Left-wing ideology, mental health and visual presentation

Abstract
There is a vast amount of research on the relationships between psychological variation and political opinion. One finding that has received attention in recent years is the relationship between mental health and left-wing opinions. A number of studies have found that positive aspects of mental health (e.g., life satisfaction) and negative (e.g., depression, anxiety) relate to left-wing ideology, such that left-wing ideology relates to worse mental health and lower positivity. We sought to test to which extent these relationships hold across a wide variety of measures of mental health and political ideology. We studied a sample of fairly representative American adults on Prolific (n = 978) who filled out 76 mental health-related questions, and 41 political ideology related questions. Overall, we found that every aspect of mental health correlated negatively with overall leftism (r's 18 diagnoses = -0.20, last 2 weeks symptoms = -0.14, MMPI-subset = -0.17, life satisfaction = -0.15, all p's < .001). Regression models showed that diagnoses were the primary predictor of leftism, as the other measures of psychopathology did not predict leftism once diagnoses were included in the same model. Controlling for age, sex, and race reduced the strength of association by about 25%. Alternatively, one might consider that leftism causes psychopathology. Controlling for self-reported measures and demographics, the diagnosis score predicted leftism (beta = 0.10, p < .001). Measurement invariance testing found only very small amounts of bias and in different directions across items. Thus, the relationship cannot be explained as resulting from biased measurement. Finally, we replicated associations between unnatural hair color, number of piercings, and tattoos, leftism, and mental health diagnoses (r's about 0.20).

Commented [1]: report average r

Keywords: political orientation, mental health, body modification, tattoo, piercing, hair color, measurement invariance, IRT, DIF, MCV, Prolific

Introduction
There is a vast amount of research on the relationships between psychological variation and political opinions or ideology (Bakker et al., 2021; Brandt et al., 2021; de Vries et al., 2022; Jost, 2021). An interesting take is that political differences between the right and the left might be caused by a family of personality differences related to an open versus closed orientation to behavior and experience (Bakker et al., 2021; Jost et al., 2003). This leaves open the question of whether leftism might be related to having more experiences that deviate from the norm, resulting in stigma and impacting mental health. If leftism reflects greater openness to non-normative stimuli, this orientation might manifest not only in abstract attitudes but also in concrete, visible behaviors—such as body modifications—that invite social stigma and, potentially, adverse mental health outcomes which, in turn, could exacerbate mental health disparities already observed among leftists (Bernardi, 2020; Kirkegaard, 2020). This possibility aligns with extensive evidence that body modification (tattoos, piercings, hair color) has a positive relationship with deviance, psychopathology and emotional distress (D'Ambrosio et al., 2013; Deschesnes et al., 2006; Dutton & Kirkegaard, 2022b; Khosla et al., 2010; King & Vidourek, 2013; Kirkegaard, 2023; Koch et al., 2010; Laumann & Derick, 2006; McCandlish & Pearson, 2023; Mortensen et al., 2019; Perrotta, 2021; Stirn et al., 2006) but also that visible tattoos carry significant stigma (Timming et al., 2017). One common

Commented [2]: take? lol. Use "One interesting perspective"

interpretation considers body art as mere self-expression (Atkinson, 2003) and need for uniqueness (Koch et al., 2010), although it is unclear whether body art such as tattooing is a marker of ideological rebellion or trendiness (Kosut, 2006).

These psychosocial risks may contribute to the well-documented relationship between leftism and poorer mental health. Indeed, many studies have found that positive aspects of mental health (e.g., life satisfaction) and negative aspects (e.g., depression, anxiety) are related to left-wing ideology such that left-wingers (leftists) have worse mental health measured in a variety of ways (Bernardi, 2020; Kirkegaard, 2020). One study in particular gives a small hint about the possibility that depression leads to political orientations but that this link may be driven by ideological/policy alignment rather than affective loyalty (Bernardi, 2020, p. 9). A question that needs to be addressed is whether leftism and body modification display a robust relationship. It is not known whether these relationships are causal and if so, in which way.

Critically, interpretations of these relationships depend on whether measurement bias along the dimension of political ideology influences the results. The reason why it is important is because measurement equivalence between groups ensures that the questions do not have different connotations across cultural groups. One large scale analysis of American and Dutch subjects across many psychological scales found the percentage of Differential Item Functioning (DIF) varies greatly across psychological scales but that the impact on the global estimates are rather minimal (Brandt et al., 2021), and one study examining depression and anxiety scales in Finnish subjects also found little bias (Dutton & Kirkegaard, 2024).

The present study therefore aims to answer multiple, yet related questions. The first goal is to examine the relationship between various mental health scales and questions as well as a large number of political ideology items, such that the generality of the pattern could be assessed. The second goal is to examine to which degree measurement bias influences the associations. The third goal is to examine whether visual presentation related to body modifications relate to leftism and psychopathology.

Commented [3]: global estimates of what? Unclear.

Commented [4]: Hypotheses? what are they? If none, saw we are just exploring. But, your intro sets up natural expectations...

Commented [5]: pattern of associations

Data and methods

We recruited American participants from Prolific (https://www.prolific.com/). Only American and British subjects are available if representative samples are desired, and the American subjects were chosen. To ensure data quality, we administered a series of 4 attention check items, instructing the participant to pick a target word among a list of words. If a participant picks another word, this participant is ineligible, and removed from the data. Only 2 participants failed any of the attention checks. The study had 2 waves of data collection. The first on February 16th 2024 (n = 495), and the second on March 10th 2025 (n = 483). The final sample was therefore relatively large (n = 978). The surveys in both cases were identical. The reason for collecting a second wave was that our analyses of the first wave indicated that key statistical model tests were underpowered with several inconclusive results (p-values near 0.05). The first wave was collected using the representative for age, sex, race' mode, while the second wave was collected using the 'representative for age, sex, and political identification' mode. This change was because online samples usually have a left-wing political skew, which lowers statistical

power and may bias some analyses. Our results confirmed a slight bias as the second sample had lower score on leftism (d = 0.22) than the first.

For measuring mental health and well-being, we chose a diverse set of items to maximize breadth in the construct of general psychopathology (p factor) as well as items measuring positive emotionality. While tendencies towards negative (depression, anxiety etc.) and positive thoughts (happiness, satisfaction etc.) correlate negatively and thus could be considered to measure the same overall aspect of personality (Dunkel et al., 2021), they are not generally considered unidimensional and thus may have different associations with political ideology. First, we chose 24 items from the MMPI-2 that had been shown to have strong loadings on the general factor in that item pool based on prior research (Kirkegaard & Nyborg, 2021). Second, we created 23 items concerning people's negative feelings in the last 14 days. These kinds of items are often found in mental health screeners. Third, we created 11 items measuring better mental health, as opposed to worse mental health. These items concerned satisfaction with various aspects of life (e.g., "Life satisfaction", and "In most ways my life is close to my ideal." on a Likert 1-7 scale). Fourth, we asked about whether the subject had received a doctor's diagnosis of any of 18 of the most common diagnoses found in the DSM-5.[1] In total, thus, we had 76 items covering mental health.

For political ideology, we similarly wanted to measure the construct with breadth. For this purpose, we created a set of 41 items concerning typical political disagreements in the USA at the time of this survey. Item examples are: "The world is suffering from overpopulation.", "Affirmative action is discrimination.", "I support Israel against Hamas.", and "Progressive taxation (higher tax rates for higher incomes) is the best way to structure a tax system.". These items are measured on a 1-7 scale, going from strongly disagree to strongly agree.

