

## Secondary Publication



Leising, Daniel; Borgstede, Matthias; Burger, Julian; u. a.

### **Why Do Judgments on Different Person-Descriptive Attributes Correlate With One Another? : A Conceptual Analysis With Relevance for Most Psychometric Research**

Date of secondary publication: 18.08.2026

Version of Record (Published Version), Article

Persistent identifier: urn:nbn:de:bvb:473-irb-116851x

#### **Primary publication**

Leising, Daniel; Borgstede, Matthias; Burger, Julian; u. a. (2025): Why Do Judgments on Different Person-Descriptive Attributes Correlate With One Another?: A Conceptual Analysis With Relevance for Most Psychometric Research, in: Collabra: Psychology, Oakland, CA: University of California Press, vol. 11, no. 1, pp. 1–24, doi: 10.1525/collabra.133683

#### **Legal Notice**

This work is protected by copyright and/or the indication of a licence. You are free to use this work in any way permitted by the copyright and/or the licence that applies to your usage. For other uses, you must obtain permission from the rights-holders.

This document is made available under a Creative Commons license.



The license information is available online:

<https://creativecommons.org/licenses/by/4.0/legalcode>

Personality-Psychology-

# Why Do Judgments on Different Person-Descriptive Attributes Correlate With One Another? A Conceptual Analysis With Relevance for Most Psychometric Research

Daniel Leising<sup>1a</sup>, Matthias Borgstede<sup>2</sup>, Julian Burger<sup>3</sup>, Johannes Zimmermann<sup>4</sup>, Martin Bäckström<sup>5</sup>, Joshua Oltmanns<sup>6</sup>, Nele Freyer<sup>1</sup>, Anne Wiedenroth<sup>1</sup>, Paula Knischewski<sup>1</sup>, Brian Connelly<sup>7</sup>

<sup>1</sup> Technische Universität Dresden, Dresden, Germany, <sup>2</sup> Universität Bamberg, Bamberg, Germany, <sup>3</sup> Yale University, New Haven, CT, USA, <sup>4</sup> Universität Kassel, Kassel, Germany, <sup>5</sup> Lunds Universitet, Lund, Sweden, <sup>6</sup> University of Kentucky, Lexington, KY, USA, <sup>7</sup> University of Toronto, Toronto, ON, Canada

Keywords: factor analysis, network model, measurement, psychometrics, response style, attitude, semantics, person perception

<https://doi.org/10.1525/collabra.133683>

## Collabra: Psychology-

Vol.-11, Issue-1, 2025-

Patterns of correlations among judgments of targets on different attributes are the basis for common psychometric procedures such as factor analysis and network modeling. The outcomes of such analyses may shape the images (i.e., theories) that we as scientists have of the phenomena that we study. However, key conceptual issues tend to be overlooked in these analyses, which is especially problematic when the items are descriptions expressed in the natural language. A correlation between judgments on two such attributes may reflect the influences of (a) a common substantive cause, (b) substantive target characteristics on another, (c) semantic redundancy, (d) the perceivers' attitudes toward the targets, (e) the perceivers' formal response styles, or (f) any mixture of these. We present a conceptual framework integrating all of these mechanisms and use it to connect formerly unrelated strands of theorizing with one another. A lack of awareness regarding the complexity involved may compromise the validity of interpretations of psychometric analyses. We also review the effectiveness of a broad range of solutions that have been proposed for dealing with the various influences, and provide recommendations for future research.

“One of the hazards of science is the ease with which it is possible to confuse propositions about the world with propositions about language” (D’Andrade, 1965)

## Introduction

A great amount of psychometric research is based on correlations between judgments of targets on different person-descriptive attributes. Scale scores are often computed from sets of items that correlate with one another, assuming that these correlations indicate that the items somehow “measure the same thing”. However, most of this research fails to fully acknowledge the complexity of everything that “sameness” may represent and thus runs a risk of leading to misinterpretations of findings. In this article, we present a conceptual framework that is supposed to help organize this complexity, in order to help avoid such misinterpretations. We also review the effectiveness of various methods that have been used to disentangle the various factors

that may influence correlation patterns among items and attempt to give best practice recommendations.

The core problem may be conceived as the risk of mistaking associations between person judgments for direct reflections of associations between the actual target characteristics underlying those judgments. In other words, there is a risk of mistaking the words one uses to describe certain target characteristics for those characteristics themselves. Person judgments may, however, reflect several other influences apart from those actual target characteristics, and those influences may partly or even completely explain why judgments correlate with one another. In order to be able to make valid inferences regarding the meaning of inter-item correlations, researchers need to be aware of those other influences, and account for them in the way they conduct their research. This does not only concern the ways in which data is analyzed, but also the ways in which data is collected (i.e., study design). Only after accomplishing that

a Correspondence: Daniel Leising, Assessment and Intervention Research Unit, Technische Universität Dresden, 01062 Dresden, Germany  
Phone: +49 176 616 93286  
[Daniel.Leising@tu-dresden.de](mailto:Daniel.Leising@tu-dresden.de)

will it be possible to draw valid inferences regarding associations between actual target characteristics.

We present a conceptual analysis of five different mechanisms that may account for correlations between person judgments on different attributes. All individual components of this analysis have been discussed in the literature before (e.g., Brunswik, 1956; Funder, 1995; Kenny, 1994, 2004; Leising et al., 2015, 2021; Podsakoff et al., 2003; West & Kenny, 2011). The unique contribution of the present paper is their integration within a single, unified formal framework that applies to different kinds of target characteristics and across different types of psychometric models. We use this analysis to highlight crucial interpretational pitfalls in the analysis of judgment data, and we provide concrete suggestions as to how these pitfalls may be avoided.

In the first section, we present the overarching conceptual framework, focusing on five distinct mechanisms that may produce correlations among judgments of target persons. Next, we discuss how these mechanisms are relevant to psychometrics, especially research on so-called “general factors” (of personality, personality pathology, and psychopathology) and on network models. We further ask how appropriately studies using these approaches in the more applied fields have so far accounted for the various mechanisms that may be responsible for correlations between attributes. We close with a detailed review of methods that have been tried to disentangle the mechanisms from one another, and attempt to derive a few evidence-based best practice recommendations.

Although we primarily use examples from the content domains of personality, personality disorders, and psychopathology to illustrate our analysis, the mechanisms outlined herein are equally applicable to any domain in which judgments using the natural language are collected, such as supervisor ratings of job performance (e.g., Viswesvaran et al., 2005), teacher ratings of children’s behavioral disinhibition (e.g., Bishop et al., 2003), or in appraising one’s subjective well-being (e.g., Eid & Diener, 2004).

## The Basic Conceptual Model<sup>2</sup>

Our primary concern in this paper are the correlations that may exist between judgments of targets ( $\tau$ ) by perceivers ( $\pi$ ) on different attributes. We will introduce the model components using a generic data structure in which each target is judged by a different perceiver, and the dyads of targets and perceivers are the same across all attributes. This includes the situation in which the perceiver and the target in each dyad are the same persons (i.e., self-rating data), as a special case.

For example, imagine that researchers were aiming to understand why judgments on two given attributes ( $J_1$  = “joyful” and  $J_2$  = “jealous”) correlate with one another. They recruit a sample of targets (Tina, Tom, and Ted), each

of whom is rated on both attributes by a different perceiver (e.g., Peter judges Tina, Paula judges Tom, and Priscilla judges Ted). It is straightforward to calculate the correlation between these judgments across the three dyads of perceivers and targets. Understanding the possible origins of that correlation is a thornier matter, and the focus of the present paper.

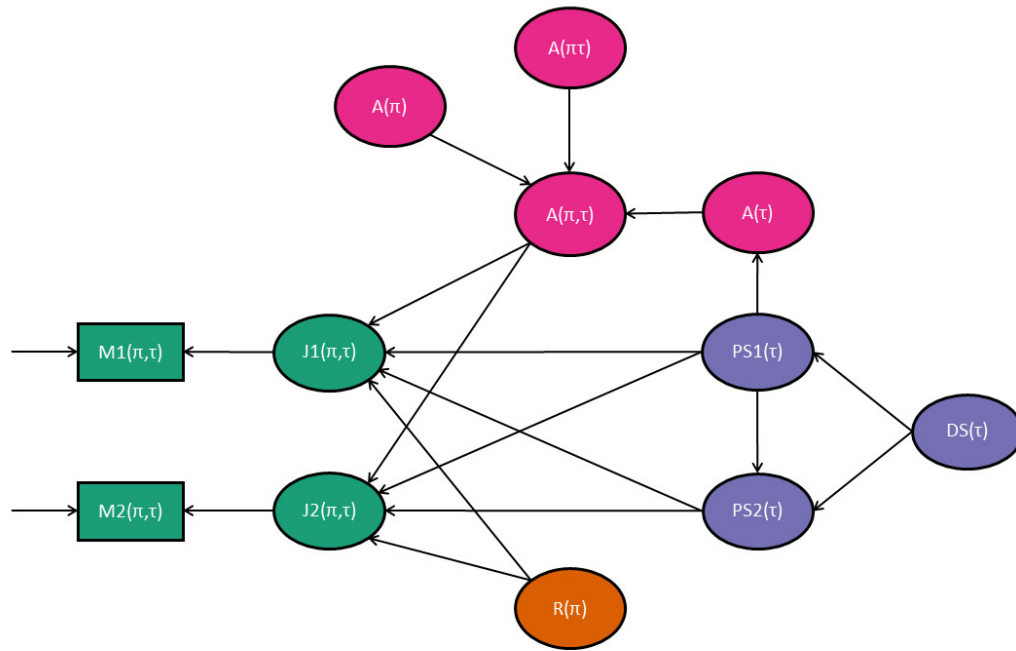
Figure 1 presents the conceptual model that we will use throughout the present paper to analyze this issue in depth. Table 1 lists and explains all parameters. The model integrates ideas from several previous models, such as Brunswik’s (1956) lens model, Funder’s (1995) Realistic Accuracy Model, Kenny’s (1994, 2004) Weighted Average-/PERSON model, and West and Kenny’s (2011) Truth and Bias model. It distinguishes between four kinds of variables: Judgments (J), Substance (S), Attitudes (A), and Response styles (R).

Throughout this paper, we use color coding<sup>1</sup> to emphasize a fundamental difference between these variables: Judgments (green) are ideas that perceivers may have regarding targets’ standings on a given attribute. Such judgments may be measured by asking perceivers (e.g., Peter, Paula, and Priscilla) to rate targets (e.g., Tina, Tom, and Ted) on items (e.g., “joyful” and “jealous”). In our model, the only thing that distinguishes measurements ( $M_1$ ,  $M_2$ ) from judgments ( $J_1$ ,  $J_2$ ) is that the former also contain some random noise. Thus, the judgment variables in the model are conceptually identical to the measurements’ true (or latent, or noiseless) scores, in line with Classical Test Theory (Novick, 1966). Whereas it is the correlation between (noisy) measurements that is used in many psychometric analyses, we focus on the correlation between (noise-free) judgments in the present paper.

Figure 1 does contain measurements ( $M_1$ ,  $M_2$ ) of the targets on two items. Following SEM conventions, they are displayed as rectangles, whereas all other variables are displayed as ovals. All subsequent figures contain only ovals, to make it clear that we are addressing structural issues and largely ignore the problem of random noise. We model the correlation  $\rho_{J_1(\pi, \tau), J_2(\pi, \tau)}$  between the true or noiseless judgments  $J_1$ ,  $J_2$  on two attributes, of which the observed correlation  $r_{M_1(\pi, \tau), M_2(\pi, \tau)}$  between measurements ( $M_1$ ,  $M_2$ ) on two items will simply be a derivative attenuated by measurement error.

Substance (purple) refers to “real” or “actual” differences that exist between the targets and that may inform the perceivers’ judgments more or less strongly. For example, the actual height of targets (S) may inform judgments of how “tall” (J) they are. Likewise, we would assume that judgments of how “joyful” ( $J_1$ ) or “jealous” ( $J_2$ ) targets are reflect some actual differences between the targets (e.g., S1: how often they smile, S2: how often they criticize others). In other words, the substance is about *what is* being de-

1- The information contained in the color coding is redundant with the capital letters that we also use (J, S, A, R). The colors are just used for emphasis.



**Figure 1.2** Full conceptual model displaying *possible* connections between variables.<sup>2</sup>

$\pi$  = perceivers;  $\tau$  = targets;  $\pi\tau$  = perceiver-target dyads;  $PS1(\tau)$ ,  $PS2(\tau)$  = two proximal substance variables;  $J1(\pi, \tau)$ ,  $J2(\pi, \tau)$  = judgments of the targets by the perceivers on two attributes;  $M1(\pi, \tau)$ ,  $M2(\pi, \tau)$  = measured judgments on two items, including random noise;  $A(\pi, \tau)$  = attitudes of the perceivers regarding the targets;  $A(\pi)$  = perceiver effects in attitudes,  $A(\tau)$  = target effects in attitudes,  $A(\pi\tau)$  = dyadic effects in attitudes;  $R(\pi)$  = formal response styles of the perceivers;  $DS(\tau)$  = a distal substance variable that may act as a common substantive cause. Arrows symbolize causal effects.

