



Full Length Article

Perceived liberalism and intellect cannot be reduced to a single openness dimension: A state-trace experiment

Johannes Titz¹*, Benedikt Quirin Gailer¹

Chemnitz University of Technology, Germany

ARTICLE INFO

Dataset link: <https://doi.org/10.5281/zenodo.3083760>

Keywords:

State-trace analysis
Dimensionality
Intellect
Liberalism
Openness to experience
Monotonicity
Personality perception
Big Five

ABSTRACT

Liberalism and intellect are commonly conceptualized as facets of openness to experience. Previous self-report evidence suggests that ratings of these facets may not be representable as monotonic indicators of a single latent dimension. We tested whether analogous nonmonotonicity emerges in personality perception using a vignette-based state-trace experiment. We experimentally manipulated descriptions of fictitious targets' self-efficacy and artistic interest, and participants judged each target's perceived liberalism and intellect. The relationship between these judgments was nonmonotonic in both between- and within-subjects analyses, requiring at least two dimensions to accommodate the observed pattern. These results demonstrate that perceived liberalism and intellect cannot be explained by a single latent dimension and illustrate how state-trace experiments can complement correlational approaches to personality perception.

Are liberals, on average, more intellectual than conservatives? This comparison may itself be too simple. Intellectually oriented conservatives — from traditionalist theologians to conservative philosophers — make it easy to see why. More generally, groups with more pronounced positions at either end of the ideological spectrum may, on average, score higher in intellect than moderate groups. This possibility matters because, in widely used five-factor inventories based on the influential taxonomy of *Costa and McCrae (2008)*, liberalism (or values) and intellect (or ideas) are assigned to the same openness domain and contribute to a single aggregate domain score.¹ If that score is interpreted as measuring a single ordered trait, a U-shaped relation would contradict this unidimensional interpretation: higher liberalism should, on average, be accompanied by higher — not lower — intellect.

This implication has received relatively little attention in personality psychology, in part because dominant approaches to dimensionality — most notably factor analysis based on Pearson correlations — are not well suited to adjudicating such questions. When a contemporary method of dimensionality analysis, state-trace analysis (*Dunn &*

Kalish, 2018), is applied, the one-dimensional interpretation of liberalism and intellect fails: across part of the conservative range, greater conservatism is, on average, associated with higher intellect.

Importantly, the present argument differs from traditional correlational critiques of openness. Factor analysis and state-trace analysis address different questions. Factor analysis examines shared variance among variables, whereas state-trace analysis tests whether variables can be represented as monotonic indicators of a single latent dimension. Accordingly, the present article does not argue that correlational approaches are generally inappropriate, but rather that they are not designed to adjudicate the stronger criterion of monotonic unidimensionality.

More broadly, the adoption of modern methods of dimensionality analysis may provide useful complementary insights into questions of personality structure, paralleling developments that are currently transforming dimensional theories in cognitive psychology (e.g. *Dunn, 2008; Dunn et al., 2012; Dunn & Rao, 2019; Hayes et al., 2018;*

* Correspondence to: Department of Psychology, Research Methods and Evaluation in Psychology, Chemnitz University of Technology, Wilhelm-Raabe-Str. 43, 09120 Chemnitz, Germany

E-mail address: johannes@titz.science (J. Titz).

¹ Although the distinction between openness and intellect is commonly acknowledged, the broad construct is still regularly described as “a dimension”, “the dimension”, or “the openness/intellect dimension” (e.g., *Oleynick et al., 2017*). More directly, widely used inventories calculate a single openness domain score by aggregating responses assigned to different facets. *Costa and McCrae (1995, p. 26)* explicitly report that NEO-PI-R raw domain scores are calculated by summing the six facet scales. The official IPIP scoring instructions likewise direct users to sum all keyed item values within a scale (*International Personality Item Pool, n.d.*). This aggregation may be adopted pragmatically and does not, by itself, assert that the facets are strictly unidimensional or deny facet-specific variance. Nevertheless, insofar as the resulting domain score is interpreted as locating individuals on a single broad trait, it is relevant to ask whether the aggregated responses are compatible with a common monotonic ordering.

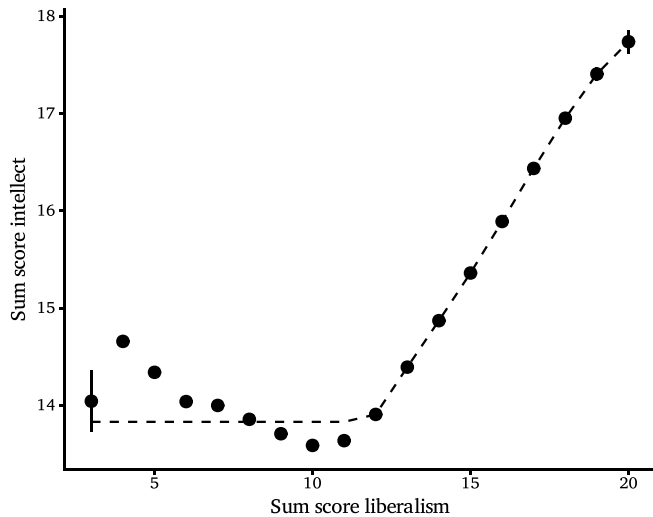


Fig. 1. *Intellect sum score plotted against liberalism sum score.* The source dataset contained 619,150 participants; after excluding liberalism sum scores of 1 (2 participants) and 2 (31 participants) due to insufficient sample size, the plotted sample contained 619,117 participants. Although 95% confidence intervals are shown, they are largely imperceptible (except at liberalism sum scores of 3 and 20) because they are narrower than the plotted data points. The dashed line represents the best-fitting isotonic model and is included as a visual aid. Source: Adapted from Titz (2025a, Figure 6).

© 2025 Johannes Titz, CC BY 4.0.

Heathcote et al., 2019; Jang et al., 2019; Liu & Jia, 2022; Robinson & Irwin, 2019).

Previous work has provided evidence for a multidimensional structure of openness to experience in self-ratings (Titz, 2025a), drawing on ideas from state-trace analysis. The present study asks whether a comparable pattern emerges when observers evaluate *experimentally manipulated* targets and examines how the information provided causally shapes their personality judgments.

The experiment thus concerns personality perception rather than the traits of real individuals. Specifically, observers infer a fictitious target's liberalism and intellect from systematically varied descriptions. Its direct conclusions therefore pertain to the organization of these judgments; implications for personality structure in real people depend on convergence with assessments of actual individuals.

In what follows, we first present strong — though non-experimental — evidence that liberalism and intellect cannot be reduced to a single underlying dimension, while providing an informal introduction to state-trace analysis. We then examine whether the same nonmonotonic pattern appears in experimentally controlled personality perceptions.

1. The problem

Titz (2025a) was among the first to investigate the personality structure of the five-factor model using ideas from state-trace analysis. His analyses yielded a consistent general finding: among all Big Five domains, openness to experience exhibited the greatest number of item pairs violating unidimensionality in the IPIP-NEO-120 (Johnson, 2014), with the most frequent and pronounced violations occurring between the facets liberalism and intellect. When plotted against one another, the relationship between liberalism and intellect was best characterized as U-shaped (see Fig. 1). On average, individuals with the lowest intellect scores exhibited liberalism scores slightly above the scale midpoint. As liberalism deviated from this midpoint in either

direction, intellect increased, with the increase being more pronounced on the liberal side. For liberalism scores below 4, intellect appeared to decline again, although the associated uncertainty was substantially greater.

The key point is that the two facets exhibit a nonmonotonic relationship and therefore cannot be meaningfully represented by a single underlying dimension. Suppose there were a single latent variable, O (openness), driving variation in both facets. As O increases, one would expect, on average, both liberalism and intellect to increase correspondingly. Under this assumption, it should never occur that a group with higher openness scores is, on average, less liberal or less intellectual than a group with lower openness scores. This, in turn, implies a monotonic relationship between the two facets themselves: greater liberalism should coincide with greater intellect, and vice versa. Fig. 1, however, directly contradicts this expectation. A strictly one-dimensional explanation is therefore difficult to reconcile with the observed pattern.