Concerning visual presentation, we asked subjects about whether they had dyed their hair an unnatural color ("As an adult, have you ever dyed your hair with an unnatural hair color? (pink, green, blue, purple, yellow, etc)"), on which bodyparts they have tattoos ("choose all bodyparts that apply from: hand, arm, chest/stomach, back, neck, leg, feet, face/head, other"), and whether they had any piercings ("choose all bodyparts that apply from: ears, eyebrow, nose (non-septum), mouth area (inside), mouth area (outside), cheek, nipple, navel, genital, other"). For the tattoos and piercings, these were converted into total counts as the study was underpowered to investigate differential relationships for each type of tattoo or piercing.

Auxiliary items included a question about age, sex (Y chromosome), sexual orientation (choose one from heterosexual, homosexual, bisexual, asexual, other), transgender, race/ethnicity (choose any that apply from Asian, American Indian/Alaskan, Black, Hispanic/Latino, Jewish, White, Mixed, Hawaiian/Pacific Islander, Other), and immigration background (foreign born, second generation, third generation, older

[1] ADHD, alcoholism, substance abuse, autism spectrum, anti-social personality, bipolar, borderline personality, depression, general anxiety, obsessive compulse, panic, paranoia, social phobia, specific phobia, post-traumatic stress, schizophrenia, schizoid personality, and sleeping disorder(s).

Commented [6]: You need to explain this in more detail. It seems like you just recruited the 2nd half to get what you wanted. Did you assess power before data collection to arrive at the N you needed for finding some size effect at a given alpha? obviously not. So what informed your first sample N? What were the results of a post-hoc power analyses that would tell you what your power for the analyses actually was. "Several inconclusive results" doesn't seem to me as necessarily indicating your power was a problem.

Commented [7]: scale? give a sample item

generation). Racial variables were combined into composite categories if multiple options were selected (e.g., White + Hispanic). The composite categories revealed that the standard US census coding was the most appropriate for this dataset, and the remaining options were collapsed into an Other category. After this classification, the sample consisted of 705 Whites (72%), 127 Blacks (13%), 58 Other (6%), 49 Hispanic (5%), and 40 Asians (4%). Thus, the numbers correspond fairly well to the US Census with the exception that Hispanics are under-represented.

Results

Items with negative loadings were reverse-coded for statistical reasons (some **mirt** functions were not compatible with negative loadings).

Commented [8]: which ones? did they influence your choice of analyses or results?

Political ideology

The 41 political questions were scored using a single factor model based on item response theory (IRT). This was done using the **mirt** package (Chalmers et al., 2020). Each item was modeled using the "graded" model for ordinal data. Most items loaded strongly on the general factor, which explained 42% of the variance. The median absolute factor loading was 0.65 and the lowest was 0.11 ("Homosexual behavior is fine when it is private and chaste"). The empirical reliability of this scale was estimated at 0.97, Cronbach's alpha was 0.96. Unidimensionality was assessed with Revelle's rho (congeneric), which was 0.95 (Revelle & Condon, 2025). This metric indicates the degree to which the item correlation matrix can be predicted from the model. For simulated data with a true unidimensional model, this value was 0.95-1.00 in their simulations (the average factor loading does not matter), while it was below 0.70 when multiple large factors were present. For real life personality scales (from the big five and ICAR), rho was between 0.96-1.00, while for mixed personality items (big five all scales), the value was 0.69-0.79. Thus, while our collection of political questions were not strictly unidimensional, they were relatively close to it. Factor scores from the model were saved for analysis. The resulting leftism scores were normally distributed (Figure S1). As a validation, we compared the scores to answers to the question "What is your party registration?", which showed that the gap between registered Democrats and Republicans was 1.75 d (this question was not used to generate the scores). In other words, 4-way party registration (Republican, Democrat, Independent, Other) could account for 33% of the variance in leftism.

Commented [9]: This section is confusing. which value was .95? Your data or the Revelle prototype? What was the value for your data? you give a set of different values based on ? their (Revelle's?) simulations and then indicate your "collection of political questions were not strictly unidimensional" but if your value was .95 then it would seem to be quite unidimensional per "their simulations."

Make this clear.

Commented [10]: 33 or 43%? chatGPT tells me the latter.

Mental health

Given the heterogeneity of items, we scored them in multiple scales to assess how this would affect the results. Items were modeled using the appropriate item model (2PL for dichotomous items, graded for ordinal items). The scales were:

1. All items (76 items), rho = 0.91, r_{xx} = 0.97
2. Self-related symptoms (58 items), rho = 0.93, r_{xx} = 0.97
3. MMPI-2 items (24 items), rho = 0.97, r_{xx} = 0.85
4. Last 2 weeks (23 items), rho = 0.98, r_{xx} = 0.92
5. Life satisfaction (11 items), rho = 1.00, r_{xx} = 0.96
6. Diagnoses (18 items), rho = 0.89, r_{xx} = 0.60

Unidimensionality (rho) was relatively high for all scales, and higher for those with relatively similar items, as expected. The lowest unidimensionality was for the 18 diagnoses, which is not surprising considering their heterogeneity and known complex factor structure (Kotov et al., 2021). Reliabilities[2] for the self-rated scales were relatively high, while for diagnoses it was relatively poor. This is because 59% of subjects had 0 diagnoses and thus received the minimum score on the construct (13% had 1, 11% had 2, 7% had 3, 5% had 4, and 4% had 5+). In this case, the internal reliability is a poor estimate of the retest reliability because if the subjects were asked again, they would very likely answer very or entirely similarly, demonstrating the high retest reliability. Each of the scales were scored. Scores on each scale were conceptualized as an imperfect measure of the general factor of psychopathology, p (Kotov et al., 2021; Watts et al., 2023). The self-report p gap between those without and with any diagnosis was 0.91 d (Pearson r = 0.47). The symptoms p score increased monotonically as a function of the number of diagnoses: 1.69 d between those with 0 and those with 5+, shown in Figure 1. The distribution of diagnoses in the sample are displayed in Figure 2.

Commented [11]: isn't there a reliability for dichotomous items that corrects for zero-inflation?

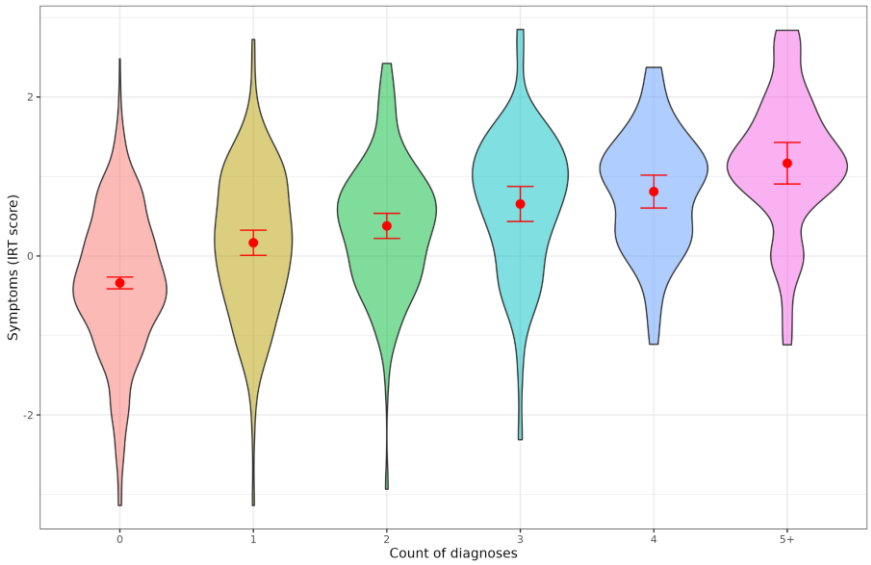

Figure 1. Psychopathology self-related symptoms (p factor scores based on IRT) as a function of the number of clinical diagnoses.