**Table 1.2** Model Parameters<sup>2</sup>

| Meaning                                                | Noisy | Parameter in Model                    | Color in Displays |
|--------------------------------------------------------|-------|---------------------------------------|-------------------|
| Perceivers                                             | No    | $\pi$                                 |                   |
| Targets                                                | No    | $\tau$                                |                   |
| Targets' standings on proximal substance variable 1    | No    | $PS1(\tau)$                           | Purple            |
| Targets' standings on proximal substance variable 2    | No    | $PS2(\tau)$                           | Purple            |
| Targets' standings on distal substance variable        | No    | $DS(\tau)$                            | Purple            |
| Attitudes (of perceivers, toward targets)              | No    | $A(\pi, \tau)$                        | Magenta           |
| Attitudes (of perceivers, averaged across targets)     | No    | $A(\pi)$                              | Magenta           |
| Attitudes (toward targets, averaged across perceivers) | No    | $A(\tau)$                             | Magenta           |
| Attitudes (specific for perceiver-target dyads)        | No    | $A(\pi\tau)$                          | Magenta           |
| Perceivers' formal response styles                     | No    | $R(\pi)$                              | Orange            |
| Judgments of targets by perceivers on attribute 1      | No    | $J1(\pi, \tau)$                       | Green             |
| Judgments of targets by perceivers on attribute 2      | No    | $J2(\pi, \tau)$                       | Green             |
| Correlation between judgments on attributes 1 and 2    | No    | $\rho_{J1(\pi, \tau), J2(\pi, \tau)}$ |                   |
| Measurements on item 1                                 | Yes   | $M1(\pi, \tau)$                       | Green             |
| Measurements on item 2                                 | Yes   | $M2(\pi, \tau)$                       | Green             |
| Correlation between measurements on items 1 and 2      | Yes   | $r_{M1(\pi, \tau), M2(\pi, \tau)}$    |                   |

scribed, as opposed to *how* (e.g., more positively or negatively) it is being described.<sup>2-</sup>

For didactic reasons we will only tie each attribute to a single substance variable in the following, although judgments on most attributes are likely to reflect compounds (“blends”) of several substance variables. For simplicity, we will also focus entirely on between-target variance in the substance variables and treat substance variation between perceivers and between dyads as noise.<sup>3-</sup>

The substance is assumed to exist irrespective of whether or how any judgments ever take place, and even if people were incapable of judging one another. Thus, we take a strong realist position, philosophically. In line with Brunswik- (1956), we also distinguish between two kinds of substance variables. The substantive variation that informs the perceivers’ judgments of the targets will be called “proximal” (PS1, PS2). Substantive variation that is just as real but not used by the perceivers in making their current judgments will be called “distal” (DS).-

Sometimes, it is necessary to account for the possibility that some distal substantive variation may drive covariation among proximal substance variables (see the PS-DS mechanism below). For example, some targets may tend to smile (S1) more often than other targets do, and to criticize others (S2) less often than other targets do. Such a correlation among substance variables may be due to the operation of a “common substantive cause” like a genetic disposition, or certain childhood experiences. Note that such differences would be just as “substantive” in nature (i.e., they are elements of the real external world), but we call them “distal” because they do not *directly* affect the perceivers’ judgments of the targets. Alternatively, the two substance variables may correlate with one another because one causally affects the other (see below).-

Attitudes (magenta) concern the extent to which perceivers consider targets to be their friends or their enemies (or neither). Another term that is commonly used in the personality literature when referring to this variable is “Liking” (e.g., Leising et al., 2010). More broadly, Osgood, Suci, and Tenenbaum- (1957)- reported finding a general “evaluation” factor in ratings that perceivers made of diverse sets of objects. We assume that all of this is basically the same thing. -

For attitudes, we use the letter A, and we distinguish between attitude components that are contributed by the targets ( $A(\tau)$ ), the perceivers ( $A(\pi)$ ), and the specific dyads of targets and perceivers ( $A(\pi\tau)$ ).<sup>4-</sup> In line with a variance-analytic “social relations” view of person judgments (Kenny, 1994), we assume that the perceivers’ overall attitudes to-

ward their respective targets ( $A(\pi, \tau)$ ) are the sum of these three influences. -

For example, targets may be judged as being more “joyful” but less “jealous” when their respective perceivers like them, and as less “joyful” but more “jealous” when their perceivers dislike them. Such an effect could be described via the interaction between the perceivers’ attitudes toward the targets and the so-called “social desirabilities” (Edwards, 1953) of the attributes. The emerging negative correlation between judgments on the two attributes would constitute what is called a “halo effect” (Thorndike, 1920). -

Formal response styles (orange) are the perceivers’ tendencies to judge people in certain ways *regardless of content*. For example, Peter may be more inclined than Paula to see *any* judgment of a target as being justified (“acquiescence”). Therefore, Peter would assign higher ratings than Paula on both items, creating a positive correlation between them (see the J-R mechanism below). Although other formal response styles are conceivable (e.g., a preference for the number 4 on a rating scale), we will only use acquiescence as an example in the present context. The letter R will be used for formal response styles, and we assume that formal response styles only vary between perceivers ( $R(\pi)$ ). -

Several more remarks are in order, to avoid possible misunderstandings. First, the numbers in the variable labels only serve to distinguish different variables of the same type from one another. They do not establish an exclusive connection between variables whose labels contain the same number. For example, J2 may not necessarily be affected only by PS2, but also by PS1 (see Figure 4). -

Second, the paths connecting the model variables with one another will later be indexed by the names of the variables they connect. This is necessary to enable precise analyses of possible effects, but is foregone in Figure 1, to preserve readability. All arrows symbolize possible causal effects. The model posits that some combination of the displayed causal effects is the reason why a correlation between the judgments exists. We assume all variables to be standardized for reasons of notational convenience. -

Third, it should be noted that the paths displayed in Figure 1 only constitute an assembly of *conceivable* influences. In a concrete dataset, some of these paths may actually be zero. Finally, even though the presentation may already seem fairly complex, it actually reflects a large number of simplifications that we had to make in order to limit this complexity. These simplifications are listed in Appendix A. -

2- For example, when people judge how “intelligent” they are, their judgments may reflect a mixture of substance (e.g., how much they know, how quickly they think) and attitude (e.g., people with higher self-esteem view themselves as being more “intelligent”). When people judge their own self-esteem, however, self-esteem would be a substance variable. -

3- If needed, the model could easily be amended with perceiver-specific and dyad-specific substance components. These could then have unique effects on corresponding attitude and/or judgment components. -

4- The question of how perceivers’ attitudes toward targets develop is a research topic of its own, with its own vast literature. In this paper, we do address the possibility that attitudes may be partly rooted in proximal substance variation, but otherwise ignore the issue. -



## Mechanisms Accounting for Correlations<sup>2</sup> Between Attributes<sup>2</sup>

We will now use the conceptual framework introduced above to highlight five different mechanisms that may bring about a correlation between attributes. Understanding these mechanisms is of crucial and direct importance in regard to how such correlations (and, by extension, scale scores, factors, and whole networks of relationships between items or scales) may be interpreted. The first two mechanisms are entirely about substance. This is what most applied psychology researchers are primarily interested in. Making substantive claims based on correlations among judgments is questionable, however, because this type of data may reflect at least three other influences that may each by itself bring about such a correlation.

We display each mechanism separately in graphical form (in Figures 2 to 7), temporarily “knocking-out” the other mechanisms. This is purely for didactic reasons. The outcome that is the focus of attention here is the “true”, unattenuated correlation between the perceivers’ judgments of the targets on two attributes ( $\rho_{J1(\pi, \tau), J2(\pi, \tau)}$ ). Each mechanism is indexed using an acronym that highlights the relevant type of path in the model explaining the emergence of this correlation.

### Mechanism 1: Common Substantive Causation<sup>2</sup> (PS-DS)<sup>2</sup>

One reason for judgments on different attributes to correlate with one another is the existence of a common substantive cause that affects the proximal substance variables which are reflected by these attributes. Figure 1 displays this situation. Here, the judgments on each attribute (J1 and J2) reflect a different proximal substance variable (PS1 and PS2), and the only reason that makes these judgments correlate with one another is the fact that the two proximal substance variables are affected by the same distal substantive cause (DS).

For example, a brain trauma (DS) may simultaneously account for the occurrence of two different kinds of symptoms (PS1, PS2), which would make judgments of those symptoms (J1: “headache”, J2: “dizziness”) correlate with one another. When looking only at targets with the trauma, or targets without the trauma, or when statistically controlling for the influence of the trauma, the correlation between the symptoms (and therefore, the judgments) should disappear (“local stochastic independence”). When a correlation between judgments is *exclusively* accounted for by this mechanism, it will reflect the following product of paths:

$$\rho_{J1(\pi, \tau), J2(\pi, \tau)} = \beta_{J1(\pi, \tau), PS1(\tau)} \cdot \beta_{PS1(\tau), DS(\tau)} \cdot \beta_{PS2(\tau), DS(\tau)} \cdot \beta_{J2(\pi, \tau), PS2(\tau)}$$

Of course, more than one causal factor may be at play, but that possibility will be ignored here, for the sake of simplicity. Also, common substantive causes (e.g., having a virus) may have yet more distal causes (e.g., having been exposed to people who carry the virus).

### Mechanism 2: One Proximal Substance Variable<sup>2</sup> Affecting Another (PS-PS)<sup>2</sup>

The proximal substance variable informing judgments on one attribute may causally affect the proximal substance variable informing judgments on another attribute. For example, a target’s tendency to get into fights more easily than other targets (PS1) may cause that target to lose their job more often than other targets (PS2). Such a pattern is displayed in Figure 2. Note that, in this example, PS1 may also be viewed as a distal substance variable with regard to J2. In such cases, we give precedence to whether a substance variable directly feeds into any of the judgment variables.

When a correlation between judgments is *exclusively* accounted for by this mechanism, it will reflect the following product of paths:

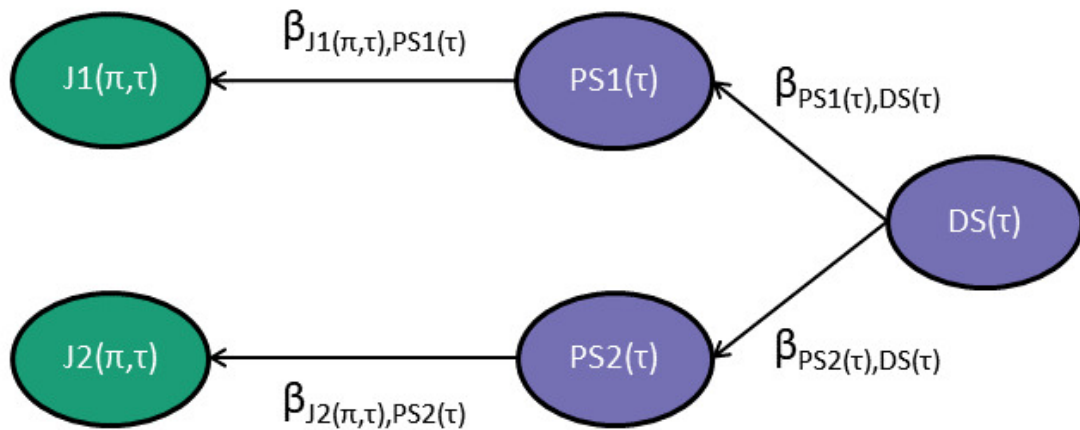
$$\rho_{J1(\pi, \tau), J2(\pi, \tau)} = \beta_{J1(\pi, \tau), PS1(\tau)} \cdot \beta_{PS2(\tau), PS1(\tau)} \cdot \beta_{J2(\pi, \tau), PS2(\tau)}$$

For simplicity, we only addressed the case in which one substance variable causally affects another. Of course, in reality the ways in which substance variables are associated with one another will often be more complicated. For example, there may be interactions, or feedback loops, or indirect effects of one variable on another via a third. All of this would make the interpretation of correlations among judgments even more difficult. Whole patterns of how variables affect other variables may be analyzed under a “network” or “dynamic systems” perspective (Borsboom & Cramer, 2013; Epskamp et al., 2018; Schmittmann et al., 2013), and similar arguments have been put forth as a “functionalist” or “dynamic” account of personality (e.g., Cramer et al., 2012; Wood et al., 2015) and intelligence (van der Maas et al., 2006).

### Mechanism 3: Shared Semantic References (J-PS)<sup>2</sup>

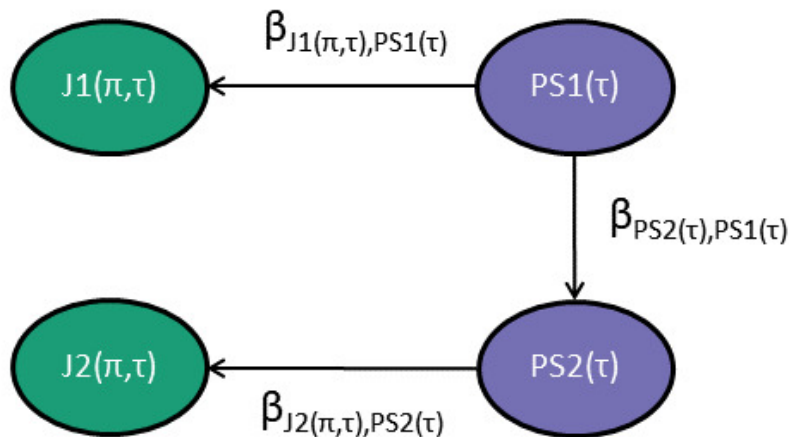
Another reason why judgments may correlate with one another is their reflecting the same proximal substance variables. For example, judgments of targets as being more or less “punctual” (J1) or “reliable” (J2) may both reflect the average number of minutes they come late to appointments (PS1). Such paths constitute the “meaning systems” in Kenny’s (1994) Weighted-Average-Model. The term “semantics” is also commonly used to refer to such paths.