A more rigorous argument proceeds as follows (for detailed discussion, see Dunn & Kalish, 2018). Consider the variables O , l (liberalism), and i (intellect), where both l and i depend solely on O : $l = f(O)$ and $i = g(O)$. Assume further that f and g are strictly monotonically increasing functions, such that higher openness entails higher liberalism and higher intellect. Under these assumptions, l can be expressed as a function of i , namely $l = f(g^{-1}(i))$. Because the inverse of a strictly monotonically increasing function is itself monotonically increasing, it follows that l and i must be monotonically related.

A helpful visualization of the fact that g^{-1} is monotonically increasing is the following: consider any increasing function plotted in a coordinate system, rotate it by 90° clockwise, and then reflect it across the horizontal axis. The resulting function remains monotonically increasing. Formally, this property is captured by the derivative of the inverse function at points where $g'(O) > 0$:

$$\frac{d}{di}(g^{-1}(i)) = \frac{1}{g'(g^{-1}(i))} = \frac{1}{g'(O)} > 0.$$

Thus, we conclude that the pattern shown in Fig. 1 would not arise if liberalism and intellect were truly governed by a single underlying dimension. This finding is supported by a comparable pattern, again statistically significant, observed for the analogous facets values and ideas in a NEO-PI-R (Costa & McCrae, 2008) dataset (Titz, 2025a).²

In discussing his results, Titz (2025a) argued that the observed violations of monotonicity between liberalism and intellect indicate a substantive theoretical issue. Given the large number of incompatible item pairs, this pattern cannot plausibly be explained by a few poorly functioning items. Rather, the findings suggest that the two facets reflect conceptually distinct traits that resist reduction to a single latent dimension. Accordingly, within the examined operationalizations, liberalism and intellect cannot be interpreted as monotonic indicators of a single openness continuum.

The possibility that openness comprises more than one dimension is not novel. Indeed, evidence for a division of openness into two dimensions exists that does not rely on state-trace analysis, and the debate itself is long-standing (e.g., Johnson, 1994; Saucier, 1992). For example, DeYoung et al. (2007) identified two aspects of openness — openness and intellect — that differ in their emphasis on engagement with perceptual or sensory information versus abstract information, respectively (DeYoung et al., 2014). Although these aspects were substantially correlated ($r = .51$; DeYoung et al., 2007), the authors argued that they reflect distinct traits on the basis of discriminant validity: openness and intellect exhibited different patterns of association with the remaining Big Five dimensions (for additional evidence of

² In another NEO-PI-R sample, the effect was not statistically significant, arguably due to the student sample, which naturally included few conservative respondents.

divergence in other areas see, for instance, Kaufman et al., 2016; Vartanian et al., 2018; Willoughby et al., 2023). Such discriminant-validity evidence is informative about differential associations, but it does not by itself determine dimensionality: under monotonic measurement models, functional dissociations and interaction patterns are neither necessary nor sufficient for inferring more than one latent variable (Dunn & Kalish, 2018; Dunn & Kirsner, 1988). What is novel, therefore, is not the proposed distinction between openness and intellect, but its evaluation using a formal test of monotonic unidimensionality.

Regardless of whether one relies on correlational patterns or the stronger argument based on monotonicity, the existing evidence remains non-experimental. Although this does not undermine the dimensionality argument itself, it leaves open whether controlled changes in target information causally produce the same nonmonotonic pattern in observers' judgments. A vignette experiment addresses this complementary question by varying the available target information under controlled conditions and examining the resulting personality impressions. It thereby permits causal conclusions about the effects of the supplied cues on personality judgments.

This gain in experimental control comes at a cost in external validity. We address this trade-off explicitly in the Discussion, but three qualifications concerning the scope of our conclusions should be noted at the outset. First, the dependent variables are observers' personality perceptions elicited by descriptions of fictitious targets, not measurements of actual traits in real individuals. Any implication for actual personality therefore rests on convergence with separate self-rating evidence rather than direct generalization (Titz, 2025a).

Second, because the vignettes provide only limited information about those targets, participants are likely to supplement the supplied cues with implicit personality theories when judging liberalism and intellect (Hakel, 1969; Mirels, 1976). The resulting pattern may therefore partly reflect culturally shared beliefs or stereotypes about liberals and conservatives. Because the political meaning of liberalism and conservatism varies across societies, the German context may limit generalization to other cultural settings.

Third, the vignette descriptions and target judgments are grounded in the IPIP-NEO-120 and therefore reflect its conceptualization of openness to experience and liberalism. The IPIP-NEO-120 is a public-domain representation of the NEO-PI-R, and its scales converge substantially with their NEO-PI-R counterparts (Johnson, 2014; Maples et al., 2014). This convergence supports broader relevance to the five-factor model but does not imply identical results for every inventory designed to assess it.

With these scope conditions in place, we next clarify how hierarchical personality models relate to the specific criterion of dimensionality tested by state-trace analysis.

2. Hierarchy and dimensionality

Mainstream personality models commonly represent personality hierarchically, with broad domains and narrower facets, sometimes connected by intermediate aspects (DeYoung et al., 2007; Soto & John, 2017). Lower-order traits within a domain share variance attributable to the broader domain while also retaining trait-specific variance. Thus, the existence of five broad domains does not preclude meaningful differentiation at lower levels of the hierarchy. Indeed, facets are intended to capture precisely this differentiation, which may be characterized in factor-analytic terms as multidimensionality within a domain. When facets are represented as distinct lower-order latent variables, the resulting factor-analytic model contains at least as many lower-order dimensions as facets.

Within this framework, unidimensionality is evaluated through the covariance structure: a set of indicators is treated as unidimensional when their covariation is adequately represented by a single common factor, without substantively important residual associations after that factor has been taken into account. The absence of such residual

associations is closely related to the requirement of local independence in item response theory and its residual-dependence diagnostics (e.g., Chen & Thissen, 1997; Edwards et al., 2018; Yen, 1984). These diagnostics are conditional on the specified factor or item-response model and its functional-form assumptions.

The present state-trace argument adopts a different perspective, asking how many latent dimensions are minimally required to account for the observed responses. Here, unidimensionality means that the relevant observed variables can be represented as monotonic functions of a single latent variable—the most basic formal requirement for measurement on a single ordered psychological dimension (Dunn & Kalish, 2018). We occasionally use the more explicit term *monotonic unidimensionality* to distinguish this definition from covariance-based uses of unidimensionality. If an indicator first increased and then decreased as the latent trait increased, it could not consistently locate persons or experimental states along the same latent continuum. Throughout this article, we therefore operationalize unidimensionality as compatibility with such a one-dimensional monotonic representation.

Correlations and factor loadings cannot directly establish a common monotonic ordering. A low or moderate correlation can occur between variables that are perfectly ordered along one dimension, whereas a substantial correlation or factor loading can conceal a nonmonotonic relation. Standard factor analysis based on Pearson correlations additionally assumes linear relationships. Such linearity is neither evident in empirical patterns nor required for unidimensionality. This distinction has been demonstrated formally and empirically for Pearson correlations and the unidimensionality index H (Titz, 2025b). Factor analysis and discriminant validity remain useful for describing shared variance, prediction, and differential associations, but they do not test the monotonic ordering that is necessary for unidimensional measurement (Dunn & Kalish, 2018).

Accordingly, evidence for a hierarchical covariance structure does not, by itself, determine the minimum number of latent dimensions required to account for the observed responses. If a single monotonic dimension accommodates the pattern across the relevant items, facets, or aspects, no additional dimension is required to explain the observed responses, regardless of how many such lower-order components are distinguished in the hierarchical model. The parsimonious one-dimensional account can therefore be retained unless evidence requiring an additional dimension emerges. This does not prove that no additional latent sources could exist: multiple sources may remain empirically indistinguishable if they do not vary independently in the observed data. The inference in the opposite direction is stronger. If the state-trace is nonmonotonic, a single latent dimension is insufficient (Dunn & Kalish, 2018). Whether an additional dimension is required is therefore a question about monotonicity, not a direct consequence of specifying facet-level structure in a hierarchical covariance model.