Commented [12]: clarify man this is confusing. these are separate scales?, should be labeled that way (e.g., MMPI items); you are conceptualizing each scale set as an imperfect measure of p, but i wouldn't call them p factor scores here. Just label the scale per your description in the methods.

OK, i see now this is the 'self-reported items' scale so label it as that 'scale' (#2 above) because you have other scales like MMPI going on; why don't you produce this for that scale if it too is an imperfect measure of p per your description.

[2] Reliabilities were computed using *empirical_rxx()* from **mirt** package. This is a kind of internal reliability estimate based on a single-factor model.

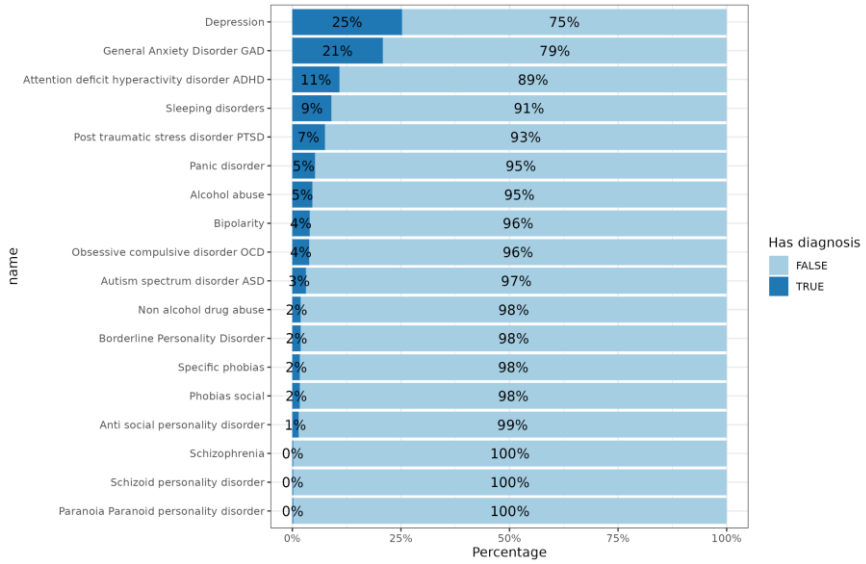

Figure 2. Distribution of diagnoses in the sample. Values rounded to the nearest percentage point.

The last 3 diagnoses were only seen in 2 individuals, hence 0.2% of the dataset. Despite the rarity of some diagnoses, 152 of the 153 correlations among the diagnoses were positive (mean latent correlation = 0.39), see Figure S2. Thus, we replicated the positive manifold observed in prior research.

Correlations among primary variables

Figure 3 shows the main correlations between variables. Latent correlations were applied when necessary using *hetcor()* from the **polycor** package (e.g., tetrachoric for binary pairs).

Commented [13]: of saved factor scores? or was this modeled jointly in latent space somehow?

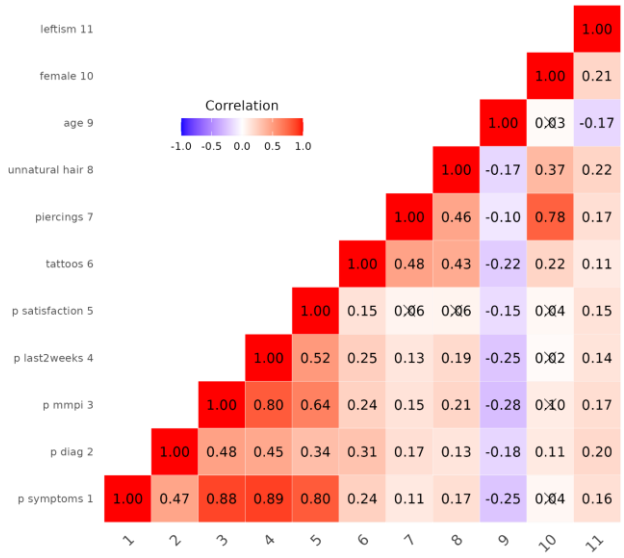

Figure 3. Correlation matrix between main variables in the study. Latent correlations used when necessary. Crossed out values are p > .01.

It can be observed that leftism correlates positively with each variant of the p factor (variables 1-5), as well as with tattoos, piercings, unnatural hair color and being female. On the other hand, leftism correlates negatively with age. Similarly, p factor variants correlate negatively with age (younger people higher in p), and slightly positively with being female. As such, it is likely that correlations between p factor and leftism are confounded by age and sex. The three variables concerning body modifications were fairly strongly correlated, at around 0.45. One notable observation is the very high correlation between piercings and being female, mainly due to relatively common piercings in women: ears r = 0.77, navel 0.55, non-septum nose 0.50, and nipple 0.46. More than 80% of women had an ear piercing (earring).

Commented [14]: You need to not use 'p-factor' here. Just use 'various mental health scales' or the p-factor variants. You conceptually make the case the they are p-factor variants, not an empirical one.

Regressions

While causality cannot be cleanly determined with the present dataset, certain causal models can be ruled out, and effect sizes can be estimated based on assumed causal models. In the first case, it is assumed that leftism is caused by psychopathology along with covariates (in R notation: leftism ~ p + [controls]). Table 1 shows the regression results.

Commented [15]: rephrase as psychopathology causes leftism. It will make the 2nd model more comprehensible below.

Predictor/Model	1	2	3	4	5	6
Intercept	0.00 (0.031, 1)	-0.01 (0.142, 0.969)	-0.01 (0.143, 0.941)	0.00 (0.143, 0.99)	0.01 (0.143, 0.931)	0.00 (0.142, 0.977)
p_{diag}	0.15 (0.036, <0.001***)	0.16 (0.032, <0.001***)				0.13 (0.036, <0.001***)

p_{mmpi}	0.08 (0.058, 0.166)		0.12 (0.032, <0.001***)			0.03 (0.059, 0.665)
$p_{last2weeks}$	-0.02 (0.052, 0.686)			0.10 (0.032, 0.001**)		-0.01 (0.052, 0.892)
$p_{satisfaction}$	0.06 (0.041, 0.16)				0.12 (0.031, <0.001***)	0.07 (0.040, 0.094)
age		-0.14 (0.032, <0.001***)	-0.14 (0.033, <0.001***)	-0.14 (0.033, <0.001***)	-0.15 (0.032, <0.001***)	-0.13 (0.033, <0.001***)
sex = Male		-0.32 (0.062, <0.001***)	-0.33 (0.062, <0.001***)	-0.34 (0.062, <0.001***)	-0.34 (0.062, <0.001***)	-0.32 (0.062, <0.001***)
US born		0.19 (0.137, 0.175)	0.21 (0.138, 0.13)	0.21 (0.138, 0.136)	0.19 (0.138, 0.164)	0.19 (0.137, 0.176)
race		(yes)	(yes)	(yes)	(yes)	(yes)
R2 adj.	0.045	0.083	0.074	0.07	0.074	0.086
N	978	973	973	973	973	973

Table 1. Regression models predicting leftism score. Standardized betas. Numbers in parentheses are standard errors and p values. * = p < .01, ** = p < 0.005, *** = p < .001.