Judgments of targets on two attributes may correlate with one another when these attributes share some of their semantic references (i.e., they are informed by the same proximal substantive variation). Figure 4 illustrates this. For example, if judgments on the attributes “punctual” (J1) and “reliable” (J2) both reflect the average number of minutes targets come late to appointments (PS1), then judgments of how “punctual” and how “reliable” targets are would correlate positively with one another. However, the correlation may be less than perfect because the term “reliable” has broader semantic connotations than the term “punctual”. For example, judgments of targets as being “reliable” may also reflect the average number of days it takes them to give back stuff that they borrowed (PS2). This type



**Figure 2. Common Substantive Causation (PS-DS).2**

$\pi$ =perceivers;  $\tau$ =targets;  $PS1(\tau)$ ,  $PS2(\tau)$ =two proximal substance variables;  $J1(\pi, \tau)$ ,  $J2(\pi, \tau)$ =judgments of the targets by the perceivers on two attributes;  $DS(\tau)$ =a distal substance variable that acts as a common substantive cause.



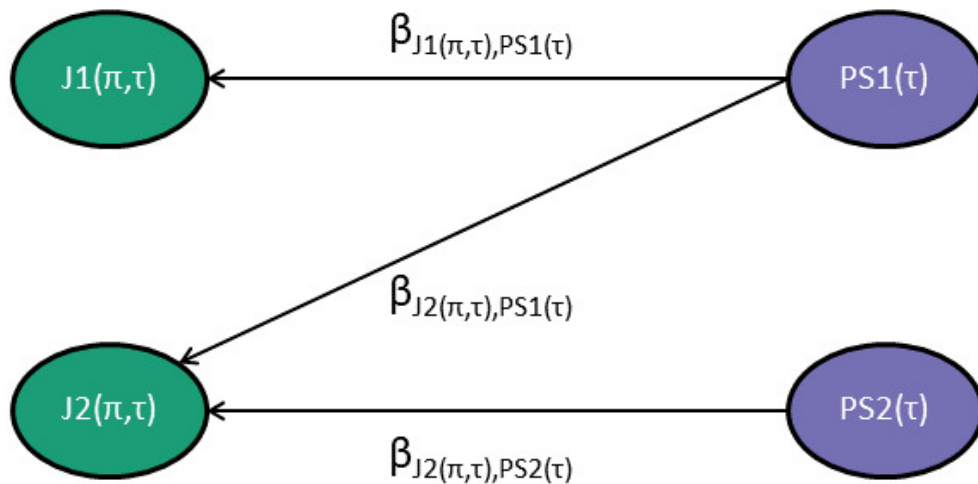
**Figure 3.2 One proximal substance variable affecting another (PS-PS).2**

$\pi$ =perceivers;  $\tau$ =targets;  $PS1(\tau)$ ,  $PS2(\tau)$ =two proximal substance variables;  $J1(\pi, \tau)$ ,  $J2(\pi, \tau)$ =judgments of the targets by the perceivers on two attributes.

of mixture of substances is what is meant by the term “blends” in the personality judgment literature (e.g., Ashton et al., 2009). The issue also touches directly on the “level of abstraction”, “category breadth”, or “bandwidth” of attributes (e.g., Hampson et al., 1986): Attributes that are broader or more abstract reflect more different substantive influences.

It is important to note the fundamental difference to the common substantive cause (PS-DS) mechanism introduced

above. The PS-DS mechanism is present when the same distal substance variable causally affects two proximal substance variables. In contrast, a shared semantic reference effect (J-PS) is present when judgments on two attributes reflect an influence of the same proximal substance variable. Although both mechanisms may appear to be similar in that they refer to substance variables that ultimately have an effect on more than one judgment, they are in fact independent of one another: J-PS may only operate when



**Figure 4.2 Shared Semantic References (J-PS).2**

$\pi$  = perceivers;  $\tau$  = targets;  $PS1(\tau)$ ,  $PS2(\tau)$  = two proximal substance variables;  $J1(\pi, \tau)$ ,  $J2(\pi, \tau)$  = judgments of the targets by the perceivers on two attributes.

judgments are being made whereas PS-DS operates *before* judgments are being made (cf. Buntins, 2014; Buntins et al., 2016).

When a correlation between judgments is *exclusively* accounted for by this Mechanism, it will reflect the following product of paths:

$$\rho_{J1(\pi, \tau), J2(\pi, \tau)} = \beta_{J1(\pi, \tau), PS1(\tau)} \cdot \beta_{J2(\pi, \tau), PS1(\tau)}$$

This product may also be interpreted directly as the semantic similarity of the two attributes. When more than one substance variable is involved and the different substance variables are uncorrelated, the formula would have to be amended by the respective products for each additional substance variable (e.g.,  $+\beta_{J1(\pi, \tau), PS2(\tau)} \cdot \beta_{J2(\pi, \tau), PS2(\tau)}$ ).

#### Mechanism 4: Shared Attitudes Effect (J-A)2

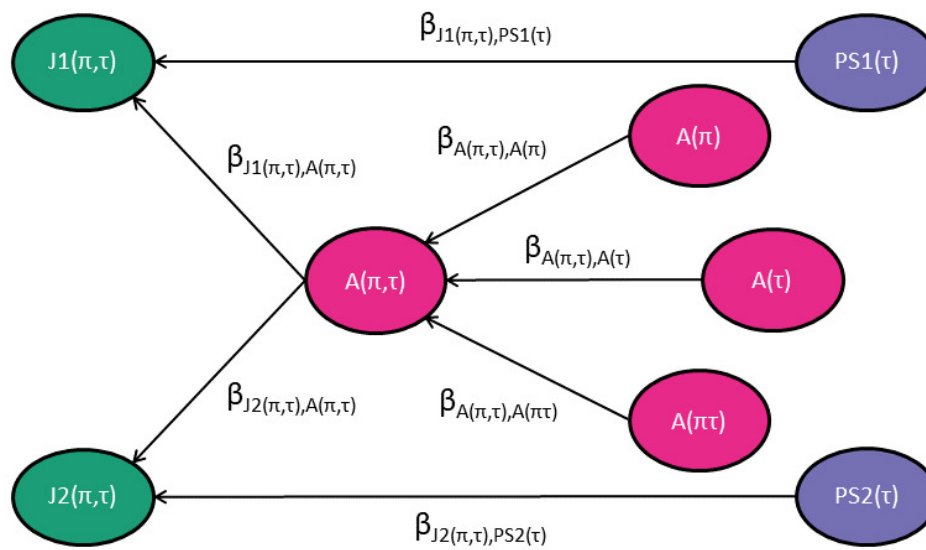
Correlations between judgments on two attributes may also come about because perceivers differ in the attitudes they have toward their respective targets, and those attitudes are reflected by each attribute. Figures 5 and 6 display two variants of this mechanism: In Figure 5, attitudes (A) are independent of the substantive target characteristics (PS1, PS2) that the attributes at hand reflect. In Figure 6, attitudes *do* depend on one of those characteristics (PS1). In reality, these mechanisms may operate at the same time, but we keep them separate here, for didactic reasons.

In Figure 5, the correlation is completely accounted for by attitude components that are unrelated to the substance which is reflected by the attributes (PS1 and PS2): A( $\pi$ ) stands for evaluative perceiver effects (i.e., some perceivers judging the average target more positively than other perceivers) (Rau et al., 2021; Srivastava et al., 2010; Wood et al., 2010). A( $\tau$ ) stands for evaluative target effects (i.e., some targets being judged more positively than other targets by the average perceiver—*irrespective* of the substance that is reflected by the attributes at hand). A( $\pi\tau$ ) stands for dyadic attitude effects (i.e., some perceivers judging their specific targets more positively than would be expected on grounds of perceiver- and target-effects). Each of these three components alone would suffice to bring about a correlation between judgments on any two attributes, as long as those attributes both reflect the perceivers' attitudes. For example, targets being judged as more joyful may be judged as being less jealous, simply because there are attitude differences and the first attribute reflects them positively and the latter reflects them negatively.

The weight by which judgments on a given attribute (J1) reflect the perceivers' attitudes ( $\beta_{J1(\pi, \tau), A(\pi, \tau)}$ ) is conceptually identical to what has been called the "social desirability" of attributes (Edwards, 1953).<sup>5</sup> The absolute difference between an attribute's social desirability and the point of neutrality on the rating scale is that attribute's "evaluativeness" (cf. John & Robins, 1993).

5- Note that in this literature the term "social desirability" is often used to denote a perceiver's tendency to portray the target positively (i.e., A( $\pi, \tau$ )). In the present paper, we use the exact same term to denote an item's evaluative tone in line with Edwards (1953) who introduced the concept into the literature.





**Figure 5. Shared Attitudes Effect (J-A), with attitudes that are independent of the substance to which the attributes refer.**

$\pi$ =perceivers;  $\tau$ =targets;  $\pi\tau$ =perceiver-target dyads;  $PS1(\tau)$ ,  $PS2(\tau)$ =two proximal substance variables;  $J1(\pi, \tau)$ ,  $J2(\pi, \tau)$ =judgments of the targets by the perceivers on two attributes;  $A(\pi, \tau)$ =attitudes of the perceivers regarding the targets,  $A(\pi)$ =perceiver effects in attitudes,  $A(\tau)$ =target effects in attitudes,  $A(\pi\tau)$ =dyadic effects in attitudes.

In fact, research has consistently shown that most person-descriptive terms in the natural language are strongly evaluative (Anderson, 1968; Dumas et al., 2002; Leising et al., 2012; Leistner et al., 2024), and the rated social desirability of attributes aligns almost perfectly with how their use reflects the perceivers' attitudes toward the targets (Leising et al., 2010, 2013). When a correlation between judgments is exclusively accounted for by this mechanism, it will reflect the following product of paths (Leising et al., 2021):

$$\rho_{J1(\pi, \tau), J2(\pi, \tau)} = \beta_{J1(\pi, \tau), A(\pi, \tau)} \cdot \beta_{J2(\pi, \tau), A(\pi, \tau)}$$

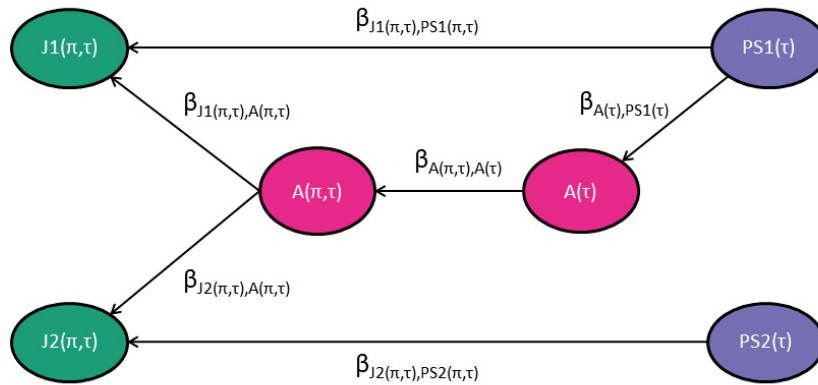
In Figure 6, the situation is fundamentally different: Here, the proximal substance that is reflected by one of the attributes ( $PS1$ ) does influence the perceivers' attitudes. This influence will then translate into a correlation to the extent that both attributes are evaluative. For example, judgments of targets on the attributes "attractive" ( $J1$ ) and "intelligent" ( $J2$ ) may be correlated not because the two proximal substance variables ( $PS1$ ,  $PS2$ ) are correlated, but because the proximal substance ( $PS1$ ) that is reflected by the first attribute ( $J1$ ) evokes positive attitudes in the average perceiver ( $A(\pi)$ ), which then affect judgments of the targets on both attributes ( $J1$  and  $J2$ ). For example,  $PS1$  may be facial symmetry (Noor & Evans, 2003).

All four ways in which perceiver attitudes may account for a correlation between judgments (displayed in Figures 5 and 6) constitute conceptually distinguishable instances of what is called a *halo effect* (Anusic et al., 2009; Gräff & Unkelbach, 2016; Lance et al., 1994; Nisbett & Wilson, 1977; Thorndike, 1920).