This point is illustrated by previous state-trace-inspired analyses of personality ratings. In IPIP-NEO-120 responses from approximately 619,000 participants, the violations of monotonicity for neuroticism and conscientiousness were each attributable to only one item in the corresponding domain (Titz, 2025a). Thus, apart from one localized exception in each domain, the responses were essentially unidimensional across items from all six facets. This is a positive result: it provides a more parsimonious representation and supports interpreting an aggregate domain score as essentially a single ordered scale.

Departures in most other Big Five domains were likewise concentrated in a small number of items or facet relations. Taken together, these findings demonstrate that the mere presence of facets in a hierarchy does not, by itself, imply that multiple latent dimensions are required to account for the observed responses. Openness differed from this general pattern: its violations were distributed broadly across facets, suggesting that even a two-dimensional division may be insufficient and that, within this operationalization, as many dimensions as facets may be required (Titz, 2025a).

The relation between hierarchy and dimensionality has direct consequences for interpreting a domain score: it can be treated either as a pragmatic composite or as a measure of a single latent continuum. As a composite, it may be useful for prediction even when its items or facets are not compatible with a single dimension. Interpreting it as a single trait score, however, requires the contributing responses to be commensurable within one monotonic ordering. State-trace analysis evaluates this stronger interpretation. A violation does not eliminate every possible use of the composite, but it shows that the composite cannot be interpreted as measurement of a single latent continuum without qualification. Having clarified how hierarchical structure relates to unidimensionality, we now turn to the design of the experiment.

3. Methods

Experimental investigations of personality perception are often conducted using vignette-based designs (e.g., Atzmüller & Steiner, 2010; Baguley et al., 2022). For instance, researchers may manipulate specific features of a vignette to make a fictitious target appear more or less open to experience, and then assess participants’ perceptions of the target’s associated traits (e.g., liberalism and intellect). The present experiment uses controlled vignette manipulations to test whether observers’ judgments of liberalism and intellect are consistent with a single latent dimension.

3.1. Design

In the design of a state-trace experiment, three sets of variables are central (Dunn & Kalish, 2018; Prince et al., 2012): independent variables (IVs), latent variables (LVs), and dependent variables (DVs). The IVs are assumed to influence the DVs via their effects on one or more latent variables. The typical objective is to determine whether a single latent variable is sufficient — or insufficient — to account for the observed empirical patterns. Applied to this study, the question is whether perceived openness, understood as a latent dimension, can adequately account for participants’ judgments of the targets’ intellect and liberalism (DVs).

Specifically, the independent variables are properties of the vignette descriptions, and the dependent variables are participants’ judgments. The manipulation therefore permits causal inference about the effect of the supplied information on those judgments.

A well-designed state-trace experiment requires careful selection and manipulation of the independent variables. Inadequate planning risks producing a non-diagnostic state-trace (Prince et al., 2012), thereby rendering the experiment uninformative. Central to this design is the inclusion of a *dimension factor*—a variable intended to challenge unidimensionality and typically motivated by theoretical considerations. In the present case, identifying such a factor is comparatively straightforward, as variables exist that positively affect judgments of intellect while negatively affecting judgments of liberalism. Notably, facet C1 (self-efficacy) satisfies this criterion, exhibiting a strong positive correlation (.41) with O5 (intellect) and a slightly negative correlation (-.11) with O6 (liberalism) in the IPIP-NEO-120 (Kajonius & Johnson, 2019). Accordingly, two levels of this factor (low vs. high self-efficacy) were selected. Importantly, self-efficacy is a facet of conscientiousness, not of openness to experience, thereby reducing direct semantic overlap between the manipulated and evaluated openness-related content.

To improve the mapping of the state-trace, it is customary to introduce a *trace factor*—a variable that shifts the level of the hypothesized single underlying dimension. Doing so helps to mitigate problems of non-diagnosticity (Prince et al., 2012). In the present study, the trace factor was designed to modulate perceived openness in observers’ judgments. In principle, any facet of openness other than intellect or liberalism could serve this function. We selected the facet *artistic interest*, which correlates $r = .41$ with intellect and $r = .23$ with

liberalism in the IPIP-NEO-120 (Kajonius & Johnson, 2019). Three levels of this factor (low, medium, high) were implemented, yielding six experimental conditions in a 2×3 design. The dependent variables were judgments of liberalism and intellect.

In choosing between a between-subjects and a within-subjects design, there are principled arguments in favor of both approaches. Between-subjects vignette designs risk confounding experimental conditions with idiosyncratic contextual or framing effects (e.g., Birnbaum, 1999). Conversely, within-subjects designs may be affected by carry-over or contrast effects, insofar as participants’ ratings of a given vignette can be influenced by preceding ones. We adopted a fully within-subjects design while using the randomized first condition as an additional between-subjects comparison. The two analyses provide complementary evidence: the complete data offer greater sensitivity for mapping the state-trace, whereas the first-condition data are unaffected by carry-over from preceding vignettes. The order of conditions was randomized separately for each participant.

3.2. Vignettes

A simple vignette was constructed in which the bolded terms “inaccurate” and “accurate” were experimentally manipulated. The English version (the study was conducted in Germany in German, see Appendix for item translations) reads as follows:

Ms. N. completes a questionnaire designed to assess her personality. She indicates the extent to which various statements apply to her, using a scale ranging from one to five (see illustration).

1	2	3	4	5
<i>inaccurate</i>	<i>moderately</i>	<i>neither</i>	<i>moderately</i>	<i>accurate</i>
	<i>inaccurate</i>	<i>inaccurate nor</i>	<i>accurate</i>	
		<i>accurate</i>		

You will now see a selection of statements from the questionnaire. Assume that Ms. N. is entirely “correct” in her self-assessments: that is, the statements apply to her exactly to the extent indicated by her ratings.

Ms. N. rates each of the following three statements as **inaccurate**:

- I excel in what I do.
- I handle tasks smoothly.
- I know how to get things done.

Ms. N. rates each of the following three statements as **accurate**:

- I believe in the importance of art.
- I see beauty in things that others might not notice.
- I like poetry.

The items for the vignettes were taken from the self-efficacy (C1, Items 35, 65, 95) and artistic interest facets (O2, Items 8, 38, 68) of the IPIP-NEO-120.

To clarify the context of the fictitious person’s self-assessment, the word “Ich” (“I”) was added at the beginning of every item.

Note that Item O68 was reversed into “like poetry” to make the group of items uniformly positively keyed and avoid confusion due to negative wording (the original is: “do not like poetry”).

3.3. Independent variables

For self-efficacy, the fictitious vignette respondent endorsed each item as either *inaccurate* (scale point 1) or *accurate* (scale point 5). For artistic interest, each item was endorsed as *inaccurate* (1), *neither inaccurate nor accurate* (3), or *accurate* (5). This manipulation yielded six within-subjects conditions.

3.4. Dependent variables

After reading the vignettes, participants had to rate the described person on the liberalism and intellect scales of the IPIP-NEO-120 (Johnson, 2014).³ The following are the English-language IPIP-NEO-120 items for the liberalism and intellect scales, respectively (see Appendix for the German translations):

- Item O28: Tend to vote for liberal political candidates
- Item O58: Believe that there is no absolute right or wrong
- Item O88: Tend to vote for conservative political candidates
- Item O118: Believe that we should be tough on crime
- Item O23: Love to read challenging material
- Item O53: Avoid philosophical discussions
- Item O83: Have difficulty understanding abstract ideas
- Item O113: Am not interested in theoretical discussions

The sum of all items constituting the facet scale was calculated to determine the facet scores. The same five-point Likert scale as the one illustrated in the vignette was used for participants' responses. Consequently, the DV values had a possible range from 4 to 20.