The table shows many results of interest. In model 1, all the p factor variants are included together to gauge relative predictive validity. However, only $p_{diagnoses}$ retains validity (p < .001). In models 2-5, models add demographic controls, showing that each variant of p retains validity, though reduced compared to the correlations in Figure 3 by about 25%. Finally, model 6 repeats model 1 but with demographic controls. Here, once again, $p_{diagnoses}$ predicts leftism (p < .001) whereas the other variants do not.

Commented [16]: significance, not validity is the word you want here.

Commented [17]: predictive significance, not validity

Alternatively, one might posit that leftism causes psychopathology instead. Since $p_{diagnoses}$ appeared to be the key variable, this was used as the outcome variable. Table 2 shows regression models based on this assumption.

Predictor/Model	**1**	**2**
Intercept	0.66 (0.107***)	0.28 (0.098**)
leftism	0.16 (0.032***)	0.10 (0.028***)
age	-0.01 (0.002***)	0.00 (0.002)
male	-0.11 (0.062)	-0.07 (0.056)
race	(yes)	(yes)
p mmpi		0.29 (0.051***)
p last2weeks		0.17 (0.046***)
p satisfaction		0.04 (0.035)

R2 adj.	0.087	0.281
N	976	976

Table 2. Regression models predicting diagnosis-based psychopathology score. Standardized betas. Numbers in parentheses are standard errors and p values. * = p < .01, ** = p < 0.005, *** = p < .001.

The results show that net of demographic controls, leftism still predicts having more diagnoses (beta = 0.16), and this is also true even controlling for self-reported symptom scales (beta = 0.10). These results are thus consistent with bidirectional effects.

Commented [18]: You need to specify in a sentence what you mean by bidirectional effects. I know what you mean, but spell it out for the reader.

Measurement invariance

One chief worry based on prior research is whether subjects high and low in leftism answer mental health questions similarly. In other words, there is a question of whether measurement invariance holds and to what extent violations of this affect the observed relationships. For this reason, we used differential item functioning (DIF) testing on the mental health scales. We used the leave-one-out approach, whereby the other items are used as anchors and the target item is tested for intercept/threshold and slope differences (Chalmers, 2015a, 2015b). Thus, for each item, a p value was calculated for whether the item differs in the parameters, and these p values were then adjusted for multiple testing using the bonferroni method. Finally, to calculate the practical importance of the detected deviations from measurement invariance, an effect size was calculated (Meade, 2010; Nye & Drasgow, 2011). In this approach, the scores on the scale are calculated using the joint invariant fit and again for the partially invariant fit, and the differences in the score gaps are compared to produce a Cohen's d effect of the deviation from measurement invariance. Table 3 gives the results.

Commented [19]: how did you create the groups here for DIFF. you need to spell it out.

Commented [20]: compared to the average of the anchor items? More detail needed

Scale	d	Bias strict	Bias lenient	Biased items strict	Biased items lenient	Item count
All items	0.29	0.02	0.02	4 vs. 4	16 vs. 17	76
Non-diagnoses	0.29	0.01	0.01	3 vs. 4	15 vs. 14	58
Diagnoses	0.34	0	0	0 vs. 0	0 vs. 1	18
MMPI	0.3	-0.02	0.01	0 vs. 3	3 vs. 4	24
Last 2 weeks	0.2	0.01	0.02	0 vs. 0	1 vs. 2	23
Satisfaction	0.34	0.01	0.02	0 vs. 0	1 vs. 0	11

Commented [21]: you need to provide detail on the different variants of bias here.

Table 3. Effect sizes of deviations from measurement invariance across scales. Positive values means higher scores for the low-leftism group.

Many items were found to have some bias (in the combined model, 33/76 items showed bias at p < .05), but overall, the impact of the biased items on the scale scores was relatively minor because the biases were small and counter-balanced. Overall biases

tended to be in the 'wrong' direction, as they generally had the effect of making the low-leftism group attain slightly higher scores on the scale, thus making the observed gaps corrected for bias slightly larger (e.g., d = 0.29 for the all items naive gap, d = 0.31 for the corrected gap).

Generalization across items

It is also possible to investigate whether the psychopathology items with stronger factor loadings showed larger gaps between the political groups (Jensen's Method of Correlated Vector, MCV; explained in Jensen, 1998, Appendix B). If the variation among item associations is caused by differences in the latent trait of interest, then items with stronger factor loadings should relate more strongly to the criterion variable, in this case, leftism. Though it should be noted that sampling error in item parameter estimates will cause a bias towards 0 (Dutton & Kirkegaard, 2022a). Given the large effects of age and sex on leftism (cf. Table 1), it seems important to control for this confounding. Thus two approaches were used. First, the latent correlation between each item and leftism was calculated (standard Jensen method). Second, to control for demographics, two linear models were fit to predict leftism as a function of the demographics alone and demographics plus the index item (expanded Jensen method). The incremental validity of the item was computed from the models in multiple R units. Figure 4 shows the results.

Commented [22]: this comes out of nowhere. factor loadings need to be defined man; it's a different model than IRT etc.

Commented [23]: need detail here. latent correlation between each item and leftism? Did you model each scale as a latent variable, and then correlate item loadings with a latent variable of leftism?

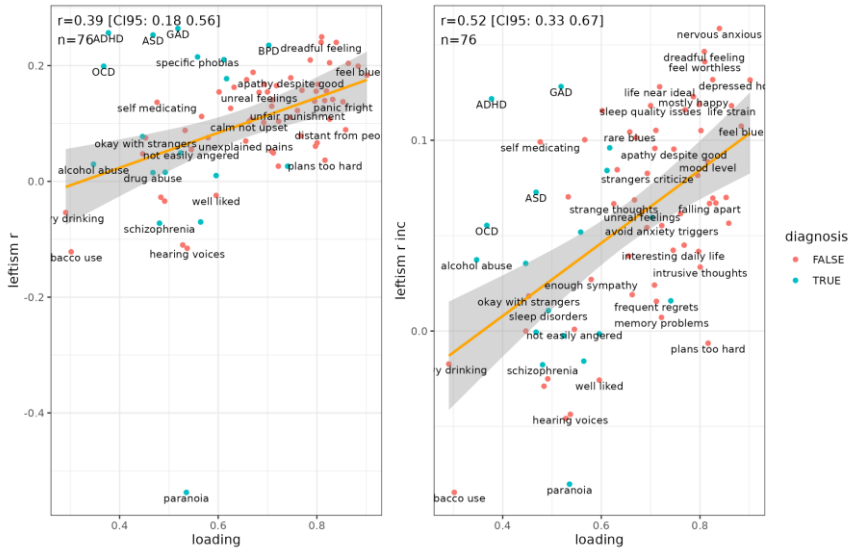

Figure 4. Jensen's method applied to the psychopathology items for political ideology. The left plot shows the latent correlations between each item and leftism, and the second plot shows incremental validity of each item net of demographics.