### Mechanism 5: Formal Response Styles Effect (J-R)2

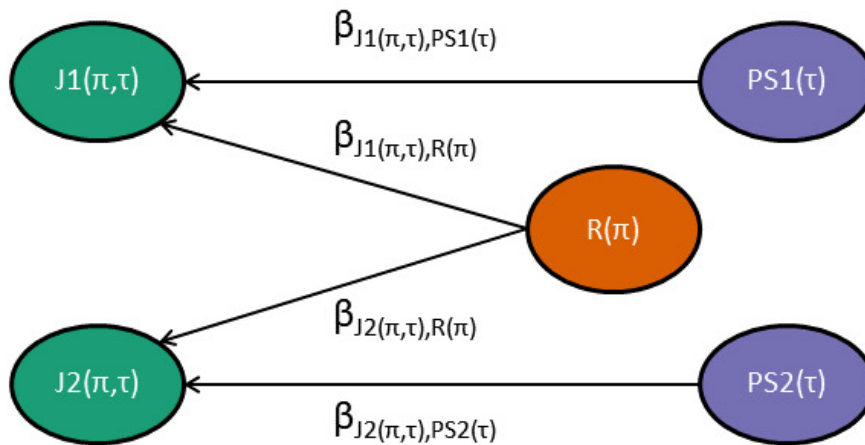
Finally, perceivers may differ from one another in which response options they prefer over other response options—irrespective of content—and this may bring about a correlation between judgments. For example, perceivers may have preferences for high vs. low numerical values, for specific locations on the response scale (left vs. middle vs. right), and/or for agreeing vs. disagreeing with any statements in general ("acquiescence"). To the extent that the items at hand permit different perceivers to act on such preferences, a correlation between judgments may emerge. With the typical Likert rating scales, this is the case.

Figure 7 displays a situation in which a correlation between attributes is brought about exclusively by a formal response style such as acquiescence. Each of the above-named formal response styles could be displayed separately.



**Figure 6. Shared Attitudes Effect (J-A), with attitudes being influenced by one proximal substance variable.**

$\pi$  = perceivers;  $\tau$  = targets.  $PS1(\tau)$ ,  $PS2(\tau)$  = two proximal substance variables;  $J1(\pi, \tau)$ ,  $J2(\pi, \tau)$  = judgments of the targets by the perceivers on two attributes;  $A(\pi, \tau)$  = attitudes of the perceivers regarding the targets.  $A(\tau)$  = target effects in attitudes.



**Figure 7. Formal Response Styles Effect (J-R).**

$\pi$  = perceivers;  $\tau$  = targets;  $\pi\tau$  = perceiver-target dyads;  $PS1(\tau)$ ,  $PS2(\tau)$  = two proximal substance variables;  $J1(\pi, \tau)$ ,  $J2(\pi, \tau)$  = judgments of targets by the perceivers on two attributes;  $R(\pi)$  = formal response styles of the perceivers.

in the Figure, but we chose to only display one for simplicity.<sup>6</sup>

When a correlation between judgments is exclusively accounted for by this mechanism, it will reflect the following product of paths:-

$$\rho_{J1(\pi, \tau), J2(\pi, \tau)} = \beta_{J1(\pi, \tau), R(\pi)} \cdot \beta_{J2(\pi, \tau), R(\pi)}$$

## Relevance of the Five Mechanisms for2 Assessment and Research2

Up to this point, we may have conveyed the impression that our conceptual analysis is mainly an academic exercise with little or no real-world implications. It is not. Quite to the contrary, this analysis is directly relevant to how (pat-

6- Technically, extreme responding (Wetzel et al., 2016) also constitutes a formal response style, but this one operates differently from the ones mentioned before. Perceivers with this response style tend to spread their ratings more across the rating scale. This may be formalized as random effects on the paths from the attributes to the measurements, but is ignored here.

terns-of)-associations-between-items-and/or-scales-may-or-may-not-be-appropriately-interpreted.-

The-key-point-to-make-here-is-that-a-correlation-between-judgments-on-any-two-attributes-is-likely-to-reflect-some-mixture-of-all five of the above-named mechanisms. Unless researchers apply specific methods to disentangle their influences from one another, they will be unable to determine the composition of the true scores underlying their measurements. Measurements on two attributes may correlate with one another because (a) they reflect two types of proximal substance variation that are both affected by a common and more distal substantive cause (PS-DS mechanism), and/or (b) at least one of those two types of proximal substance variation has a causal effect on the other type (PS-PS mechanism), and/or (c) the substantive between-target variation that the two attributes refer to is at least partially the same (J-PS mechanism), and/or (d) both attributes also reflect the perceivers' attitudes toward the targets (J-A mechanism) and/or (e) both attributes also reflect the perceivers' formal response styles (J-R mechanism).-

In this section, we will briefly touch on how all of this complicates the interpretation of factor analyses, scale scores, and network analyses. Before we do so, however, we have to talk at least briefly about some recent advancements in psychometric research that are relevant for most of this. Most important, judgments may be conceived of as resulting from interactions between properties of perceivers, targets, and dyads on the one hand, and properties of attributes on the other hand.-

Research has shown that (a) person-descriptive items may be very reliably rated by relatively small groups of (15-20) raters on a variety of dimensions: observability, social desirability, importance, base rate, traitness, abstractness, pairwise semantic similarity (Edwards, 1953, 1957; Funder & Dobroth, 1987; John & Robins, 1993; Leising et al., 2014), and (b) that these dimensions are in fact distinguishable from one another, although some seem to be meaningfully related (e.g., Leising et al., 2014). Furthermore, (c) they may predict the extent to which judgments reflect properties of perceivers, targets, and dyads (i.e., the above-mentioned interactions). For example, the perceivers' attitudes toward the targets will only affect their judgments to the extent that the respective attribute is evaluative. Many more such interactions are conceivable but empirical research in this area has just begun. More recent research suggests that (d) at least some of these ratings may be obtained in an automated fashion (Arnulf et al., 2014; Garcia et al., 2020; Leistner et al., 2024; Sikström & Garcia, 2019), which increases the speed and efficiency of studies very significantly. For example, Leistner et al. (2024) showed that hundreds of items may be quickly rated regarding several characteristics of attributes using Large Language Models, with the automated estimates being reasonably similar to ones that were provided by human raters.-

## Relevance for Factor Analytic Research<sup>2</sup>

Three largely independent literatures have emerged regarding general factors of (a) personality, (b) general psychopathology, and (c) personality pathology (Caspi et al.,

2014; Caspi & Moffitt, 2018; Davies et al., 2015; Hörz-Sagstetter et al., 2021; Irwing, 2013; Sharp et al., 2015). Most recently, a general factor of disease ('d') has also been proposed (Brandt et al., 2023).-

Each of these relatively unconnected literatures grapples with the meaning of one common superordinate factor extracted from a group of items with oftentimes different substantive referents. Several interpretations of this superordinate factor have been put forward by researchers in each area (e.g., Caspi & Moffitt, 2018; Oltmanns et al., 2018; Revelle & Wilt, 2013; Sharp et al., 2015; van der Linden et al., 2010; Watts et al., 2024), but there has been little integration across the three domains (exception: Oltmanns et al., 2018). In addition, the various mechanisms that may bring about a general factor in a dataset (see above) are not systematically accounted for in most of the respective studies. Understanding these mechanisms, however, including the ways in which their operation depends on the specific properties of a study's research design and data, is a necessary prerequisite for being able to properly interpret the nature of the general factor.-

Research on General Factors in person judgment originally evolved from higher-order research on the factorial structure of psychopathology (Achenbach, 1966) and personality (Digman, 1997). In terms of conceptual approach, it aligns quite closely with research on "g," the general factor of intelligence (Spearman, 1904). The starting point for that research was Spearman's observation of a "positive manifold" in intelligence testing data: Persons who were better at solving certain kinds of tasks tended to be better at solving other kinds of tasks, as well. Spearman thus proposed the existence of a common cause (g) accounting for that manifold. In terms of our conceptual analysis, this may be seen as an example of the supposed operation of the PS-DS mechanism. Since then, intelligence research has evolved considerably, and shown that models comprising several unrelated causal factors, or hierarchical factors, tend to fit the data better than a single common factor. More recently, dynamic network explanations (PS-PS mechanism) for why intellectual capacities tend to correlate with one another have been advanced as well (van der Maas et al., 2006).-

While intelligence researchers have the luxury of being able to deal with relatively objective data (i.e., test results), researchers interested in social judgments also have to consider the various additional influences that may shape such judgments apart from the substance. We will now briefly review some of the major contributions to the literature in this field, focusing specifically on whether and how these intervening processes were accounted for, and what the respective outcome was.-

Musek (2007) extracted a general factor of personality (GFP) from self-reported indicators of the five-factor model (FFM) of personality. Across three samples using three separate measures of the FFM, a GFP model had good fit in indices and the GFP explained a significant amount of variance. Moreover, the GFP correlated with life satisfaction, self-esteem, and positive affect, and negatively with negative affect. Musek wrote that the GFP represents "positive

versus-negative-aspects-of-personality”-(p.-1228).-Similar-factors-have-been-extracted-from-self-reports-using-multiple-personality-measures-and-models,-with-some-of-this-research-(e.g.,-Bono-&-Judge,-2003;-Furr-&-Funder,-1998;-Judge-et-al.,-2002)-predating-the-coining-of-the-term-“general-factor-of-personality”.-They-are-usually-found-to-be-associated-with-important,-self-reported-life-outcomes-such-as-job-performance-and-subjective-well-being-(Burns-et-al.,-2017;-Kallio-Strand-et-al.,-2021;-Pelt-et-al.,-2017;-van-der-Linden-et-al.,-2010).-Some-researchers-do-interpret-the-GFP-in-line-with-the-PS-DS-mechanism.-

There-is,-however,-a-sizable-literature-calling-such-substantive-interpretations-of-the-GFP-into-question,-for-various-reasons:-First,-research-has-shown-that-perceiver-attitudes-are-reflected-in-most-person-judgements-and,-accordingly,-in-the-factors-that-may-be-extracted-from-such-judgments.-For-example,-Anusic-et-al.(2009)-demonstrated-the-existence-of-an-evaluative-“halo”-factor-reflecting-the-overall-positivity-of-targets’-self-judgments-independent-of-the-variance-that-is-shared-between-judgments-by-the-targets-and-by-informants.-They-also-showed-that-this-factor-closely-aligns-with-the-targets’-self-esteem,-which-is-basically-the-targets’-attitude-toward-themselves.-

Second,-when-the-GFP-is-modeled-using-multi-source-rather-than-single-source-data,-it-tends-to-disappear,-or-at-least-be-substantially-weakened-(e.g.,-Chang-et-al.,-2012;-Danay-&-Ziegler,-2011;-Revelle-&-Wilt,-2013;-Riemann-&-Kandler,-2010).-For-example,-for-the-research-by-Chang-et-al.-(-2012),-45-independent-samples-from-44-studies-using-self-and-informant-ratings-of-the-same-target-persons-were-collected-and-jointly-analyzed-according-to-the-standards-of-multi-trait-multi-method-analysis-(MTMM;-Campbell-&-Fiske,-1959).-The-focus-lay-on-the-extent-to-which-correlations-between-different-personality-ratings-would-depend-on-whether-those-ratings-are-provided-by-the-same-persons,-or-by-different-persons-(e.g.,-self-vs.-informant).-Hetero-trait-mono-method-correlations-within-informant-ratings-(=same-data-source)-turned-out-to-be-substantially-higher-( $\rho = .16$ )-than-hetero-trait-hetero-method-correlations-between-self-and-informant-ratings-( $\rho = .03$ ),-and-between-different-informants-( $\rho = .04$ )-(=different-data-sources).-This-means-that-most—but-not-all—of-the-variance-that-was-apparently-shared-between-judgements-on-different-attributes-was-only-shared-as-long-as-the-ratings-came-from-the-same-persons,-strongly-suggesting-the-operation-of-rater-specific-biases-(such-as-unique-attitudes-and-formal-response-styles).-

Third,-when-the-GFP-is-modeled-with-items-that-have-been-“neutralized”-in-terms-of-evaluative-content,-the-GFP-is-also-greatly-reduced-in-saturation-(Bäckström-et-al.,-2009).-For-example,-Bäckström-and-Björklund-(2016)-were-able-to-show-that-the-general-factor-of-personality-could-be-significantly-weakened-by-rephrasing-items-in-a-more-neutral/less-evaluative-fashion-(from- $\omega_H = .53$ -to- $\omega_H = .27$ ).-Given-that-the-evaluativeness-of-items-basically-equals-the-extent-to-which-they-reflect-the-perceivers’-attitudes-toward-the-targets-(Leising-et-al.,-2015),-this-finding-also-supports-the-view-that-much-of-the-GFP-variance-is-in-deed-attitude-variance.-

Fourth,-it-has-been-shown-that-items-loading-in-the-same-direction-on-the-GFP-often-seem-to-have-opposite-semantic-references:-Pettersson-et-al.-(-2012)-had-a-sample-of-participants-describe-themselves-using-30-sets-of-“quadruplets”,-each-comprising-four-items-that-were-balanced-in-terms-of-valence-and-content-(cf.-Borkenau-&-Ostendorf,-1989;-Peabody,-1967):-For-each-quadruplet,-items-a-and-b-were-assumed-to-reflect-high-levels-whereas-items-c-and-d-were-assumed-to-reflect-low-levels-of-the-respective-content-dimension.-In-addition,-whereas-items-a-and-c-had-a-positive-evaluative-tone,-items-b-and-d-had-a-negative-evaluative-tone-(e.g.,-a:-“assertive”,-b:-“overbearing”,-c:-“modest”,-d:-“submissive”).-Note-that-these-specifics-directly-concern-the-signs-and-strengths-of-paths-from-the-PS-and-A-variables-to-the-J-variables-in-Figures-5-and-6.-