3.5. Manipulation and attention checks

In addition to the dependent measures, participants rated the fictitious target on Item C05 ("Completes tasks successfully") and a positively keyed version of Item O98 ("Enjoys going to art museums") from the IPIP-NEO-120 (Johnson, 2014). These items assess self-efficacy and artistic interest, respectively, and served as manipulation checks; neither appeared in the vignettes. If the manipulations worked as intended, the manipulated level of self-efficacy should be strongly associated with ratings on Item C05, and the manipulated level of artistic interest with ratings on Item O98. Successful checks would indicate that participants interpreted the vignette descriptions consistently with the intended IPIP-NEO facets, thereby strengthening the conceptual link between the elicited perceptions and the corresponding constructs assessed in real individuals.

In addition, attention checks were implemented to maintain participants' engagement with the survey and to identify inattentive respondents for exclusion from the analyses. The first attention check was administered immediately after the first vignette and required participants to reproduce the fictitious person's self-rating on item C35 ("I excel in what I do") from the vignette they had just read. The remaining attention checks were embedded within the items assessing the vignette person's intellect and liberalism. These checks required participants to select a specific response option (e.g., "This question checks your attention. Please select 5 = accurate.").

For the dependent variables and manipulation checks, items were rephrased in third person to ensure grammatical correctness and clarify the context of participants' assessments of the fictitious person (e.g., "Ms. N. loves to read challenging materials" instead of "I love to read challenging materials").

3.6. Translations

Since the study was conducted in German, we used German translations of the IPIP-NEO-120 items. We prioritized the translations by Dege (2018), as they provide the most comprehensive validation evidence. For Item C35, however, we adopted the translation by Treiber (2013), because Dege reported a low item-scale correlation ($r = .21$).

³ Note that these scales are called openness to values (short: values) and openness to ideas (short: ideas) in the NEO-PI-R (Costa & McCrae, 2008).

For Item O88, we also used Treiber's translation, as it is less ambiguous.⁴

Item O28 was modified to remove explicit references to the German "Sozialdemokratische Partei Deutschlands" (SPD).⁵

All translations of the IPIP-NEO-120 items used in the main study are provided in Appendix.

3.7. Software and statistics

For state-trace analysis, the STACMR-R package (Dunn & Kalish, 2020) was used, which implements the coupled monotonic regression (CMR) algorithm (Kalish et al., 2016). The significance level α is conventionally set to .05; however, because two statistical tests were conducted, it was adjusted to .025. A total of 100,000 bootstrap samples were drawn, yielding resolution for p -values down to 0.00001.

The null hypothesis states that the relationship between liberalism and intellect is monotonic in the population, which is equivalent to the unidimensional hypothesis. The alternative hypothesis posits that the relationship is nonmonotonic, implying that at least two dimensions are required.

The state-trace was plotted using the package ggplot2 (Wickham, 2016).

Data, analysis code, and materials availability are reported in the Transparency subsection and the Open Practices section.

For the manipulation check, Items C05 and O98 were each correlated with the independent variables. Item C05 was correlated with a dummy variable representing the level of self-efficacy, and Item O98 was correlated with an ordinal variable representing the level of artistic interest.

3.8. Power based on pilot study

Currently, no closed-form power analysis is available for state-trace analysis. To inform sample size determination, we conducted a pilot study with 42 participants, in which both facets were manipulated at two levels (high/low self-efficacy and high/low artistic interest) in a 2×2 between-subjects design.

The goodness-of-fit (or, more precisely, badness-of-fit) statistic for the CMR model was 2.94, with $p = .081$, and therefore did not reach significance at $\alpha = .05$. Bootstrap power-analysis simulations based on the pilot data indicated that the original sample size would need to be increased by a factor of approximately 2.5 (to $N = 105$) to achieve at least 80% power to reject a one-dimensional account of the elicited judgments, assuming that the effect estimated from the pilot data reflected the population effect.

After the pilot study, we also decided to add an additional level of artistic interest based on inspection of the state-trace plot. A smaller number of experimental conditions makes it easier to account for the state-trace with a monotonic model (Prince et al., 2012). By increasing the number of conditions, a larger portion of the state-trace space is mapped. Moreover, if two dimensions underlie the judgments, the inclusion of this additional level is expected to increase statistical power.

⁴ Treiber: "Ich tendiere dazu, konservative Politiker zu wählen", compared with Dege's "Ich neige dazu, konservative bzw. rechtsliberale Politiker zu wählen." *Rechtsliberal* is an uncommon German political term denoting a combination of right-oriented and liberal positions (Dudenredaktion, n.d.); it is not synonymous with *konservativ*. Treiber's wording therefore corresponds more directly to the original item.

⁵ Specifically, "Neigt dazu, sozialdemokratische bzw. links-liberale Politiker zu wählen" was changed to "Neigt dazu, links-liberale Politiker zu wählen".

3.9. Sample

A total of 116 participants took part in the state-trace experiment. Three participants were excluded from all analyses because they failed to correctly answer two or more attention-check questions, resulting in a final sample size of $N = 113$. One additional participant was excluded from the within-subjects analyses, but retained in the between-subjects analyses. This participant reported having interpreted several vignettes as describing the same fictitious person. While this is unproblematic for the first vignette, it compromises the interpretation of the subsequent ones.

Because the order of vignettes was automatically randomized for each participant, the first vignette was not evenly distributed across conditions, resulting in an unbalanced design for the between-subjects analysis ($n_1 = 14, n_2 = 15, n_3 = 23, n_4 = 19, n_5 = 17, n_6 = 25$). Note that a balanced design is not a requirement for state-trace analysis.

Participants had a mean age of 28.47 years ($SD = 13.06$). Regarding biological sex, 69 participants (61.06%) identified as female and 41 (36.28%) as male; two participants identified as diverse, and one did not select any of the provided categories.

The majority of participants (79; 69.91%) were students. Among these, most were studying psychology (56 students; 70.89%).

Regarding highest educational attainment, most participants reported either a German high school diploma (67; 59.29%) or a university degree (41; 36.28%).

3.10. Recruitment and incentive

Participants were eligible to participate if they were at least 18 years old, possessed native-level proficiency in German, and had access to a quiet, undisturbed location with a stable internet connection for the estimated duration of the experiment (30 min). Participants were recruited via the study participation mailing list of Chemnitz University of Technology and through social media posts (Instagram, Facebook) published on the second author's private account. Informed consent was obtained electronically prior to participation. Upon completing the online survey, participants could choose between receiving credit for 0.5 h of research participation toward a curricular requirement or entering a draw for one of five cash prizes (€10 each).

3.11. Ethics

The experiment received approval from the Chemnitz University of Technology ethics committee under number #101651960.

3.12. Transparency

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study, and we follow JARS reporting standards. Analyses were conducted in R using the packages STACMR-R (Dunn & Kalish, 2020) and ggplot2 (Wickham, 2016). The data needed to reproduce all reported results, the analysis code, study materials, and supplementary materials are available in the Zenodo repository: <https://doi.org/10.5281/zenodo.23083760>. The study was not preregistered. Additional repository details are provided in the Open Practices section.

4. Results

4.1. Manipulation check

The Pearson correlation between the self-efficacy manipulation and Item C05 was $r = .79, p < 2.2 \times 10^{-16}$.

The Spearman correlation between the artistic-interest manipulation and Item O98 was $\rho = .72, p < 2.2 \times 10^{-16}$.

4.2. Between-subjects state-trace analysis

The empirical mean values underlying the state-trace analysis, together with the best-fitting CMR model, are depicted in the state-trace plot (Fig. 2). The CMR model yielded a Sum of Squares Error of 11.30, corresponding to $p \approx .0143$.

4.3. Within-subjects state-trace analysis

The empirical mean values of the state-trace as well as the best-fitting CMR model are illustrated in a state-trace plot (Fig. 3). The Sum of Squares Error for the best-fitting CMR model was 85.32, $p < .00001$ (not a single bootstrap sample was as extreme as the empirical data).