It can be observed that items with stronger loadings on the p factor showed more robust associations with leftism. This was true without (r = 0.39) and with controls (r = 0.52),

Commented [24]: clarify bro. For each scale of interest? You didn't model a p-factor. You conceptualized each scale as a p-factor

though it was stronger after controls. This comparison suggests that the pattern was distorted by the demographic variables, not enhanced by them (paired correlation test, p = .01). The plots also reveal that diagnoses show stronger associations with leftism than would be expected from their factor loadings in the common item-pool. Table 4 shows the item-level regressions.

Commented [25]: common item pool?

Outcome:	*Correlations*			*Incremental R*		
Predictor	1	2	3	1	2	3
(Intercept)	-0.10 (0.056, 0.086)	-0.13 (0.068, 0.069)	-0.17 (0.072, 0.02)	-0.07 (0.024, 0.006*)	-0.08 (0.030, 0.011)	-0.11 (0.031, <0.001***)
loading	0.30 (0.084, <0.001***)	0.34 (0.096, <0.001***)	0.40 (0.103, <0.001***)	0.19 (0.036, <0.001***)	0.20 (0.042, <0.001***)	0.25 (0.044, <0.001***)
diagnosis		0.02 (0.034, 0.469)	0.28 (0.150, 0.069)		0.01 (0.015, 0.589)	0.17 (0.064, 0.01)
loading:diagnosis			-0.46 (0.267, 0.088)			-0.29 (0.113, 0.012)
R2 adj.	0.138	0.132	0.155	0.261	0.254	0.307
N	76	76	76	76	76	76

Table 4. Item-level regressions using Jensen's method. Unstandardized slopes.

The regression models showed that a hypothetical item with loading = 1 would show a leftism gap of 0.21 d (0.34-0.13) and 0.12 d (0.20-0.08) for the models without interactions. The interactions between being a diagnosis and factor loading were borderline significant (p = 0.09 and p = 0.01), suggesting that, in the pooled-item set, loadings among diagnosis items were not related to items' association strength with leftism.

Alternatively, one might approach the question from the other perspective and ask: does it matter how leftism is measured? Here, the analysis is reversed such that factor loadings for political ideology items are used instead and the outcome variable is psychopathology. Figure 5 shows the results.

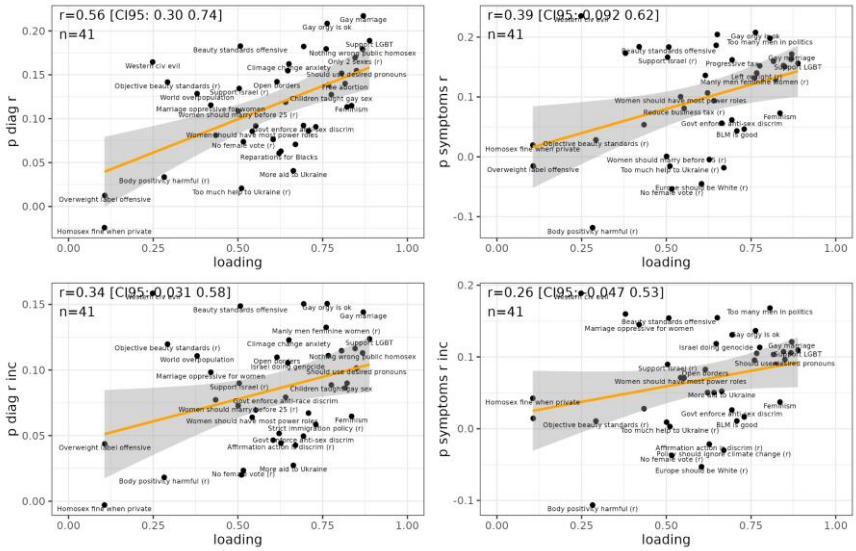

Figure 5. Jensen's method applied to the political ideology items for psychopathology. The top plots show the latent correlations (each item's association with psychopathology, p), the bottom plots show their incremental validity net of demographics. (r) indicates the item was reversed.

Results show a consistent pattern. The more a given question measures leftism (or negatively rightism), the more it is associated with psychopathology. This was true whether this was measured using diagnoses or symptoms (though the correlation had p > .05 for incremental validity for the symptoms scale). Table 5 shows the political items most and least correlated with psychopathology.

Item	Factor loading	mean r with psychopathology
Body positivity harmful (r)	0.28	-0.04
No female vote (r)	0.52	0
Europe should be White (r)	0.6	0.01
Too much help to Ukraine (r)	0.51	0.01
Homosex fine when private	0.11	0.01
Climate change anxiety	0.65	0.16
Gay marriage	0.87	0.16

Beauty standards offensive	0.51	0.17
Gay orgy is ok	0.76	0.18
Western civ evil	0.25	0.19

Table 5. Political ideology items most and least associated with psychopathology. Factor loadings are for the political factor of leftism. (r) indicates the item was reversed.

Body modifications and psychology

In Figure 3, it was shown that piercings, tattoos, and unnatural hair colors are associated with both leftism and psychopathology. This suggests that perhaps these features would be useful visual cues to judging the psychology of a person at distance (stereotyping). However, given that they also correlated with age and sex, which are usually easier to observe, this might fail if they do not have incremental validity. For this reason, we sought to see whether they had incremental validity to predict leftism and psychopathology beyond demographic variables. Table 6 shows the results.

Predictor	**Outcome**					
	Leftism			*p diagnoses*		
Intercept	-0.15 (0.042, <0.001***)	0.60 (0.125, <0.001***)	0.52 (0.124, <0.001***)	-0.18 (0.041, <0.001***)	0.46 (0.122, <0.001***)	0.37 (0.122, 0.002**)
tattoo count	0.04 (0.030, 0.241)	0.03 (0.030, 0.391)	-0.01 (0.031, 0.852)	0.23 (0.030, <0.001***)	0.20 (0.029, <0.001***)	0.20 (0.029, <0.001***)
piercing count	0.09 (0.043, 0.03)	-0.01 (0.050, 0.767)	-0.02 (0.049, 0.747)	0.03 (0.042, 0.436)	0.01 (0.048, 0.88)	0.01 (0.048, 0.843)
unnatural hair	0.29 (0.080, <0.001***)	0.24 (0.079, 0.003**)	0.23 (0.078, 0.003**)	0.05 (0.078, 0.528)	0.03 (0.077, 0.67)	0.00 (0.076, 0.976)
age		-0.01 (0.002, <0.001***)	-0.01 (0.002, <0.001***)		-0.01 (0.002, <0.001***)	-0.01 (0.002, <0.001***)
male		-0.31 (0.075, <0.001***)	-0.30 (0.074, <0.001***)		-0.07 (0.073, 0.317)	-0.03 (0.073, 0.702)
race		(yes)	(yes)		(yes)	(yes)
p_diag			0.15 (0.033, <0.001***)			
leftism						0.15 (0.031, <0.001***)
R2 adj.	0.031	0.068	0.089	0.079	0.117	0.136
N	978	976	976	978	976	976

Table 6. Visual cues to leftism and psychopathology from piercings, tattoos, and unnatural hair color. Unstandardized betas. Numbers in parentheses are standard errors and p values. * = p < .01, ** = p < 0.005, *** = p < .001.