When-factored,-the-general-factor-that-emerged-clearly-assembled-items-with-similar-valence-but-opposite-semantic-references.-Based-on-this-finding,-Pettersson-et-al.-(-2012)-concluded-that-“it-is-unlikely-that-a-behavioral-style-could-be-characterized-by-so-many-paradoxical-pairs-of-descriptors,-so-this-dimension-may-be-better-interpreted-as-a-response-bias-rather-than-as-a-description-of-a-consistent-behavioral-propensity”-(p.-299).-We-would-argue-that-this-“response-bias”-reflects-the-perceivers’-attitudes-toward-the-targets.-

Taken-together,-all-of-these-findings-consistently-suggest-that-a-substantial-proportion-of-GFP-variance-is-attitudinal-( $A(\pi, \tau)$ )-in-nature.-Recent-research-also-shows-that-the-different-general-factors-overlap-so-strongly-that-they-may-be-largely-indistinguishable-from-one-another-(Oltmanns-et-al.,-2018).-Based-on-their-analyses,-the-same-authors-speculate-that-at-least-some-of-the-general-factor-variance-may-“reflect-extent-of-impairment-or-dysfunction-within-the-respective-persons’-lives,-irrespective-of-the-basis-of-that-dysfunction-or-impairment”.-In-the-language-of-our-model,-that-would-be-the- $A(\tau)$ -component.-

## Relevance for Scale Score Computation<sup>2</sup>

Most-research-in-personality-and-social-psychology-bases-its-analyses-on-the-scores-of-scales-comprising-many-items,-rather-than-on-individual-items.-The-reason-is-that,-by-aggregating-across-items,-one-may-expect-reliability-to-increase,-because-unsystematic-error-is-likely-to-cancel-out-more-across-a-greater-number-of-“parallel”-assessments.-The-primary-approach-to-identifying-items-that-should-go-on-the-same-scale-is,-to-this-day,-factor-analysis.-

Even-if-an-aggregate-score-of-several-correlated-items-tends-to-be-more-reliable,-however,-the-noiseless-(“true”,-“latent”)-scale-score-will-be-highly-ambiguous-in-terms-of-how-it-should-be-interpreted.-Unless-appropriate-methods-for-disentangling-possible-sources-of-covariation-are-employed,-a-scale-score-may-reflect-any-combination-of-the-five-influences-described-above.-Importantly,-the-relative-contribution-of-these-mechanisms-will-depend-directly-on-how-much-their-influences-on-the-individual-items-of-the-scale-are-the-same-and-will-thus-stack.-In-fact,-computing-scale-scores-by-aggregating-across-items-may-sometimes-exacerbate-rather-than-mitigate-the-validity-problem.-

For example, when a self-report scale comprises numerous items asking participants whether they are “good” or “bad” at various things, and the substantive skills that the individual items refer to are uncorrelated, then the overall scale score will basically become a highly reliable measure of the targets’ (= perceivers) self-esteem, with little or no substantive basis anymore (e.g., Leising et al., 2011). This is also relevant for prediction: When a correlation between a predictor and a criterion is high, a common-but-naïve interpretation goes that the substance feeding into the predictor is associated with the substance feeding into the criterion. But when both predictor and criterion contain the same attitude variance more than anything else (e.g., due to aggregation), then this interpretation would miss the mark entirely. Likewise, predictor and criterion may correlate highly with one another because (e.g., due to aggregation) they constitute highly reliable measures of the same thing, only under different names.

## Relevance for Research Using Network Models<sup>2</sup>

Another major strand of research literature to which our conceptual analysis directly applies is the literature on “network models” (including so-called “dynamical systems”). Here, the core idea is that correlations between items are not necessarily attributable to common substantive causes (Mechanism PS-DS), but to how the substance components that are reflected by the items causally affect each other (Mechanism PS-PS). For example, the symptoms of Bulimia Nervosa should not be viewed so much as expressions of some common underlying disposition, but as a network of causally interconnected variables: patients respond to emotional distress with eating binges (S1), which then cause fear of weight gain and feelings of guilt (S2), which then lead to induced vomiting (S3) etc. Such dynamics may exist entirely in the *absence* of any common underlying cause. Over the past decade or so, this approach has gained considerable traction in both personality (Cramer et al., 2012; Wood et al., 2015) and clinical psychology (Borsboom & Cramer, 2013; Robinaugh et al., 2020), as well as their intersection (Lunansky et al., 2020).

Again, the key thing to notice is that this research relies more or less exclusively on judgment data. Therefore, correlation patterns underlying these analyses may not directly be interpreted in terms of causal substantive effects, as they may just as well reflect shared semantic references (J-PS mechanism), and/or shared attitude effects (J-A mechanism), and/or formal response style effects (J-R mechanism).

The J-PS and J-A mechanisms in particular may pose the most trouble for the validity of conclusions drawn from network analyses, because these mechanisms directly affect the *differential* strengths of associations between individual items or scales. Items may differ from one another in terms of how strongly they reflect perceiver attitudes and/or in terms of their (shared) semantic references. These differences alone would suffice to create a differentiated pattern of correlations between items, which may easily be misinterpreted in terms of substantive causality. For example, “self-doubt” may be erroneously viewed as having a

stronger causal influence on “depressivity” than on “hostility”, whereas in reality the former two terms simply share more of their substantive references. This issue is compounded by the fact that most scale items used to estimate networks are traditionally formulated such that they reflect a shared underlying construct, i.e., a latent variable, as opposed to semantically distinct entities.

Moreover, analyses used for network estimation often aim to take dependency structures into account, by (e.g.) calculating partial correlations between items (Epskamp et al., 2018; Epskamp & Fried, 2018). The pairwise relations that are estimated between items this way then reflect their relationship after controlling for all other items in the system. Due to this dependency, influences of mechanisms J-PS, J-A and J-R would threaten not only the validity of the estimated relationship between a given pair of items, but the validity of the estimated relationships between all other items, as well.

The field of network psychometrics borrows concepts from social network analysis (Borgatti et al., 2009), such as analyzing which nodes are most “central” to a given network structure. A node’s *strength* (sometimes referred to as *degree* or *expected influence*), for instance, is a measure of centrality that sums up all of the absolute incoming/outgoing edge weights for a given node in the network. Such measures have gained increasing popularity, especially in clinical psychology (Fried et al., 2016). The rationale is that more central nodes may be predestined to become a focus of therapeutic treatment, because by influencing them one may have the greatest impact on the dynamics of the (maladaptive) network as a whole. While some studies did find predictive effects of centrality estimates for treatment outcome (Elliott et al., 2018), others call for caution in using centrality estimates to identify targets of intervention (Bringmann et al., 2019).

According to the current conceptual analysis, such caution is clearly warranted because patterns of correlations between verbal judgments may reflect different degrees of pair-wise semantic redundancy (Mechanism J-PS; Fried & Cramer, 2017) and/or different degrees of joint evaluation (Mechanism J-A; Leising et al., 2021). As a consequence, the estimated centrality of an item in a network of items may simply be a reflection of how much the semantic references of that item overlap with those of all other items in that network (with the most “central” item being the most redundant with the average other item), and/or of the extent to which that item reflects the same perceiver attitudes that all other items also reflect to some degree (with the most “central item” reflecting those attitudes most strongly).

## Conclusion<sup>2</sup>

Our brief, eclectic review of the literature showed that some of the research on general factors of personality *has* accounted for attitude effects, and shown them to be of considerable strength. However, these findings have been largely ignored so far in the fields of psychopathology and personality pathology. Furthermore, all three strands of general factor literature pay little attention so far to the



problem-of-semantic-redundancy,-and-the- admittedly-less-urgent- problem-of-formal-response-styles. Shared-attitude-effects,-formal-response-styles-and-semantic-redundancy-also-have-been-largely-ignored-as-potential-validity-threats-in-the-network-models-literature. Thus,-the-problem-of-how-to-correctly-interpret-patterns-of-correlations-between-person-judgments-on-different-attributes-is-far-from-having-been-solved.-

## A Review of Methods for Disentangling the Influences of the Five Mechanisms

Most-applied-research-in-(e.g.,-clinical,-I/O,-educational)-psychology-aims-to-assess-the-targets'-substantive-characteristics,-in-order-to-compare-them,-to-correlate-them-with-one-another-or-with-other-variables,-or-to-investigate-how-they-react-to-interventions. The-vast-majority-of-this-research-uses-judgment-data-from-questionnaires-or-interviews.-

Researchers-working-in-these-fields-should-thus-be-aware-of-the-psychological-mechanisms-involved-in-producing-this-type-of-data-(as-described-in-the-current-paper). Before-judgment-data-may-be-used-as-a-basis-for-reasoning-about-phenomena-at-the-substance-level,-all-of-the-other-possible-influences-need-to-be-appropriately-accounted-for. Only-then-will-it-be-appropriate-to-(e.g.)-speculate-about-the-biological-underpinnings-of-a-general-factor-of-personality-(or-psychopathology),-or-to-make-a-particular-variable-in-a-network-of-symptoms-the-focus-of-clinical-attention-due-its-functional-centrality-in-that-network.-

In-the-remainder-of-this-paper,-we-will-review-a-variety-of-methods-that-have-been-used- or-could-be-used- for-disentangling-the-effects-of-the-five-mechanisms-described-above. The-present-paper-is-not-the-first-to-address-these-important-issues. Most-notably,-Podsakoff-et-al.-(2003)-reviewed-the-issue-of-"common-method-bias",-including-possible-ways-of-dealing-with-such-biases-in-practical-research. Our-present-analysis-partly-overlaps-with-theirs,-but-also-goes-beyond-it-in-several-important-respects:-Specifically,-Podsakoff-et-al.-(2003)-did-not-account-for-(1)-the-distinction-between-formative-and-reflective-measurement,-(2)-the-distinction-between-common-causation-and-dynamic-interplay-of-substance-variables,-(3)-some-of-the-most-basic-and-influential-social-desirability-research-(Edwards,-1953,-1957;-McCrae-&-Costa,-1983),-(4)-item-properties-more-broadly,-including-possible-interactions-between-them-and-properties-of-perceivers,-targets,-or-dyads,-(5)-effects-of-aggregation-(across-items,-across-time)-and-the-key-role-that-the-relative-stability-of-all-contributing-factors-plays-in-it,-(6)-the-independence-of-perceiver,-target,-and-dyadic-contributions. All-of-this-is-covered-in-the-present-paper.-

## Collecting Objective Data

If-a-researcher-is-interested-in-phenomena-at-the-substance-level-and-aims-to-avoid-the-possible-pitfalls-that-are-associated-with-using-judgment-data-(see-above),-then-the-most-straightforward-methodological-approach-would-be-to-measure-substance-variables-directly-(i.e.,-unfiltered-by-human-perception). Despite-the-considerably-greater-effort

typically-associated-with-collecting-such-data,-examples-do-exist: Besides-the-aforementioned-intelligence-research,-some-studies-employed-objective-measures-of-participants'-behavior-in-economic-games-(Baumert-et-al.,-2014),-or-monitored-certain-parameters-of-their-smartphone-usage-(Stachl-et-al.,-2020). When-a-correlation-emerges-among-such-assessments,-it-may-not-be-accounted-for-by-semantic-redundancy,-perceiver-attitudes,-or-formal-response-styles,-because-no-human-raters-were-involved-in-obtaining-the-data.-

When-a-PS-DS-mechanism-is-of-interest,-one-would-have-to-show-that-two-or-more-substance-variables-correlate-with-one-another-but-cess-doing-so-once-a-third-substance-variable-is-controlled-for. It-also-needs-to-be-shown-or-at-least-made-plausible-that-variation-in-the-common-cause-*preceded* variation-in-the-other-two-substance-variables. We-are-not-aware-of-any-such-studies-in-personality-research. When-a-DSIV-mechanism-is-of-interest,-one-would-have-to-show-that-changes-in-one-substance-variable-precede-and-predict-changes-in-another-proximal-substance-variable,-while-ideally-holding-everything-else-constant.-

However,-collecting-this-type-of-data-is-often-exceedingly-effortful-and-difficult,-and-sometimes-fraught-with-ethical-problems-(e.g.,-regarding-consent). Therefore,-the-number-of-studies-applying-this-strategy-continues-to-be-small. Moreover,-the-results-tend-to-be-somewhat-ambiguous-in-terms-of-what-they-*mean* psychologically. The-use-of-objective-data-does-not-obviate-the-need-for-researchers-to-somehow-make-sense-of-what-they-find. We-are-therefore-confident-that,-for-the-foreseeable-future,-judgments-that-are-expressed-using-the-natural-language-and-provided-by-human-raters-will-continue-to-play-an-important-role-as-research-data. This-is-even-desirable-because-highly-consequential-person-judgments-in-"the-real-world"-(e.g.,-in-companies,-schools,-therapists'-offices)-will-continue-to-be-of-this-type. Moreover,-the-substance-that-*matters* the-most-to-us-humans-will-often-consist-of-highly-complex-*compounds* of-several-variables-that-are-only-tied-together-in-the-heads-of-perceivers-(i.e.,-in-the-process-of-forming-a-judgment). By-using-judgment-data-as-the-explanans,-it-becomes-possible-to-leave-the-specific-ingredients-of-those-compounds-in-the-"black-box" and-simply-rely-on-the-typical-perceiver's-meaning-making-ability.-

The-presence-of-PS-DS-and-PS-PS-mechanisms-may-still-be-investigated-using-such-data. The-ideal-approach-to-this-would-be-an-experimental-one. Below,-we-describe-an-example-(Orehek-et-al.,-2020). But-in-personality-research,-some-of-the-suspected-causal-influences-(e.g.,-childhood-experiences,-social-status,-genetic-factors)-may-not-be-manipulated. To-corroborate-their-influence,-other-means-of-causal-inference-making-will-be-necessary-(Grosz-et-al.,-2024;-Rohrer,-2018).-

## Using Multi-Source Judgment Data

Another-strategy-is-to-collect-judgment-data-on-the-same-target-persons-from-different-perceivers-(Anusic-et-al.,-2009;-Podsakoff-et-al.,-2003). For-example,-instead-of-correlating-self-reports-on-item-A-with-self-reports-on-item-B,-one-may-correlate-self-reports-on-item-A-with-informant-reports-on-item-B-(or-vice-versa). If-an-association-between-as-

assessments is still found using the latter approach, one may be tempted to interpret it in substantive terms.