5. Discussion

The aim of this experiment was to examine whether participants' perceptions of targets' intellect and liberalism could be represented by a single latent dimension. This question relates to the long-standing debate concerning the multidimensionality of openness to experience (e.g. DeYoung et al., 2007; Johnson, 1994; Saucier, 1992). Although strong evidence for a violation of unidimensionality based on self-report data has already been reported (Titz, 2025a), the present study provides a confirmatory experimental test of whether manipulated target information produces an analogous pattern in personality judgments.

Participants rated fictitious targets on liberalism and intellect after reading short vignettes about them. The vignettes described the targets as having different levels of self-efficacy and artistic interest. We used state-trace analysis to examine the relation between the resulting liberalism and intellect judgments across experimental conditions. A monotonic relation — whereby higher liberalism judgments would always be accompanied by equal or higher intellect judgments — provided a poor fit to the data.

At each level of artistic interest, participants judged targets described as having low rather than high self-efficacy as more liberal but less intellectual. This opposing pattern appeared in both the between- and within-subjects analyses. The resulting nonmonotonic state-traces are incompatible with a one-dimensional account of participants' liberalism and intellect judgments under the present experimental conditions.

5.1. Between- versus within-subjects analyses

Both the between- and within-subjects analyses rejected the monotonic model, although the statistical evidence was stronger in the within-subjects analysis. The within-subjects analysis was expected to provide greater statistical power because it used all six vignette conditions rated by each participant, whereas the between-subjects analysis used only the randomized first vignette. The condition means also showed greater differentiation of intellect judgments in the within-subjects data. Compared with the corresponding between-subjects means, targets described as low in both self-efficacy and artistic interest were judged less intellectual in the within-subjects analysis (7.62 vs. 9.84), whereas targets described as high in both characteristics were judged more intellectual (16.79 vs. 15.06).

A possible explanation for this pattern is the presence of contrast effects. For example, when a vignette describing a person with low artistic interest and low self-efficacy is followed by one describing a person with high artistic interest and high self-efficacy, the latter may receive a particularly extreme rating.

Another possible explanation is that vignettes in between-subjects designs may be evaluated differently because they evoke different frames of reference (Birnbaum, 1999). For example, vignette conditions describing a moderate level of artistic interest may evoke the frame of reference of an average “everyday-life” person, against which the

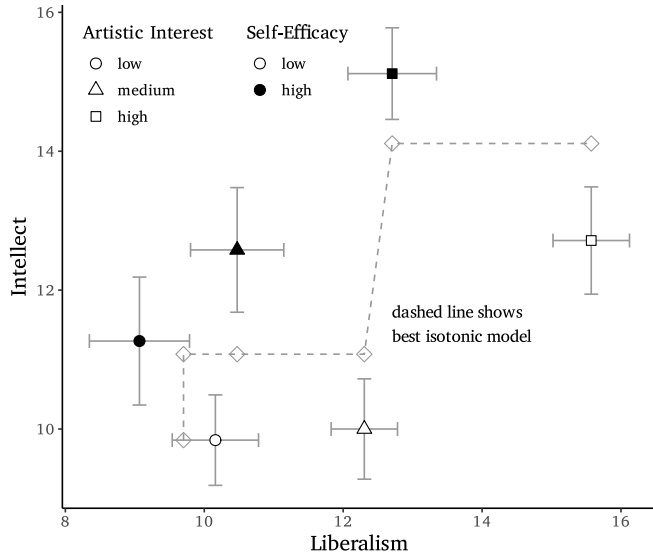


Fig. 2. Results of between-subjects analysis. Circles, triangles, and squares with horizontal and vertical error bars represent condition means and standard errors for low, medium, and high artistic interest, respectively. Among these condition markers, open symbols denote low self-efficacy and filled symbols high self-efficacy. Gray diamonds connected by the dashed line denote the best-fitting isotonic model.

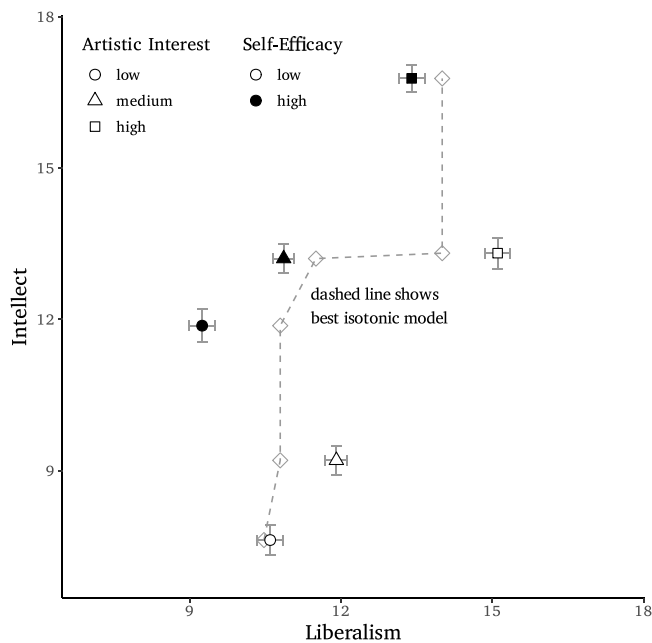


Fig. 3. Results of within-subjects analysis. Circles, triangles, and squares with error bars represent condition means and standard errors for low, medium, and high artistic interest, respectively. Among these condition markers, open symbols denote low self-efficacy and filled symbols high self-efficacy. Gray diamonds connected by the dashed line denote the best-fitting isotonic model.

vignette is then judged. In contrast, a vignette describing high self-efficacy and strong artistic interest may evoke a frame of reference involving high achievers or exceptionally talented individuals, to which the vignette is then compared. In other words, between-subjects designs may lead participants to compare each vignette with its own baseline within a given condition. Because previously rated vignettes provide a shared frame of reference, this problem is less likely to arise in within-subjects designs.

Note that the converging evidence for a nonmonotonic state-trace from both between-subjects and within-subjects data makes it difficult to attribute the results to a pure artifact of either design strategy.

5.2. Dimensionality of personality perceptions

The present results indicate that at least two dimensions are required to account for participants' liberalism and intellect judgments under these experimental conditions. At first glance, this finding might be interpreted as further support for hierarchical accounts that posit multiple lower-order dimensions within each Big Five domain (e.g., DeYoung et al., 2007; Soto & John, 2017). However, the present account differs from these proposals in important ways. Whereas DeYoung et al. (2007) and Soto and John (2017) rely primarily on correlational evidence, the present experiment provides two complementary forms of evidence. The manipulation permits causal inference about how the supplied information shaped participants' judgments, and state-trace analysis provides a formal test of whether those judgments can be represented by a single latent dimension.

As discussed in the Introduction, Titz (2025a) reported evidence that directly challenges the general expectation of multiple lower-order dimensions in every domain. In approximately 619,000 IPIP-NEO-120 respondents, neuroticism and conscientiousness were nearly unidimensional across all six facets, providing little support for additional dimensions in these domains. The clear exception was openness: within this operationalization, its widely distributed violations suggested that accounting for the observed response pattern may require as many dimensions as facets.

Nevertheless, regarding multidimensionality more broadly, the present results are consistent with the arguments advanced by DeYoung et al. (2007) and Soto and John (2017). In particular, as they note, correlations among facets do not constitute sufficient evidence for unidimensionality (DeYoung et al., 2007), and no facet of openness can be considered factor-pure (Soto & John, 2017).

Importantly, violations of monotonicity do not imply that broad openness scores lack predictive or practical utility. They do, however, pose a substantive measurement problem for interpreting domain scores formed by summing across such facets. These scores may remain useful as pragmatic composites, but when the contributing facet scores are incompatible with a common monotonic ordering, the resulting sum cannot be interpreted straightforwardly as locating respondents on a single latent continuum. Predictive usefulness therefore does not resolve the theoretical question of unidimensionality.