Interesting divergent validities were observed. For predicting leftism, unnatural hair color (0.24 z) in addition to being younger (-0.01) and female (0.30) were significant. And these associations hold when controlling for p_{diag}. For predicting p_{diag}, tattoos (each tattoo adds 0.20 z) as well as being younger (-0.01) were significant.

Commented [26]: divergent validities? No. You want different predictive significance; patterns of significant associations. Very confusing.

In addition to the whole body counts of tattoos and piercings, one might wonder if various types of body modifications have different validities. To examine this, two further regression models were fit using each specific type of tattoo and piercing as a predictor. The results were too large to be presented in full here, but the predictors with beyond chance level results ($p < .01$) were as follows. For predicting leftism, we have septum piercing (1.02), genital piercing (1.37), unnatural hair (0.23). For predicting p_{diag}, we have arm tattoo (0.41), other/uncommon tattoo (0.63), and piercing outside the mouth area (1.08). Due to the rarity of some specific tattoos and piercings, a larger sample size would no doubt find more useful cues. However, these cues are quite large in effect size (e.g., a person with a septum and genital piercing with unnatural hair color would have an expected leftism of 2.62 z above the mean).

Commented [27]: supplemental table please.

Differential validity by diagnosis

It is possible that different aspects of diagnosed psychopathology predict political orientation to different degrees. To examine this question, models were fit for each diagnosis alone plus the demographic controls (16 models), as well as a joint model. Moreover, the 3 diagnoses concerning schizoid-like problems were merged given their very low case counts and similarity (each $n_{cases} = 2$, combined $n_{cases} = 4$). Figure 6 shows the results.

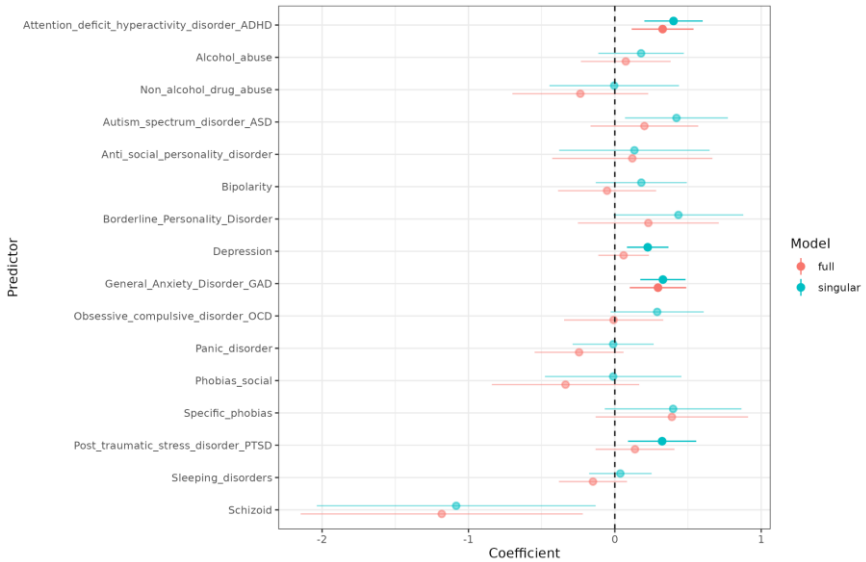

Figure 6. Regression models predicting leftism by diagnosis. Control variables are age, sex, and race. Highlighted coefficients have p < .05 false discovery rate adjusted p values. 95% confidence intervals.

It can be observed that while many diagnoses predicted politics at the 5% level when controlling for demographics (ADHD, ASD, DEP, GAD, PTSD, Schizoid-like), only 2 predicted leftism when other diagnoses were included and multiple testing adjusted p-values were used: ADHD and GAD. Given the case counts for many of these diagnoses, this analysis probably had relatively low power, however. This is especially the case for the schizoid-like diagnosis (n=4 cases).

Discussion

The present study was able to confirm many prior findings from the literature. We found that measures of mental health related to political ideology, such that worse mental health was related to left-wing views. This is in line with most research on the topic (Bernardi, 2020; Dutton & Kirkegaard, 2022b). The better a given question measured mental health, the more strongly it associated with leftism. Similarly, the better a given question measured leftism, the more it is associated with psychopathology. We also observed that mental health diagnoses were stronger predictors of leftism than self-reported symptoms. In fact, when controlling for diagnoses, symptoms no longer predicted leftism. The results are open to multiple interpretations. In our favored interpretation, other-report measures are better measures in general and show stronger predictive validities and heritabilities across different traits (Connelly & Ones, 2010; Faraone & Larsson, 2019; Kendler et al., 1993; Riemann et al., 1997; Riemann & Kandler, 2010; Willems et al., 2019). Thus, it is expected that diagnoses are more valid predictors of leftism because they are simply better measures of psychopathology. Another interpretation is that people with left-wing views are more favorable to psychiatry and may value getting a diagnosis for a variety of reasons such as sympathy, economic benefits, or Machiavellian intersectionality points. If so, left-wing individuals may accumulate more diagnoses than their right-wing counterparts, not due to greater psychopathology, but because of differential help-seeking or diagnostic bias. The present study could not rule out this interpretation based on the available data. A third option is that the measurement scales are psychometrically biased such that leftists are more likely to rate themselves higher. Our measurement invariance analyses, however, did not find much bias, and the small biases pointed in the opposite direction (the scales were slightly biased in favor of leftists). This relative lack of bias in scales across political groups confirms prior research.

In the analysis of differential validity of the different diagnosis for predicting leftism, it was found that anxiety and ADHD were the most reliable predictors. The first finding mirrors a previous study which found that anxiety symptoms were a stronger predictor of leftism than depression symptoms (Dutton & Kirkegaard, 2024). There was a possible signal of schizoid-like disorder predicting rightism, but because there were only 4 cases of this, this is only a suggestive finding. If true, it may confirm patterns that in the current American political climate, conservatives tend to attract people prone to conspiracy thinking, which is common among schizophrenics (Enders et al., 2022).

Commented [28]: reword this language.

Commented [29]: predictive significance

Commented [30]: The Enders abstract doesn't seem to make this case re: conservatives; it reads as if it depends on the particular theory in question (not generally)

We found that body modifications predict both leftism and psychopathology, though the effects depended on the specific body modification (hair color, piercing location, tattoo location) and which psychological trait was predicted. Prior research similarly found a substantial correlation between body modification such as tattoos/piercing and various psychopathological measures such as self-injurious behaviors, use of substances, negative quality of life, reduced social interaction, emotional distress (D'Ambrosio et al., 2013; Deschesnes et al., 2006; King & Vidourek, 2013; Koch et al., 2010; Stirn et al., 2006). Too little research investigated the relationship between political orientation and body modification. One study found that Democrats (left-leaning) are more likely than Republicans (right-leaning) to ever had a tattoo, a body piercing, or both (Laumann & Derick 2006, Table 1), whereas another study found no difference between Democrats and Republicans in the likelihood of having a tattoo (Schaeffer & Dinesh, 2023). Both of these results however did not adjust for potential confounders such as demographics. More research is still needed on this matter.

Commented [31]: reword this

Limitations

This study has a number of limitations. First, being a survey study based on paid subjects, it has the same sampling bias as other such research. Since the hourly wage of survey subjects is close to the minimum wage in the USA, this presumably results in selection bias for who spends their time participating. We are not aware of any particular problems this would cause for the present study, but neither can they be ruled out.

