haviland, M. G. (2013). Multidimensionality and structural coefficient bias in structural equation modeling: A bifactor perspective. *Educational and Psychological Measurement*, *73*(1), 5–26. https://doi.org/10.1177/0013164412449831

Rogers, K. H., Le, M. T., Buckels, E. E., Kim, M., & Biesanz, J. C. (2018). Dispositional malevolence and impression formation: Dark Tetrad associations with accuracy and positivity in first impressions. *Journal of Personality*, *86*(6), 1050–1064. https://doi.org/10.1111/jopy.12374

Rosenthal, R., & DiMatteo, M. R. (2001). Meta-analysis: Recent developments in quantitative methods for literature reviews. *Annual Review of Psychology*, *52*(1), 59–82. https://doi.org/10.1146/annurev.psych.52.1.59

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *ETS Research Bulletin Series*, *1968*(1), i–169. https://doi.org/10.1002/j.2333-8504.1968.tb00153.x

Spreng, R. N., McKinnon, M. C., Mar, R. A., & Levine, B. (2009). The Toronto Empathy Questionnaire: Scale development and initial validation of a factor-analytic solution to multiple empathy measures. *Journal of Personality Assessment*, *91*(1), 62–71. https://doi.org/10.1080/00223890802484381

Ten Berge, J. M. F., & Sočan, G. (2004). The greatest lower bound to the reliability of a test and the hypothesis of unidimensionality. *Psychometrika*, *69*(4), 613–625. https://doi.org/10.1007/bf02289858

Verschuere, B., & te Kaat, L. (2019). What are the core features of psychopathy? A prototypicality analysis using the Psychopathy Checklist-Revised (PCL-R). *Journal of Personality Disorders*, *34*(3), 1–10. https://doi.org/10.1521/pedi_2019_33_396

Vize, C. E., Lynam, D. R., Collison, K. L., & Miller, J. D. (2018). Differences among dark triad components: A meta-analytic investigation. *Personality Disorders*, *9*(2), 101–111. https://doi.org/10.1037/per0000222

Webster, G. D., & Jonason, P. K. (2013). Putting the "IRT" in "Dirty": Item response theory analyses of the Dark Triad Dirty Dozen-An efficient measure of narcissism, psychopathy, and Machiavellianism. *Personality and Individual Differences*, *54*(2), 302–306. https://doi.org/10.1016/j.paid.2012.08.027

Wetzel, E., & Frick, S. (2020). Comparing the validity of trait estimates from the multidimensional forced-choice format and the rating scale format. *Psychological Assessment*, *32*(3), 239–253. https://doi.org/10.1037/pas0000781

# SUPPLEMENTARY MATERIALS

## Peer Review Comments

Download: https://collabra.scholasticahq.com/article/17799-reducing-the-overlap-between-machiavellianism-and-subclinical-psychopathy-the-m7-and-p7-scales/attachment/46235.docx?auth_token=UO7R70924VhVICKg7cqN