5.3. Relation to actual personality

The present experiment assessed how participants mapped experimentally manipulated personality descriptions onto judgments of liberalism and intellect within an IPIP-based semantic framework. The resulting state-trace therefore provides evidence of a nonmonotonic pattern in participants' implicit personality theories rather than evidence about the trait structure of real individuals. The experiment alone cannot establish generalization to actual personality and should instead be interpreted as complementary to evidence from assessments of real people. Research on implicit personality theory suggests that correspondence between perceived and actual personality structures can vary with the method used to elicit judgments (for a summary, see Borkenau, 1992).

One potential methodological concern is semantic similarity. If items used as independent and dependent variables are highly similar, participants may respond by recalling their previous ratings from short-term memory rather than by making independent judgments. [Semin and Greenslade \(1985\)](#) and [Semin and Fiedler \(1988\)](#) argued that distortions due to semantic similarity are particularly pronounced when trait terms are operationalized using adjectives rather than verbs or behavioral descriptions. In the present experiment, items from the IPIP-NEO-120 were employed, which rely on verbal and/or behavioral descriptions, thereby reducing this risk. Furthermore, self-efficacy was selected as the dimensional factor because its superordinate trait is conscientiousness and should therefore be semantically distant from openness to experience.

In ordinary social perception, observers draw on multiple cues, including age, gender, and outward appearance ([Albright et al., 1988](#)). In the present experiment, the nonmanipulated background information was deliberately sparse and held constant across vignettes. This strengthened experimental control by preventing uncontrolled target differences, but it also left participants with relatively little information on which to base their judgments. Participants may therefore have relied more heavily on prior beliefs about relations among personality traits.

Simply adding demographic or appearance cues would not necessarily reduce this reliance. Such cues can themselves activate stereotypes. A stronger test would provide richer information about the individual target without merely restating the manipulated IPIP items. This might include concrete behavior across several situations, a sample of the target's reasoning, or responses to value-relevant scenarios. Such evidence could constrain trait inferences and show whether the nonmonotonic pattern persists when judgments are based on a broader behavioral profile. Indeed, previous research suggests that diagnostic individuating information tends to reduce stereotype bias ([Rubinstein et al., 2018](#)).

Reliance on shared beliefs, however, would not necessarily make the resulting judgments wholly inaccurate. Research on moral stereotypes of liberals and conservatives shows that such stereotypes can reproduce the direction of observed group differences while exaggerating their magnitude ([Graham et al., 2012](#)). The observed pattern could therefore reflect a mixture of genuine covariation and stereotype-based amplification rather than an arbitrary misconception. At the same time, the content and strength of political stereotypes depend on cultural context. Because the present participants were German, it remains unclear whether observers in other societies would make the same inferences.

For the present findings, the most relevant candidate belief is not that liberals are more intellectual. That belief would, if anything, make the observed pattern more difficult to obtain. A more relevant belief is that higher self-efficacy implies greater intellect but lower liberalism. This belief is not ad hoc; it is consistent with associations observed in self-ratings ([Kajonius & Johnson, 2019](#)), in which self-efficacy correlates positively with intellect ($r = .41$) and slightly negatively with liberalism ($r = -.11$).

Evidence that political conservatism is positively associated with conscientiousness (e.g., [Hirsh et al., 2010](#); [Xu et al., 2021](#)), together with the positive association between conscientiousness and intellect ([DeYoung et al., 2007](#)), suggests a possible indirect pathway linking conservative orientation to higher intellect through conscientiousness. At the same time, higher intellect is generally associated with greater liberalism ([Xu et al., 2021](#)).⁶ Thus, intellect may be positively associated with both conservative and liberal orientation through different pathways. The self-efficacy manipulation was designed to test this possibility under controlled conditions and produced the expected opposing shifts in intellect and liberalism judgments. The experimental judgments may therefore reflect a tension that is also observed in real-person correlational data, although the present design cannot establish that both settings have the same underlying structure.

⁶ [Jost et al. \(2003\)](#) reported a meta-analytic association of approximately $r = -.32$ between political conservatism and openness to experience.

5.4. Homogeneous sample

The sample was predominantly female, young, and highly educated, and most participants were students. The findings may therefore not generalize to populations with different demographic or educational profiles.

Most student participants were studying psychology and may have been familiar with Big Five models and with the convention of classifying liberalism and intellect as facets of openness to experience. Such familiarity could influence the personality schema they applied to the targets.

Participants may also have used themselves as a frame of reference, particularly when they perceived a vignette target as similar to themselves, and projected aspects of their own personality onto the target. Although self-target similarity and projection were not assessed directly, the strong manipulation-check correlations ($r = .79$ and $\rho = .72$) show that participants' ratings closely tracked the manipulated characteristics. Any disruptive self-referential influence therefore appears limited relative to the intended manipulation effects and did not substantially obscure participants' sensitivity to the manipulations.

To summarize, the present findings may not generalize without qualification to personality perceptions elicited by other conceptualizations of openness, non-IPIP frameworks, or more representative perceiver populations.

5.5. Implications and recommendations for future research

Several methodological improvements could enhance the validity and generalizability of future vignette-based studies examining the relationship between liberalism and intellect. Personality judgments are known to be more accurate when observers can directly perceive behavior in unstructured situations than when traits are inferred indirectly from written descriptions ([Funder, 2012](#)). Accordingly, a more naturalistic mode of vignette presentation may help address potential validity concerns. Future studies could present targets in audio or video, for example in a simulated free conversation. LLM-assisted dialogue generation combined with synthetic speech or video could make the presentation more naturalistic than written vignettes while retaining experimental control over the information conveyed. Related approaches have already been used to generate experimental text materials ([Lechuga & Karle, 2026](#)), manipulate perceived Big Five traits through synthetic speech ([Aylett et al., 2020](#)), and alter perceived personality, including openness, in generated videos ([Yurtoğlu et al., 2024](#)). Such stimuli should nevertheless be pretested for naturalness and unintended vocal or visual differences between conditions.

In addition, future studies should aim to recruit samples that are more representative of the general population. Doing so would help rule out potential biases associated with convenience samples and allow researchers to assess whether the observed dimensional violations are robust across different demographic groups.

Beyond liberalism and intellect, it would be informative to extend the present experimental paradigm to other problematic facet relations within openness to experience, as well as to other Big Five factors. Openness to experience is a particularly promising candidate, as numerous violations of unidimensionality have been reported for this factor. However, other theoretically interesting cases also exist, such as the relationship between modesty and trust within agreeableness, or items such as "I like to take it easy", which can plausibly reflect both introversion and extraversion ([Titz, 2025a](#)).

The present study may serve as a blueprint for such extensions by illustrating how the conceptual assumptions and semantics underlying state-trace analysis can be implemented.

Table A1
German item translations used in the main experiment.

item	scale	German translation	Source of translation
C35	Self-efficacy	Ich bin hervorragend in dem, was ich tue	Treiber, 2013
C65	Self-efficacy	Ich erledige Aufgaben elegant/reibungslos ^a	Dege, 2018
C95	Self-efficacy	Ich verstehe es, Dinge zu erledigen	Dege, 2018
C05	Self-efficacy	Schließt Aufgaben erfolgreich ab ^b	Dege, 2018
O08	Artistic interests	Ich glaube an die Wichtigkeit von Kunst ^a	Dege, 2018
O38	Artistic interests	Ich sehe Schönheit in Dingen, die anderen möglicherweise nicht auffällt ^a	Dege, 2018
O68	Artistic interests	Ich mag Gedichte ^{a,c}	Dege, 2018
O98	Artistic interests	Geht gerne in Kunstmuseen ^{b,c}	Dege, 2018
O28	Liberalism	Neigt dazu, linksliberale Politiker zu wählen ^{b,d}	Dege, 2018
O58	Liberalism	Glaubt, dass es kein absolutes richtig oder falsch gibt ^b	Dege, 2018
O88	Liberalism	Tendiert dazu, konservative Politiker zu wählen ^b	Treiber, 2013
O118	Liberalism	Glaubt, dass wir Verbrechen hart ahnden sollten ^b	Dege, 2018
O23	Intellect	Liest gerne anspruchsvolle/komplexe Texte ^b	Dege, 2018
O53	Intellect	Vermeidet philosophische Diskussionen ^b	Dege, 2018
O83	Intellect	Hat Schwierigkeiten, abstrakte Ideen zu verstehen ^b	Dege, 2018
O113	Intellect	Ist wenig an theoretischen Diskussionen interessiert ^b	Dege, 2018

^a The word “Ich” (German for “I”) was added to the translation to clarify the context of self-assessment.
^b Item was reformulated in third person for grammatical correctness and clarity.
^c The original translation source item is inverted.
^d Item was modified to avoid associations with specific German political parties.