Unfortunately, the situation is a bit more complicated than that. Multi-source assessment is in fact likely to rule out an inflation of correlations between items that is due to non-substantive variation *between* perceivers, such as formal ( $R(\pi)$ ) and attitudinal ( $A(\pi)$ ) perceiver effects. It will also rule out an inflation of correlations due to dyadic attitude effects ( $A(\pi\tau)$ ). However, multi-source assessment will *not* do away with the problem of target-based halo ( $A(\tau)$ ): Some targets are liked more by the average perceiver than other targets, either due to the same substantive characteristics that the items at hand also reflect ( $PS(\tau)$ ), or due to other, unknown substantive influences. This may result in an inflation of correlations between assessments, even across data sources. Multi-source assessment will also *not* eliminate the problem of possible semantic redundancy: Even if judgments on attribute A come from one perceiver and judgments on attribute B come from another perceiver, any association that is found between them may still be rooted simply in their being synonyms for the same facts about the targets.

Likewise, one may *average* ratings on the *same* items across data sources and assume that this average will reflect more of the substantive reality (e.g., a common cause) because systematic and unsystematic error should be averaged out (at least to some extent). This is the approach that Funder (1995) advocates. Even with this approach, however, target-based halo ( $A(\tau)$ ) may remain influential. In fact, a highly reliable aggregate of judgments by different perceivers may owe much or even all of its reliability to shared attitude variance. The extent to which this is the case will remain unknown unless one lets the same perceivers describe the same targets on other items that assess the same substance but no attitudes.

## Using Longitudinal Designs

One of the strategies that Podsakoff et al. (2003) advocated to mitigate the influence of “method bias” is the temporal separation of assessments. This, however, will only be helpful to the extent that the bias in question is temporally unstable. For example, there may be fluctuations in perceivers’ moods over time. These would contribute to  $A(\pi)$  and thus introduce some level of correlatedness among judgments that are provided at a given time. When judgments are provided at *different* times, the perceivers’ moods may be expected to have changed in the meantime, making the emergence of a mood-based correlation unlikely. However, to the extent that sources of bias are *stable* over time, the use of longitudinal designs will not help mitigate their influences but may in fact *amplify* them.

For example, most self-assessments reflect the targets’ self-esteem, because most attributes are evaluative (i.e.,  $\beta_{J(\pi,\tau),A(\pi,\tau)} \neq 0$ ). Given that self-esteem is fairly stable (Kuster & Orth, 2013), repeated assessments would still be expected to correlate with one another just based on this one influence (J → A mechanism). An aggregate of several repeated self-assessments on evaluative attributes will also reflect this same influence. To the extent that the *other* an-

fluences on the targets’ self-ratings (e.g., their actual behavior) are *unstable* over time, their effects will average out, thus an aggregate score may come to reflect self-esteem variance more than anything else.

Temporal separation will also not solve the problem of possible semantic redundancy: Even if an interval is inserted between assessments on two different scales, it is still possible that a correlation that emerges between them is simply owed to both scales referencing the same substance.

## Using Experimental Manipulation

There is a broad consensus in the scientific community that experiments provide the strongest evidence for causality. The same holds for research into person judgments. For our present purposes, we have to distinguish between experimental manipulations that target substance variables (i.e.,  $PS(\tau)$  or  $DS(\tau)$ ), and manipulations that target the perceivers’ attitudes toward the targets ( $A(\pi,\tau)$ ). The effects of both types of manipulations may be observed at the level of the judgments ( $J(\pi,\tau)$ ), and our model enables making relatively concise predictions in that regard.

### Manipulating Proximal Substance Information

If judgments do reflect substantive target characteristics to some extent, then manipulating the substantive information that perceivers receive about targets should affect their judgments. Studies of person judgment have used this approach, but mainly in regard to (a) the *overall quantity* of substantive information, and (b) the *modality* by which the substantive information is delivered (e.g., visual and/or auditory channel) (Borkenau et al., 2004; Borkenau & Liebler, 1992; Wiedenroth & Leising, 2020a). The gist of this research is that substantive information does in fact influence person judgments (i.e.,  $\beta_{J(\pi,\tau),PS(\tau)} \neq 0$ ).

However, almost no research has looked directly into the *specific* substance variables that inform judgments on specific attributes. This is likely due to the aforementioned effort that is required to obtain such data. Accordingly, the use of verbal behavior descriptions at a low level of abstraction (see below) is more common. This approach, however, reintroduces possible influences of formal response styles, shared attitudes, and semantic redundancy. Even when ratings of the proximal substance variables and attributes are provided by different raters, semantic redundancy ( $\beta_{J1(\pi,\tau),PS1(\tau)} \cdot \beta_{J2(\pi,\tau),PS1(\tau)}$ ) and target-based halo ( $A(\tau)$ ) may still contribute to a correlation between these measurements.

There is, however, an intriguing possibility that we think has not been used before. It concerns the semantic similarity issue. Rather than having research participants *rate* the pairwise semantic similarity among items, it would be possible to determine it empirically, based on judgments at zero acquaintance. When we let observers judge targets they have never seen before from a single video, the correlations among attributes that emerge will be informative in this regard: When attributes are correlated, we will not know exactly why that is the case. But when they are *not*

correlated, -we-can-be-sure-that-neither-semantic-redundancy-nor-any-of-the-other-mechanisms-can-be-at-play.- Thus, this approach would be suitable to derive sets of semantically non-redundant items. Using these may not only be advantageous for economic reasons, -but-it-would-also-help-“knock-out”-a-prominent-source-of-co-variation-in-judgment-data,-if-one-wants-to-focus-on-one-of-the-other-mechanisms.-To-do-so,-one-would-first-have-to-establish-this-non-redundancy,-and-then-use-the-same-items-in-a-new-set-of-judgment-data-that-might-reflect-one-of-the-other-influences-(e.g.,-one-in-which-an-experimental-manipulation-at-the-substance-level-is-introduced,-see-next-section).-

Manipulations-of-proximal-substance-information-may-be-used-to-determine-the-strength-of-some- $\beta_{A(\pi,\tau),PS(\tau)}$ -paths,-as-there-may-be-chunks-of-substantive-information-that-can-be-shown-to-affect-the-perceivers’-attitudes.-When-these-chunks-are-presented-to-perceivers,-they-should-report-liking-the-respective-targets-more-(or-less).-Their-judgments-of-the-targets-on-any-attribute-should-also-change,-to-the-degree- $\beta_{J(\pi,\tau),A(\pi,\tau)}$ .-Moreover,-all-evaluative-items-should-become-more-strongly-correlated-with-one-another-if-such-substantive-information-about-targets-is-only-presented-to-some-perceivers-(because-this-will-increase-variance-in- $A(\pi,\tau)$ ).-

### **Manipulating Distal Substance Variables I**

Clearly-the-strongest-evidence-for-the-operations-of-PD-DS-and-PS-PS-mechanisms-would-come-from-experiments-in-which-proximal-or-distal-substance-variables-are-experimentally-manipulated-and-then-the-effects-of-these-manipulations-(on-other-substance-variables-or-on-judgments)-are-observed.-This-requires-a-longitudinal-design-with-at-least-two-phases.-

In-a-study-by-Orehek-et-al.-(2020),-targets-were-randomly-assigned-to-consume-alcoholic,-placebo-or-control-drinks-and-then-the-effects-of-this-manipulation-on-observer-judgments-of-the-targets’-behavior-in-the-lab-were-observed.-Using-profile-analyses,-Orehek-et-al.-could-show-that-judgments-of-intoxicated-targets-became-more-positive-overall,-but-not-more-accurate.-Although-correlations-among-individual-items-were-not-investigated-in-this-study,-such-a-design-does-enable-an-actual-investigation-of-a-PS-DS-effect- albeit-one-whose-operation-is-fully-mediated-by-the- $\beta_{A(\pi,\tau),PS(\tau)}$ -paths:-Being-intoxicated-causes-targets-to-be-have-in-different-ways,-and-these-different-behaviors-affect-the-average-perceivers’-attitudes- $A(\tau)$ -positively,-which-is-then-expressed-in-their-judgments-of-the-targets- $J(\pi,\tau)$ -on-any-attribute,-to-the-extent- $\beta_{J(\pi,\tau),A(\pi,\tau)}$ .-

### **Manipulating Perceiver Attitudes I**

It-lies-close-at-hand-that-substantive-and-attitude-effects-in-person-judgment-may-be-disentangled-by-attempting-to-influence-the-latter-while-holding-the-former-constant.-All-three-attitude-components-may-become-the-target-of-such-experimental-treatment.-Influencing- $A(\pi)$ -may-be-the-easiest-to-accomplish.-For-example,-before-the-perceivers-in-a-study-get-to-learn-anything-about-their-targets,-they-may-have-a-good-meal,-listen-to-nice-music,-be-complimented-on-

their-recent-achievements-or-good-looks,-etc.-Influencing- $A(\tau)$ -is-more-difficult,-because-it-requires-that-perceivers-develop-certain-attitudes-towards-specific-targets,-independent-of-the-substantive-information-that-they-also-receive-about-them.-In-a-study-by-Zimmermann-et-al.-(2018),-perceivers-took-part-in-a-sham-session-during-which-their-behavior-in-a-lab-task-was-recorded.-Later,-they-were-presented-with-videos-showing-other-persons-engaging-with-the-same-tasks,-and-asked-to-judge-those-others’-behaviors.-All-perceivers-were-told-that-those-others-had-also-watched-a-few-such-videos,-including-theirs,-and-that-they-had-rated-some-participants-more-negatively-than-others.-Participants-who-were-told-that-the-more-negative-ratings-concerned-them-~~strongly~~-reciprocated-by-rating-the-respective-targets-accordingly.-Although-it-was-not-tested-in-the-study,-this-manipulation-would-be-expected-to-introduce-additional-attitude-variance-into-the-data,-which-in-turn-should-increase-correlations-among-items-in-accordance-with-the-items’-social-desirabilities.- $A(\tau)$ -variance-may-be-experimentally-increased-by-giving-all-perceivers-a-reason-to-like-certain-targets-more-than-others.-

## **Manipulating Item Characteristics 2**

### **Using Items Low in Evaluativeness I**

Given-that-the-social-desirability-of-attributes-moderates-the-influence-of-attitudes-on-judgments-more-or-less-completely,-one-may-attempt-to-eliminate-the-influence-of-attitudes-by-reducing-the-items’-evaluativeness-while-still-targeting-the-same-substantive-dimensions.-This-was-successfully-attempted-by-Bäckström-et-al.-(2009)-who-demonstrated-that-a-set-of-more-neutrally-phrased-items-did-produce-a-much-weaker-general-factor.-Later-research-from-the-same-group-showed-that-this-is-possible-without-significant-losses-in-external-validity-(Bäckström-et-al.,-2014;-Kallio-Strand-et-al.,-2021).-

### **Using Items Low in Abstractness I**

The-direct-measurement-of-substance-variables-tends-to-be-very-effortful-and-difficult-(see-above).-As-a-proxy,-one-may-use-verbal-ratings-on-attributes-that-are-low-in-abstractness.-This-has-been-done-in-studies-within-the-Act4-Frequency4-Approach4(Borkenau,-1986;-Buss-&-Crak,-1981,-1983).-Items-can-be-reliably-judged-in-terms-of-how-broad-vs.-narrow-their-range-of-behavioral-referents-is,-and-broader-ones-are-uniquely-tied-to-narrower-descriptions-of-specific-behaviors-(Hampson-et-al.,-1986;-Wiedenroth-&-Leising,-2020b).-However,-even-verbal-descriptions-that-are-low-in-abstractness-still-constitute-judgments.-Thus,-the-question-of-how-much-correlations-between-them-and-other-judgments-reflect-the-J-PS,-J-A-and-J-R-mechanisms-is-an-empirical-one.-One-might-expect-that-semantic-redundancy-among-items-should-be-lower-at-lower-levels-of-abstraction,-which-should-in-turn-reduce-the-influence-of-the-J-PS-mechanism-on-correlations.-But-this-has-not-yet-been-empirically-investigated,-according-to-our-knowledge.-

## Using Items That Are Semantically Independent

Ratings of pairwise semantic similarity between items have been used in personality research for quite some time. Most notably, D'Andrade (1965) was able to show that such ratings could predict the outcome of factor analyses almost perfectly, making it unnecessary to collect actual ratings of targets by perceivers on the same items. This strongly suggests that the semantic similarity among items should be routinely accounted for in all analyses of inter-item correlation patterns. Given that D'Andrade used relatively abstract items in his analyses, it may also be advisable to experiment with using less abstract ones (i.e., items referencing fewer substantive characteristics) and see how well these findings hold up.