Second, our measures of political ideology likely include different ideological dimensions, such as leftism (e.g., progressive taxation), sociocultural leftism (e.g., affirmative action, overpopulation as an environmental issue), geopolitical ideology or even cultural alignment (e.g., supporting Israel versus Hamas). Although such a criticism is on point, the unidimensionality estimate based on Revelle's rho is high (0.95). This still provides a strong rationale for the use of the general factor. Although we retain the general factor for parsimony, we caution that it can mask nuanced ideological differences.

Third, since we relied mainly on self-report, the research is beset with the problems common to self-report measurements (such as over-reporting bias due to awareness or stigma or errors due to mental instability). As we discussed above, self-report measurements tend to be worse than other-report or objectively scored measurements (cognitive tests).

Supplementary materials

Data, figures, R notebook are available at https://osf.io/fh685/, and the R notebook is furthermore available at https://rpubs.com/EmilOWK/politics_p_physiognomy_2025.

References

Atkinson, M. (2003). *Tattooed: The sociogenesis of a body art*. Toronto:

University of Toronto Press.

Bakker, B. N., Lelkes, Y., & Malka, A. (2021). Reconsidering the Link Between

Self-Reported Personality Traits and Political Preferences. *American Political*

Science Review, *115*(4), 1482–1498. https://doi.org/10.1017/S0003055421000605

Bernardi, L. (2020). Depression and political predispositions: Almost blue? *Party Politics*, 1354068820930391. https://doi.org/10.1177/1354068820930391

Brandt, M. J., He, J., & Bender, M. (2021). Registered Report: Testing Ideological Asymmetries in Measurement Invariance. *Assessment*, *28*(3), 687–708. https://doi.org/10.1177/1073191120983891

Chalmers, P. (2015a, February 10). *Multidimensional Item Response Theory Workshop in R*. https://philchalmers.github.io/mirt/extra/mirt-Workshop-2015_Day-1.pdf

Chalmers, P. (2015b, February 11). *Multidimensional Item Response Theory Workshop in R (day 2)*. https://philchalmers.github.io/mirt/extra/mirt-Workshop-2015_Day-2.pdf

Chalmers, P., Pritikin, J., Robitzsch, A., Zoltak, M., Kim, K., Falk, C. F., Meade, A., Schneider, L., King, D., Liu, C.-W., & Oguzhan, O. (2020). *mirt: Multidimensional Item Response Theory* (Version 1.32.1) [Computer software]. https://CRAN.R-project.org/package=mirt

Connelly, B. S., & Ones, D. S. (2010). An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin*, *136*(6), 1092–1122. https://doi.org/10.1037/a0021212

D'Ambrosio, A., Casillo, N., & Martini, V. (2013). Piercings and tattoos: psychopathological aspects. *Activitas Nervosa Superior Rediviva, 55*(4), 143–148.

Deschesnes, M., Demers, S., & Finès, P. (2006). Prevalence and characteristics of body piercing and tattooing among high school students. *Canadian Journal of Public Health, 97*, 325–329. https://doi.org/10.1007/BF03405614

de Vries, R. E., Wesseldijk, L. W., Karinen, A. K., Jern, P., & Tybur, J. M. (2022). Relations between HEXACO personality and ideology variables are mostly

genetic in nature. *European Journal of Personality*, *36*(2), 200–217. https://doi.org/10.1177/08902070211014035

Dunkel, C. S., van der Linden, D., Kawamoto, T., & Oshio, A. (2021). The General Factor of Personality as Ego-Resiliency. *Frontiers in Psychology*, *12*, 5074. https://doi.org/10.3389/fpsyg.2021.741462

Dutton, E., & Kirkegaard, E. (2022a). The Negative Religiousness-IQ Nexus is a Jensen Effect on Individual-Level Data: A Refutation of Dutton et al.'s 'The Myth of the Stupid Believer.' *Journal of Religion and Health*, *61*(4), 3253–3275. https://doi.org/10.1007/s10943-021-01351-1

Dutton, E., & Kirkegaard, E. O. W. (2022b). *Blue Hair and the Blues: Dying Your Hair Unnatural Colours is Associated with Depression*. *6*(2). https://www.pjp.psychreg.org/wp-content/uploads/2022/06/3-edward-dutton-blue_hair_blues_26-35.pdf

Dutton, E., & Kirkegaard, E. (2024). Do conservatives really have an advantage in mental health? An examination of measurement invariance. *Scandinavian Journal of Psychology*, *66*. https://doi.org/10.1111/sjop.13065

Enders, A., Farhart, C., Miller, J., Uscinski, J., Saunders, K., & Drochon, H. (2022). Are Republicans and Conservatives More Likely to Believe Conspiracy Theories? *Political Behavior*. https://doi.org/10.1007/s11109-022-09812-3

Faraone, S. V., & Larsson, H. (2019). Genetics of attention deficit hyperactivity disorder. *Molecular Psychiatry*, *24*(4), 562. https://doi.org/10.1038/s41380-018-0070-0

Jensen, A. R. (1998). *The g factor: The science of mental ability*. Praeger.

Jost, J. T. (2021). *Left and right: The psychological significance of a political distinction*. Oxford University Press.

Jost, J. T., Glaser, J., Kruglanski, A. W., & Sulloway, F. J. (2003). Political conservatism as motivated social cognition. *Psychological Bulletin, 129*(3), 339–375. https://doi.org/10.1037/0033-2909.129.3.339

Kendler, K. S., Neale, M. C., Kessler, R. C., Heath, A. C., & Eaves, L. J. (1993). The Lifetime History of Major Depression in Women: Reliability of Diagnosis and Heritability. *Archives of General Psychiatry*, *50*(11), 863–870. https://doi.org/10.1001/archpsyc.1993.01820230054003

Khosla, V., Joseph, V., & Gordon, H. (2010). Tattoos: What is their significance? *Advances in Psychiatric Treatment*, *16*(4), 281–287. https://doi.org/10.1192/apt.bp.108.006205

King, K. A., & Vidourek, R. A. (2013). Getting inked: Tattoo and risky behavioral involvement among university students. *The Social Science Journal, 50*(4), 540–546.

Kirkegaard, E. O. W. (2020). Mental Illness and the Left. *Mankind Quarterly*, *60*(4). http://mankindquarterly.org/upcoming/issue/60-4/3

Kirkegaard, E. O. W. (2023, January 26). Tattoos and piercings, what do they tell you about someone? *Just Emil Kirkegaard Things*. https://www.emilkirkegaard.com/p/tattoos-and-piercings-what-do-they

Kirkegaard, E. O. W., & Nyborg, H. (2021). Intelligence and General Psychopathology in the Vietnam Experience Study: A Closer Look. *Mankind Quarterly*, *61*(4). https://doi.org/10.46469/mq.2021.61.4.2

Koch, J. R., Roberts, A. E., Armstrong, M. L., & Owen, D. C. (2010). Body art, deviance, and American college students. *The Social Science Journal, 47*(1), 151–161. https://doi.org/10.1016/j.soscij.2009.10.001

Kosut, M. (2006). An Ironic Fad: The Commodification and Consumption of Tattoos. *Journal of Popular Culture, 39*(6), 1035–48. https://doi.org/10.1111/j.1540-5931.2006.00333.x