5.6. Conclusion

The present study provides experimental state-trace evidence that perceived liberalism and intellect may not function as monotonic indicators of a single openness dimension within personality descriptions derived from the IPIP-NEO-120. This conclusion is consistent with previous findings, but the present study extends the existing evidence by testing confirmatory hypotheses with an advanced method for assessing unidimensionality and by relying on experimental data. Writing about controversies over the number and nature of personality dimensions, Eysenck (1997) argued that purely correlational, factor-analytic approaches could not resolve them because the causal nexus was missing, and called for theoretical deductions to be tested experimentally.

State-trace analysis is particularly well positioned to help meet this challenge. When combined with controlled manipulations, it provides a direct test of dimensionality and offers personality research a new route toward resolving structural controversies that have traditionally been approached through correlational methods.

Open practices

The data needed to reproduce all reported results, the analysis code, study materials, and supplementary materials are available in the Zenodo repository: <https://doi.org/10.5281/zenodo.23083760>. The repository includes a standalone materials file documenting the vignette procedure, all six experimental conditions, exact item wording, response scales, outcome measures, manipulation checks, and attention checks. The study was not preregistered.

CRedit authorship contribution statement

Johannes Titz: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Benedikt Quirin Gailer:** Writing – review & editing, Writing – original draft, Visualization, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of Generative AI and AI-Assisted Technologies in the Manuscript Preparation Process

During the preparation of this work, the authors used OpenAI’s ChatGPT with the GPT-5.5 model and Codex with the GPT-5.6-Sol model to assist with language editing and stylistic revision. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix. Translations of IPIP-NEO-120 items used in the study

This appendix lists all German IPIP item translations used in the main study (Table A1). The column “German translation” presents the items as they were administered in the study. In many cases, these items do not correspond exactly to those in the original translation sources. Instead, most source items were adapted to better align with the objectives of this study. These adaptations are indicated by footnotes.

Data availability

Data, code, and materials sufficient to reproduce the reported results and methodology are available at <https://doi.org/10.5281/zenodo.23083760>.

References

Albright, L., Kenny, D. A., & Malloy, T. E. (1988). Consensus in personality judgments at zero acquaintance. *Journal of Personality and Social Psychology*, 55(3), 387–395. <http://dx.doi.org/10.1037/0022-3514.55.3.387>.
Atzmüller, C., & Steiner, P. M. (2010). Experimental vignette studies in survey research. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 6(3), 128–138. <http://dx.doi.org/10.1027/1614-2241/a000014>.
Aylett, M. P., Vinciarelli, A., & Wester, M. (2020). Speech synthesis for the generation of artificial personality. *IEEE Transactions on Affective Computing*, 11(2), 361–372. <http://dx.doi.org/10.1109/TAFFC.2017.2763134>.
Baguley, T., Dunham, G., & Steer, O. (2022). Statistical modelling of vignette data in psychology. *British Journal of Psychology*, 113(4), 1143–1163. <http://dx.doi.org/10.1111/bjop.12577>.
Birnbaum, M. H. (1999). How to show that 9 > 221: Collect judgments in a between-subjects design. *Psychological Methods*, 4(3), 243–249. <http://dx.doi.org/10.1037/1082-989X.4.3.243>.
Borkenau, P. (1992). Implicit personality theory and the five-factor model. *Journal of Personality*, 60(2), 295–327. <http://dx.doi.org/10.1111/j.1467-6494.1992.tb00975.x>.