## Balancing Items Within Scales

Aggregating judgments across the items of a scale may result in scale scores that confound the substance one wants to measure with the perceivers' attitudes toward the targets and/or with the perceivers' tendencies to agree with any statement (acquiescence). As a remedy, psychometricians have constructed scales that were *balanced* such that (a) half of the items described the presence of something and the other half described its absence (e.g., Soto & John, 2017), or that (b) half of the items described something (e.g., a target's trait) positively and the other half described it negatively. The feasibility of the latter approach again depends on the extent to which the same substance variation may be described in equally positive or negative terms.

## Using Forced-Choice Response Formats

Forced-choice response formats have been used in personality research for decades. Obviously, doing so enables researchers to effectively eliminate the influences of formal response styles such as acquiescence (Soto & John, 2017) and extreme responding (e.g., Block, 2008). However, forced-choice formats may also be used to disentangle substance and attitudes. For example, in the so-called Sets of Four approach (Borkenau & Ostendorf, 1992), raters are presented with two pairs of bipolar items and asked to decide which pole describes the target better on each item. The first item contains a combination of low substance and positive evaluation on the one end (e.g., "relaxed") and a combination of high substance and negative evaluation on the other end (e.g., "tense"). The second item contains a combination of high substance and positive evaluation on the one end (e.g., "energetic") and a combination of low substance and negative evaluation on the other end (e.g., "lazy"). By comparing raters' responses to such item pairs, it becomes possible to determine the extent to which responses are driven by substance (endorsing either "energetic" and "tense" or

"relaxed" and "lazy"), or by attitude (endorsing either "energetic" and "relaxed" or "tense" and "lazy"). The logic behind this approach is basically the same as in the factor-analytic study by Pettersson et al. (2012). These authors concluded that the general factor was largely attitude-based, given that highly-loading items on the factor were highly diverse in terms of content, but highly consistent in terms of evaluation (see above).

## Assessing Properties of Perceivers, Targets, and Dyads

For decades, psychologists have been grappling with the question of how to measure and control those "other" sources of variance that may affect person judgments, apart from the targets' substance levels (see Podsakoff et al., 2003, for an earlier review). Various methods for capturing attitudes and formal response styles have been tried.

## Using Direct Assessments

One common example of a direct attitude ( $A(\pi, \tau)$ ) assessment is asking perceivers how much they like their respective targets. This information may then be used to predict the same perceivers' ratings of the same targets on a set of attributes. Leising et al. (2010) showed that judgments on Big-Five attributes were partly predictable this way and that, in line with the model presented here, this predictability aligned very closely with the rated social desirability of the items. Also in line with the model, Heynicke et al. (2022) showed that estimated perceiver effects in positivity ( $A(\pi)$ ) aligned relatively well ( $r = .55$ ) with how much perceivers said they liked the average target.

## Using Random Content

Another way of assessing these sources of variance is to present perceivers with many items that vary widely in content and then determine the extent to which they (a) choose positive descriptions over negative ones (as a measure of attitudes), and/or (b) agree with all statements (as a measure of acquiescence), and/or (c) vary their responses across items (as a measure of extreme responding) regardless of the items' more specific content (e.g., Baird et al., 2017).<sup>7</sup>

The classic "lie-scales" or "social-desirability-scales" exemplify this for the assessment of attitudes. Here, perceivers are presented with all sorts of positive and negative statements about a target, and patterns of responses that consistently agree with the former and disagree with the latter are interpreted as being "too good to be true" ("self-enhancement"). Empirical research into this approach has yielded fairly disappointing results, however: If judgments are in fact "contaminated" with attitudes, then partialling

7- Note again that perceiver differences in response variability are not covered by the version of the model that is presented here. They may, however, easily be added as random effects on the slope of the path leading from  $J$  to  $M$ .

out an attitude measure should improve accuracy. Most attempts at demonstrating this with social desirability scales have failed, however (Borkenau & Ostendorf, 1992). Paunonen and LeBel (2012) explained this in terms of the relatively low saturation of typical items with social desirability. In several studies, inter-rater agreement (which was used as a proxy for accuracy) even *dropped* when participants' scores on social desirability scales were partialled out. McCrae and Costa (1983) explained this in terms of the fact that some targets do in fact evoke more positive overall impressions than other targets ( $A(\tau)$ ), thus partialling out social desirability scale scores must have this effect.

### Using Confirmatory Factor Analysis for Multilevel Profile Analysis

The most comprehensive methodological approaches to dealing with the issues highlighted in the present paper are Confirmatory Factor Analysis (CFA) and Multilevel Profile Analyses (MPA). They make it possible to simultaneously account for variation at the perceiver, target and dyad level, and for variation at the level of items. When the item sample is diverse and representative, one may use these methods to derive conclusions that are valid for *typical* items, not just for a given and often rather specific selection of items. That way, both approaches may help improve on generalizability.

According to our knowledge, Anusic et al. (2009) were the first to incorporate an attitude ("halo") factor into their factor analyses of multi-source judgment data. To do so, one only needs to specify a factor that will have positive (negative) loadings on all items with positive (negative) valence. Likewise, an acquiescence factor may be specified on which all affirmative statements load in the same direction, regardless of any other content they may also have (Anusic et al., 2009; Maydeu-Olivares & Coffman, 2006).

Multilevel Profile Analysis does something very similar, with the exception that here the predictors at the perceiver, target and dyad levels are not estimated post-hoc from the data, but measured independently. Several variables characterizing the same units of observation may be used in parallel as predictors (e.g., how well the perceivers say they *know* and *like* the targets). Item properties (e.g. social desirability ratings) may also be incorporated, as separate profiles. Parallel use of various item level profiles is possible and enables a disentangling of effects such as normative substance (i.e., base rate), distinctive substance, and attitudes (Biesanz, 2010; Wessels et al., 2020). Acquiescence may be represented by an intercept.

## Overall Conclusion

We hope that the present paper demonstrates how much clarity and efficiency may be gained by way of a thorough conceptual analysis. When the subject matter is as complex as person judgment, having a relatively lean conceptual model may go a long way in helping us not get lost in that complexity. For example, the last part of the current paper showed how a large variety of methodological approaches basically attempts to tackle the same rather limited set of interrelated problems. As long as theoretical considerations are only expressed in the natural language, the extent to which this is the case will remain opaque, because even when two researchers use the exact same term (e.g., "social desirability"), they may still not mean the same thing.

However, the third part of the present paper also shows cases how much complexity remains even after a clarifying formal analysis. In fact, a proper disentangling of the various mechanisms that may account for correlations among measurements may more often than not require a *combination* of several approaches. Based on our conceptual analysis and review of the evidence, we conclude that it will almost always be helpful to assess (a) the perceivers' attitudes toward the targets, (b) the social desirability of the items, and (c) the semantic redundancy among all items. Acquiescence should be controlled for. If we are dealing with judgment data, causal analyses pertaining to PS-DS or PS-PS mechanisms only become feasible once these competing influences have been properly accounted for.

## Competing Interests

The authors state that neither of them has a conflict of interest associated with a potential publication of this manuscript.

## Author Contributions

Conceptualization: DL, MBo. Formal Analysis: DL, MBo. Writing - Original Draft: DL.

Writing - Review & Editing: DL, BC, JZ, NF, AW, MBä, MBo, JB, JO. Visualization: DL, MBo, PK.

Submitted: May-05, 2020-PDT. Accepted: March-20, 2025-PDT.

Published: April-17, 2025-PDT.



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-4.0). View this license's legal deed at <http://creativecommons.org/licenses/by/4.0> and legal code at <http://creativecommons.org/licenses/by/4.0/legalcode> for more information.



## References2

- Achenbach, T. M. (1966). The classification of children's psychiatric symptoms: A factor-analytic study. *Psychological Monographs: General and Applied*, 80(7), 1–37. <https://doi.org/10.1037/h0093906>
- Anderson, N. H. (1968). Likeableness ratings of 555 personality-trait words. *Journal of Personality and Social Psychology*, 9, 272–279. <https://doi.org/10.1037/h0025907>
- Anusic, I., Schimmack, U., Pinkus, R. T., & Lockwood, P. (2009). The nature of correlations among big-five ratings: The Halo-Alpha-Beta-Model. *Journal of Personality and Social Psychology*, 97, 1142–1156. <https://doi.org/10.1037/a0017159>
- Arnulf, J. K., Larsen, K. R., Martinsen, O. L., & Bong, C. H. (2014). Predicting survey responses: how and why semantics shape survey statistics on organizational behavior. *PLOS ONE*, 9(9), e106361. <https://doi.org/10.1371/journal.pone.0106361>
- Ashton, M. C., Lee, K., Goldberg, L. R., & de Vries, R. E. (2009). Higher-Order Factors of Personality: Do They Exist? *Personality and Social Psychology Review*, 13(2), 79–91. <https://doi.org/10.1177/1088868309338467>
- Bäckström, M., & Björklund, F. (2016). Is the general factor of personality inventories based on evaluative responding? *Personality and Individual Differences*, 96, 31–35. <https://doi.org/10.1016/j.paid.2016.02.058>
- Bäckström, M., Björklund, F., & Larsson, M. R. (2009). Five-factor inventories have a major general factor related to social desirability which can be reduced by framing items neutrally. *Journal of Research in Personality*, 43, 335–344. <https://doi.org/10.1016/j.jrp.2008.12.013>
- Bäckström, M., Björklund, F., & Larsson, M. R. (2014). Criterion validity is maintained when items are evaluatively neutralized: evidence from a full-scale five-factor model inventory. *European Journal of Personality*, 28, 620–633. <https://doi.org/10.1002/per.1960>
- Baird, B. M., Lucas, R. E., & Donnellan, M. B. (2017). The role of response styles in the assessment of intraindividual personality variability. *Journal of Research in Personality*, 69, 170–179. <https://doi.org/10.1016/j.jrp.2016.06.015>
- Baumert, A., Schlösser, T., & Schmitt, M. (2014). Economic games: A performance-based assessment of fairness and altruism. *European Journal of Psychological Assessment*, 30(3), 178–192. <https://doi.org/10.1027/1015-5759/a000183>
- Biesanz, J. C. (2010). The social accuracy model of interpersonal perception: Assessing individual differences in perceptive and expressive accuracy. *Multivariate Behavioral Research*, 45(5), 853–885. <https://doi.org/10.1080/00273171.2010.519262>
- Bishop, G., Spence, S. H., & McDonald, C. (2003). Can parents and teachers provide a reliable and valid report of behavioral inhibition? *Child Development*, 74(6), 1899–1917. <https://doi.org/10.1046/j.1467-8624.2003.00645.x>
- Block, J. (2008). *The Q-sort in character appraisal: 4 Encoding subjective impressions of persons 4 quantitatively*. American Psychological Association.
- Bono, J. E., & Judge, T. A. (2003). Core self-evaluations: A review of the trait and its role in job satisfaction and job performance. *European Journal of Personality*, 17, 5–18. <https://doi.org/10.1002/per.481>
- Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network analysis in the social sciences. *Science*, 323(5916), 892–895. <https://doi.org/10.1126/science.1165821>
- Borkenau, P. (1986). Toward an understanding of trait interrelations: Acts as instances for several traits. *Journal of Personality and Social Psychology*, 51(2), 371–381. <https://doi.org/10.1037/0022-3514.51.2.371>
- Borkenau, P., & Liebler, A. (1992). Trait inferences: Sources of validity at zero acquaintance. *Journal of Personality and Social Psychology*, 62(4), 645–657. <https://doi.org/10.1037/0022-3514.62.4.645>
- Borkenau, P., Mauer, N., Riemann, R., Spinath, F. M., & Angleitner, A. (2004). Thin Slices of Behavior as Cues of Personality and Intelligence. *Journal of Personality and Social Psychology*, 86(4), 599–614. <https://doi.org/10.1037/0022-3514.86.4.599>
- Borkenau, P., & Ostendorf, F. (1989). Descriptive consistency and social desirability in self and peer reports. *European Journal of Personality*, 3, 31–45. <https://doi.org/10.1002/per.2410030105>
- Borkenau, P., & Ostendorf, F. (1992). Social desirability scales as moderator and suppressor variables. *European Journal of Personality*, 6(3), 199–214. <https://doi.org/10.1002/per.2410060303>
- Borsboom, D., & Cramer, A. O. J. (2013). Network analysis: An integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology*, 9, 91–121. <https://doi.org/10.1146/annurev-clinpsy-050212-185608>
- Brandt, V., Zhang, Y., Carr, H., Golm, D., Correll, C. U., Arrondo, G., Firth, J., Hassan, L., Solmi, M., & Cortese, S. (2023). First evidence of a general disease (“d”) factor, a common factor underlying physical and mental illness. *World Psychiatry*, 22, 335–337. <https://doi.org/10.1002/wps.21097>
- Bringmann, L. F., Elmer, T., Epskamp, S., Krause, R. W., Schoch, D., Wichers, M., ... Snippe, E. (2019). What do centrality measures measure in psychological networks? *Journal of Abnormal Psychology*, 128(8), 892–903. <https://doi.org/10.1037/abn0000446>
- Brunswick, E. (1956). *Perception and the representative 4 design of psychological experiments*. University of California Press. <https://doi.org/10.1525/9780520350519>
- Buntins, M. (2014). *Psychologische Tests und mehrwertige 4 Logik - Ein alternativer Ansatz zur Quantifizierung 4 psychologischer Konstrukte*. Springer VS. <https://doi.org/10.1007/978-3-658-05507-3>