Kotov, R., Krueger, R. F., Watson, D., Cicero, D. C., Conway, C. C., DeYoung, C. G., Eaton, N. R., Forbes, M. K., Hallquist, M. N., Latzman, R. D., Mullins-Sweatt, S. N., Ruggero, C. J., Simms, L. J., Waldman, I. D., Waszczuk, M. A., & Wright, A. G. C. (2021). The Hierarchical Taxonomy of Psychopathology (HiTOP): A

Quantitative Nosology Based on Consensus of Evidence. *Annual Review of Clinical Psychology*, *17*(Volume 17, 2021), 83–108. https://doi.org/10.1146/annurev-clinpsy-081219-093304

Laumann, A., & Derick, A. (2006). Tattoos and body piercings in the United States: A national data set. *Journal of the American Academy of Dermatology, 55*(3), 413–421. https://doi.org/10.1016/j.jaad.2006.03.026

McCandlish, C., & Pearson, M. (2023). Tattoos as symbols–an exploration of the relationship between tattoos and mental health. *The Journal of Mental Health Training, Education and Practice, 18*(3), 217–227.

Meade, A. W. (2010). A taxonomy of effect size measures for the differential functioning of items and scales. *The Journal of Applied Psychology*, *95*(4), 728–743. https://doi.org/10.1037/a0018966

Mortensen, K., French, M. T., & Timming, A. R. (2019). Are tattoos associated with negative health-related outcomes and risky behaviors?. *International Journal of Dermatology, 58*(7), 816–824. https://doi.org/10.1111/ijd.14372

Nye, C. D., & Drasgow, F. (2011). Effect size indices for analyses of measurement equivalence: Understanding the practical importance of differences between groups. *The Journal of Applied Psychology*, *96*(5), 966–980. https://doi.org/10.1037/a0022955

Perrotta, G. (2021) Massive use of tattoos and psychopathological clinical evidence. *Arch Community Med Public Health, 7*(2), 079–085. https://doi.org/10.17352/2455-5479.000144

Revelle, W., & Condon, D. (2025). Unidim: An index of scale homogeneity and unidimensionality. *Psychological Methods*. https://doi.org/10.1037/met0000729

Riemann, R., Angleitner, A., & Strelau, J. (1997). Genetic and Environmental Influences on Personality: A Study of Twins Reared Together Using the Self- and Peer Report NEO-FFI Scales. *Journal of Personality*, *65*(3), 449–475. https://doi.org/10.1111/j.1467-6494.1997.tb00324.x

Riemann, R., & Kandler, C. (2010). Construct validation using multitrait-multimethod-twin data: The case of a general factor of personality. *European Journal of Personality*, *24*(3), 258–277. https://doi.org/10.1002/per.760

Schaeffer, K., & Dinesh, S. (2023). 32% of Americans have a tattoo, including 22% who have more than one. *Pew Research Center*. Retrieved at: https://www.pewresearch.org/short-reads/2023/08/15/32-of-americans-have-a-tattoo-including-22-who-have-more-than-one/

Stirn, A., Hinz, A., & Brähler, E. (2006). Prevalence of tattooing and body piercing in Germany and perception of health, mental disorders, and sensation seeking among tattooed and body-pierced individuals. *Journal of Psychosomatic Research, 60*(5), 531–534. https://doi.org/10.1016/j.jpsychores.2005.09.002

Timming, A. R., Nickson, D., Re, D., & Perrett, D. (2017). What do you think of my ink? Assessing the effects of body art on employment chances. *Human Resource Management, 56*(1), 133–149.

Watts, A. L., Greene, A. L., Bonifay, W., & Fried, E. I. (2023). *A critical evaluation of the p-factor literature*. PsyArXiv. https://doi.org/10.31234/osf.io/7yrnp

Willems, Y. E., Boesen, N., Li, J., Finkenauer, C., & Bartels, M. (2019). The heritability of self-control: A meta-analysis. *Neuroscience & Biobehavioral Reviews*, *100*, 324–334. https://doi.org/10.1016/j.neubiorev.2019.02.012

Appendix

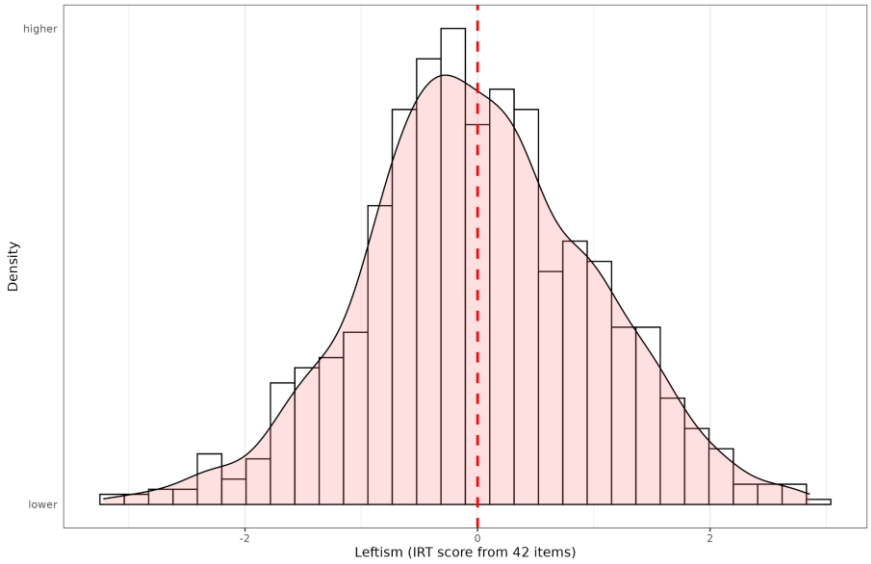

Figure S1. Distribution of political leftism based on 42 items.

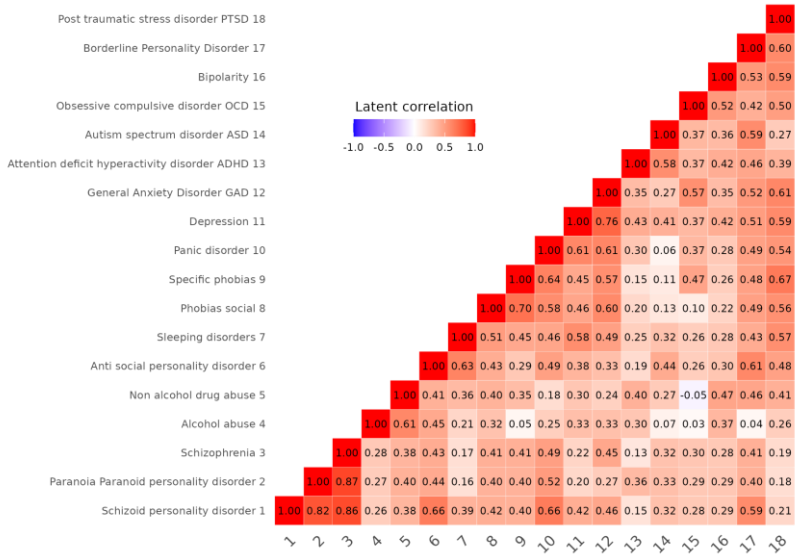

Figure S2. Correlation matrix for mental health diagnoses (latent correlation, i.e., tetrachoric).