- Chen, W.-H., & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics*, 22(3), 265–289. <http://dx.doi.org/10.3102/10769986022003265>.
- Costa, P. T., & McCrae, R. R. (1995). Domains and facets: Hierarchical personality assessment using the revised NEO personality inventory. *Journal of Personality Assessment*, 64(1), 21–50. http://dx.doi.org/10.1207/s15327752jpa6401_2.
- Costa, P. T., & McCrae, R. R. (2008). The revised NEO personality inventory (NEO-PI-R). In G. J. Boyle, G. Matthews, & D. H. Saklofske (Eds.), *The SAGE handbook of personality theory and assessment*, vol. 2: personality measurement and testing (pp. 179–198). SAGE Publications, <http://dx.doi.org/10.4135/9781849200479.n9>.
- Dege, D. C. (2018). *Adaptation des IPIP-NEO-PI-R ins Deutsche* [BA thesis], University of Erfurt.
- DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology*, 93(5), 880–896. <http://dx.doi.org/10.1037/0022-3514.93.5.880>.
- DeYoung, C. G., Quilty, L. C., Peterson, J. B., & Gray, J. R. (2014). Openness to experience, intellect, and cognitive ability. *Journal of Personality Assessment*, 96(1), 46–52. <http://dx.doi.org/10.1080/00223891.2013.806327>.
- Dudenredaktion (n.d.). *Rechtsliberal*. In Duden online. Retrieved October 7, 2026, from <https://www.duden.de/node/157958/revision/1274935>.
- Dunn, J. C. (2008). The dimensionality of the remember-know task: A state-trace analysis. *Psychological Review*, 115(2), 426–446. <http://dx.doi.org/10.1037/0033-295X.115.2.426>.
- Dunn, J. C., & Kalish, M. L. (2018). *State-trace analysis*. Springer International Publishing, <http://dx.doi.org/10.1007/978-3-319-73129-2>.
- Dunn, J. C., & Kalish, M. L. (2020). *Guide to the STACMR-R package*. GitHub, Retrieved November 29, 2025, from <https://github.com/michaelkalish/STA/blob/master/STACMR-R/STACMR-R.pdf>.
- Dunn, J. C., & Kirsner, K. (1988). Discovering functionally independent mental processes: The principle of reversed association. *Psychological Review*, 95, 91–101. <http://dx.doi.org/10.1037/0033-295X.95.1.91>.
- Dunn, J. C., Newell, B. R., & Kalish, M. L. (2012). The effect of feedback delay and feedback type on perceptual category learning: The limits of multiple systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(4), 840–859. <http://dx.doi.org/10.1037/a0027867>.
- Dunn, J. C., & Rao, L.-L. (2019). Models of risky choice: A state-trace and signed difference analysis. *Journal of Mathematical Psychology*, 90, 61–75. <http://dx.doi.org/10.1016/j.jmp.2018.12.005>.
- Edwards, M. C., Houts, C. R., & Cai, L. (2018). A diagnostic procedure to detect departures from local independence in item response theory models. *Psychological Methods*, 23(1), 138–149. <http://dx.doi.org/10.1037/met0000121>.
- Eysenck, H. J. (1997). Personality and experimental psychology: The unification of psychology and the possibility of a paradigm. *Journal of Personality and Social Psychology*, 73(6), 1224–1237. <http://dx.doi.org/10.1037/0022-3514.73.6.1224>.
- Funder, D. C. (2012). Accurate personality judgment. *Current Directions in Psychological Science*, 21(3), 177–182. <http://dx.doi.org/10.1177/0963721412445309>.
- Graham, J., Nosek, B. A., & Haidt, J. (2012). The moral stereotypes of liberals and conservatives: Exaggeration of differences across the political spectrum. *PLoS One*, 7(12), Article e50092. <http://dx.doi.org/10.1371/journal.pone.0050092>.
- Hakel, M. D. (1969). Significance of implicit personality theories for personality research and theory. *Proceedings of the Annual Convention of the American Psychological Association*, 4, 403–406.
- Hayes, B. K., Stephens, R. G., Ngo, J., & Dunn, J. C. (2018). The dimensionality of reasoning: Inductive and deductive inference can be explained by a single process. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(9), 1333–1351. <http://dx.doi.org/10.1037/xlm0000527>.
- Heathcote, A., Holloway, E., & Sauer, J. (2019). Confidence and varieties of bias. *Journal of Mathematical Psychology*, 90, 31–46. <http://dx.doi.org/10.1016/j.jmp.2018.10.002>.
- Hirsh, J. B., DeYoung, C. G., Xu, X., & Peterson, J. B. (2010). Compassionate liberals and polite conservatives: associations of agreeableness with political ideology and moral values. *Personality and Social Psychology Bulletin*, 36(5), 655–664. <http://dx.doi.org/10.1177/0146167210366854>.
- International Personality Item Pool (n.d.). Converting IPIP item responses to scale scores, <https://ipip.ori.org/newScoringInstructions.htm>.
- Jang, Y., Lee, H., & Huber, D. E. (2019). How many dimensions underlie judgments of learning and recall redux: Consideration of recall latency reveals a previously hidden nonmonotonicity. *Journal of Mathematical Psychology*, 90, 47–60. <http://dx.doi.org/10.1016/j.jmp.2018.10.006>.
- Johnson, J. A. (1994). Clarification of factor five with the help of the AB5C model. *European Journal of Personality*, 8(4), 311–334. <http://dx.doi.org/10.1002/per.2410080408>.
- Johnson, J. A. (2014). Measuring thirty facets of the Five Factor Model with a 120-item public domain inventory: Development of the IPIP-NEO-120. *Journal of Research in Personality*, 51, 78–89. <http://dx.doi.org/10.1016/j.jrp.2014.05.003>.
- Jost, J. T., Glaser, J., Kruglanski, A. W., & Sulloway, F. J. (2003). Political conservatism as motivated social cognition. *Psychological Bulletin*, 129(3), 339–375. <http://dx.doi.org/10.1037/0033-2909.129.3.339>.
- Kajonius, P. J., & Johnson, J. A. (2019). Assessing the structure of the Five Factor Model of personality (IPIP-NEO-120) in the public domain. *Europe's Journal of Psychology*, 15(2), 260–275. <http://dx.doi.org/10.5964/ejop.v15i2.1671>.
- Kalish, M. L., Dunn, J. C., Burdakov, O. P., & Sysoev, O. (2016). A statistical test of the equality of latent orders. *Journal of Mathematical Psychology*, 70, 1–11. <http://dx.doi.org/10.1016/j.jmp.2015.10.004>.
- Kaufman, S. B., Quilty, L. C., Grazioplene, R. G., Hirsh, J. B., Gray, J. R., Peterson, J. B., & DeYoung, C. G. (2016). Openness to experience and intellect differentially predict creative achievement in the arts and sciences. *Journal of Personality*, 84(2), 248–258. <http://dx.doi.org/10.1111/jopy.12156>.
- Lechuga, J., & Karle, N. N. (2026). Generating experimental text stimuli for psychological research using ChatGPT. *Advances in Methods and Practices in Psychological Science*, 9(1), Article 25152459251407631. <http://dx.doi.org/10.1177/25152459251407631>.
- Liu, W., & Jia, Y. (2022). Multiple latent variables but functionally dependent output mappings underlying the recognition of own- and other-race faces for Chinese individuals: Evidence from state-trace analysis. *Frontiers in Psychology*, 13, Article 968956. <http://dx.doi.org/10.3389/fpsyg.2022.968956>.
- 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. <http://dx.doi.org/10.1037/pas0000004>.
- Mirels, H. L. (1976). Implicit personality theory and inferential illusions. *Journal of Personality*, 44(3), 467–487. <http://dx.doi.org/10.1111/j.1467-6494.1976.tb00133.x>.
- Oleynick, V. C., DeYoung, C. G., Hyde, E., Kaufman, S. B., Beaty, R. E., & Silvia, P. J. (2017). Openness/intellect: The core of the creative personality. In G. J. Feist, R. Reiter-Palmon, & J. C. Kaufman (Eds.), *The Cambridge handbook of creativity and personality research* (pp. 9–27). Cambridge University Press, <http://dx.doi.org/10.1017/9781316228036.002>.
- Prince, M., Brown, S., & Heathcote, A. (2012). The design and analysis of state-trace experiments. *Psychological Methods*, 17(1), 78–99. <http://dx.doi.org/10.1037/a0025809>.
- Robinson, M. M., & Irwin, D. E. (2019). Are there two visual short-term memory stores? A state-trace analysis. *Journal of Mathematical Psychology*, 90, 23–30. <http://dx.doi.org/10.1016/j.jmp.2019.02.001>.
- Rubinstein, R. S., Jussim, L., & Stevens, S. T. (2018). Reliance on individuating information and stereotypes in implicit and explicit person perception. *Journal of Experimental Social Psychology*, 75, 54–70. <http://dx.doi.org/10.1016/j.jesp.2017.11.009>.
- Saucier, G. (1992). Openness versus intellect: Much ado about nothing? *European Journal of Personality*, 6(5), 381–386. <http://dx.doi.org/10.1002/per.2410060506>.
- Semin, G. R., & Fiedler, K. (1988). The cognitive functions of linguistic categories in describing persons: Social cognition and language. *Journal of Personality and Social Psychology*, 54(4), 558–568. <http://dx.doi.org/10.1037/0022-3514.54.4.558>.
- Semin, G. R., & Greenslade, L. (1985). Differential contributions of linguistic factors to memory-based ratings: Systematizing the systematic distortion hypothesis. *Journal of Personality and Social Psychology*, 49(6), 1713–1723. <http://dx.doi.org/10.1037/0022-3514.49.6.1713>.
- Soto, C. J., & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology*, 113(1), 117–143. <http://dx.doi.org/10.1037/pspp0000096>.
- Titz, J. (2025a). State-trace analysis meets personality measurement: Why the Big Five tests are not based on five latent dimensions and how to fix them. *PLoS One*, 20(2), Article e0317144. <http://dx.doi.org/10.1371/journal.pone.0317144>.
- Titz, J. (2025b). The relationship between the phi coefficient and the unidimensionality index H: Improving psychological scaling from the ground up. *Psychological Methods*, <http://dx.doi.org/10.1037/met0000736>, Advance online publication.
- Treiber, L. (2013). *Entwicklung psychometrischer Überprüfung und konvergente Validierung der deutschsprachigen 30-Facetten-IPIP-Skala* [BA thesis], Universität Koblenz-Landau.
- Vartanian, O., Wertz, C. J., Flores, R. A., Beatty, E. L., Smith, I., Blackler, K., Lam, Q., & Jung, R. E. (2018). Structural correlates of openness and intellect: Implications for the contribution of personality to creativity. *Human Brain Mapping*, 39(7), 2987–2996. <http://dx.doi.org/10.1002/hbm.24054>.
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer, <http://dx.doi.org/10.1007/978-3-319-24277-4>.
- Willoughby, E. A., Kim, Y., Lee, J. J., & DeYoung, C. G. (2023). The construct validity of intellect and openness as distinct aspects of personality through differential associations with reaction time. *Journal of Intelligence*, 11(2), 30. <http://dx.doi.org/10.3390/jintelligence11020030>.
- Xu, X., Soto, C. J., & Plaks, J. E. (2021). Beyond openness to experience and conscientiousness: Testing links between lower-level personality traits and American political orientation. *Journal of Personality*, 89(4), 754–773. <http://dx.doi.org/10.1111/jopy.12613>.
- Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement*, 8(2), 125–145. <http://dx.doi.org/10.1177/014662168400800201>.
- Yurtoğlu, A., Sonlu, S., Doğan, Y., & Güdükbay, U. (2024). Personality perception in human videos altered by motion transfer networks. *Computers & Graphics*, 119, Article 103886. <http://dx.doi.org/10.1016/j.cag.2024.01.013>.