- Buntins, M., Buntins, K., & Eggert, F. (2016). Psychological tests from a (fuzzy-)logical point-of-view. *Quality and Quantity*, 50(6), 2395–2416. <https://doi.org/10.1007/s11135-015-0268-z>
- Burns, G. N., Morris, M. B., Periard, D. A., LaHuis, D., Flannery, N. M., Carretta, T. R., & Roebke, M. (2017). Criterion-related validity of a Big Five general factor of personality from the TIPI to the IPIP. *International Journal of Selection and Assessment*, 25, 213–222. <https://doi.org/10.1111/ijsa.12174>
- Buss, D. M., & Craik, K. H. (1981). The act-frequency-analysis-of-interpersonal-dispositions: Aloofness, gregariousness, dominance and submissiveness. *Journal of Personality*, 49(2), 175–192. <https://doi.org/10.1111/j.1467-6494.1981.tb00736.x>
- Buss, D. M., & Craik, K. H. (1983). The act-frequency-approach-to-personality. *Psychological Review*, 90(2), 105–126. <https://doi.org/10.1037/0033-295X.90.2.105>
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81–105. <https://doi.org/10.1037/h0046016>
- Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., ... Moffitt, T. E. (2014). The p-factor: one general psychopathology factor in the structure of psychiatric disorders? *Clinical Psychological Science*, 2, 119–137. <https://doi.org/10.1177/2167702613497473>
- Caspi, A., & Moffitt, T. E. (2018). All for one and one for all: Mental disorders in one dimension. *American Journal of Psychiatry*, 175, 831–844. <https://doi.org/10.1176/appi.ajp.2018.17121383>
- Chang, L., Connelly, B. S., & Geeza, A. A. (2012). Separating method factors and higher-order traits of the Big Five: A meta-analytic multitrait-multimethod approach. *Journal of Personality and Social Psychology*, 102, 408–426. <https://doi.org/10.1037/a0025559>
- Chmielewski, M., & Watson, D. (2009). What is being assessed and why it matters: the impact of transient error on trait research. *Journal of Personality and Social Psychology*, 97(1), 186–202. <https://doi.org/10.1037/a0015618>
- Cramer, A. O. J., van der Sluis, S., Noordhof, A., Wichers, M., Geschwind, N., Aggen, S. H., Kendler, K. S., & Borsboom, D. (2012). Dimensions of Normal Personality as Networks in Search of Equilibrium: You Can't Like Parties if You Don't Like People. *European Journal of Personality*, 26, 414–431. <https://doi.org/10.1002/per.1866>
- Danay, E., & Ziegler, M. (2011). Is there really a single factor of personality? A multirater approach to the apex of personality. *Journal of Research in Personality*, 45(6), 560–567. <https://doi.org/10.1016/j.jrp.2011.07.003>
- D'Andrade, R. G. (1965). Trait psychology and componential analysis. *American Anthropologist*, 67(5), 215–228. <https://www.jstor.org/stable/668763>
- Davies, S. E., Connelly, B. S., Ones, D. S., & Birkland, A. S. (2015). The General Factor of Personality: The “Big One,” a self-evaluative trait, or a methodological gnat that won't go away? *Personality and Individual Differences*, 81, 13–22. <https://doi.org/10.1016/j.paid.2015.01.006>
- Digman, J. M. (1997). Higher-order factors of the Big Five. *Journal of Personality and Social Psychology*, 73(6), 1246–1256. <https://doi.org/10.1037/0022-3514.73.6.1246>
- Dumas, J. E., Johnson, M., & Lynch, A. M. (2002). Likableness, familiarity, and frequency of 844 person-descriptive words. *Personality and Individual Differences*, 32, 523–531. [https://doi.org/10.1016/S0191-8869\(01\)00054-X](https://doi.org/10.1016/S0191-8869(01)00054-X)
- Edwards, A. L. (1953). The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *Journal of Applied Psychology*, 37, 90–93. <https://doi.org/10.1037/h0058073>
- Edwards, A. L. (1957). *The social desirability variable in personality assessment and research*. The Dryden Press.
- Eid, M., & Diener, E. (2004). Global judgments of subjective well-being: Situational variability and long-term stability. *Social Indicators Research*, 65(3), 245–277.
- Elliott, H., Jones, P. J., & Schmidt, U. (2018). Central symptoms predict post-treatment outcomes and clinical impairment in anorexia nervosa: A network analysis. *Clinical Psychological Science*, 8(1), 139–154. <https://doi.org/10.1177/2167702619865958>
- Epskamp, S., Borsboom, D., & Fried, E. I. (2018). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods*, 50(1), 195–212. <https://doi.org/10.3758/s13428-017-0862-1>
- Epskamp, S., & Fried, E. I. (2018). A tutorial on regularized partial correlation networks. *Psychological Methods*, 23(4), 617–634. <https://doi.org/10.1037/met0000167>
- Fried, E. I., & Cramer, A. O. (2017). Moving forward: challenges and directions for psychopathological network theory and methodology. *Perspectives on Psychological Science*, 12(6), 999–1020. <https://doi.org/10.1177/17456916177058>
- Fried, E. I., Epskamp, S., Nesse, R. M., Tuerlinckx, F., & Borsboom, D. (2016). What are “good” depression symptoms? Comparing the centrality of DSM and non-DSM symptoms of depression in a network analysis. *Journal of Affective Disorders*, 189, 314–320. <https://doi.org/10.1016/j.jad.2015.09.005>
- Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, 102(4), 652–670. <https://doi.org/10.1037/0033-295X.102.4.652>
- Funder, D. C., & Dobroth, K. M. (1987). Differences between traits: Properties associated with interjudge agreement. *Journal of Personality and Social Psychology*, 52(2), 409–418. <https://doi.org/10.1037/0022-3514.52.2.409>

- Furr, R. M., & Funder, D. C. (1998). A multimodal analysis of personal negativity. *Journal of Personality and Social Psychology*, 74, 1580–1591. <https://doi.org/10.1037/0022-3514.74.6.1580>
- Garcia, D., Rosenberg, P., Al-Nima, A., Granjard, A., Cloninger, K. M., & Sikström, S. (2020). Validation of two short-personality-inventories using self-descriptions in natural language and quantitative semantics test theory. *Frontiers in Psychology*, 11, 16. <https://doi.org/10.3389/fpsyg.2020.00016>
- Gräf, M., & Unkelbach, C. (2016). Halo-effects in trait assessment depend on information valence: Why being honest makes you industrious, but lying does not make you lazy. *Personality and Social Psychology Bulletin*, 42, 290–310. <https://doi.org/10.1177/0146167215627137>
- Grosz, M. P., Ayaita, A., Arslan, R. C., Buecker, S., Ebert, T., Hünermund, P., Müller, S. R., Rieger, S., Zapko-Wilmes, A., & Rohrer, J. M. (2024). Natural experiments: missed opportunities for causal inference in psychology. *Advances in Methods and Practices in Psychological Science*, 7(1). <https://doi.org/10.1177/25152459231218610>
- Hampson, S. E., John, O. P., & Goldberg, L. R. (1986). Category breadth and hierarchical structure in personality: Studies of asymmetries in judgments of trait implications. *Journal of Personality and Social Psychology*, 51, 37–54. <https://doi.org/10.1037/0022-3514.51.1.37>
- Heynicke, M., Rau, R., Leising, D., Wessels, N., & Wiedenroth, A. (2022). Perceiver Effects in Person-Perception Reflect Acquiescence, Positivity, and Trait-Specific Content: Evidence From a Large-Scale Replication Study. *Social Psychological and Personality Science*, 13(4), 839–848. <https://doi.org/10.1177/19485506211039101>
- Hörz-Sagstetter, S., Volkert, J., Rentrop, M., Benecke, C., Gremaud-Heitz, D. J., Unterrainer, H. F., Schauenburg, H., Seidler, D., Buchheim, A., Doering, S., Feil, M. G., Clarkin, J. F., Dammann, G., & Zimmermann, J. (2021). A bifactor model of personality organization. *Journal of Personality Assessment*, 103, 149–160. <https://doi.org/10.1080/00223891.2019.1705463>
- Irwing, P. (2013). The general factor of personality: Substance or artefact? *Personality and Individual Differences*, 55, 234–242. <https://doi.org/10.1016/j.paid.2013.03.002>
- John, O. P., & Robins, R. W. (1993). Determinants of interjudge agreement on personality traits: The Big Five domains, observability, evaluativeness, and the unique perspective of the self. *Journal of Personality*, 61(4), 521–551. <https://doi.org/10.1111/j.1467-6494.1993.tb00781.x>
- Judge, T. A., Erez, A., Bono, J. E., & Thoresen, C. J. (2002). Are measures of self-esteem, neuroticism, locus of control, and generalized self-efficacy indicators of a common core construct? *Journal of Personality and Social Psychology*, 83, 693–710. <https://doi.org/10.1037/0022-3514.83.3.693>
- Kallio-Strand, K., Bäckström, M., & Björklund, F. (2021). Accounting for the evaluative factor in self-ratings provides a more accurate estimate of the relationship between personality traits and well-being. *Journal of Research in Personality*, 93. <https://doi.org/10.1016/j.jrp.2021.104120>
- Kenny, D. A. (1994). *Interpersonal perception. A social relations analysis*. The Guilford Press.
- Kenny, D. A. (2004). PERSON: A general model of interpersonal perception. *Personality and Social Psychology Review*, 8, 265–280. [https://doi.org/10.1207/s15327957pspr0803\\_3](https://doi.org/10.1207/s15327957pspr0803_3)
- Kuster, F., & Orth, U. (2013). The long-term stability of self-esteem: its time-dependent decay and nonzero asymptote. *Personality and Social Psychology Bulletin*, 39(5), 677–690. <https://doi.org/10.1177/0146167213480189>
- Leising, D., Erbs, J., & Fritz, U. (2010). The letter-of-recommendation effect in informant ratings of personality. *Journal of Personality and Social Psychology*, 98, 668–682. <https://doi.org/10.1037/a0018771>
- Leising, D., Krause, S., Köhler, D., Hinsen, K., & Clifton, A. (2011). Assessing interpersonal functioning: Views from within and without. *Journal of Research in Personality*, 45(6), 631–641. <https://doi.org/10.1016/j.jrp.2011.08.011>
- Leising, D., Ostrovski, O., & Borkenau, P. (2012). Vocabulary for describing disliked persons is more differentiated than vocabulary for describing liked persons. *Journal of Research in Personality*, 46, 393–396. <https://doi.org/10.1016/j.jrp.2012.03.006>
- Leising, D., Ostrovski, O., & Zimmermann, J. (2013). Are we talking about the same person here? Inter-rater agreement in judgments of personality varies dramatically with how much the perceivers like the targets. *Social Psychological and Personality Science*, 4, 468–474. <https://doi.org/10.1177/1948550612462414>
- Leising, D., Scharloth, J., Lohse, O., & Wood, D. (2014). What types of terms do people use when describing an individual's personality? *Psychological Science*, 25(9), 1787–1794. <https://doi.org/10.1177/0956797614541285>
- Leising, D., Scherbaum, S., Locke, K., & Zimmermann, J. (2015). A model of “substance” and “evaluation” in person judgments. *Journal of Research in Personality*, 57, 61–71. <https://doi.org/10.1016/j.jrp.2015.04.002>
- Leising, D., & Schilling, R. L. (2025). A mathematical model of person judgment part 1: Cue emergence. *Personality Science*, 6. <https://doi.org/10.1177/27000710241291543>
- Leising, D., Vogel, D., Waller, V., & Zimmermann, J. (2021). Correlations between person-descriptive items are predictable from the product of their midpoint-centered social desirability values. *European Journal of Personality*, 35(5), 667–689. <https://doi.org/10.1177/0890207020962331>
- Leistner, M., Hommel, B. E., Wendt, L. P., & Leising, D. (2024). *Properties of Person Descriptors in the Natural German Language: A Preregistered Replication and Extension*. <https://doi.org/10.31234/osf.io/s8h7u>



- Lunansky, G., van Borkulo, C., & Borsboom, D. (2020). Personality, Resilience, and Psychopathology: A Model for the Interaction between Slow and Fast Network Processes in the Context of Mental Health. *European Journal of Personality*, 34(6), 969–987. <https://doi.org/10.1002/per.2263>
- Maydeu-Olivares, A., & Coffman, D. L. (2006). Random intercept item factor analysis. *Psychological Methods*, 11(4), 344–362. <https://doi.org/10.1037/1082-989X.11.4.344>
- McCrae, R. R., & Costa, P. T. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology*, 51(6), 882–888. <https://doi.org/10.1037/0022-006X.51.6.882>
- Musek, J. (2007). A general factor of personality: Evidence for the Big One in the five-factor model. *Journal of Research in Personality*, 41, 1213–1233. <https://doi.org/10.1016/j.jrp.2007.02.003>
- Nisbett, R. E., & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology*, 35, 250–256. <https://doi.org/10.1037/0022-3514.35.4.250>
- Noor, F., & Evans, D. C. (2003). The effect of facial symmetry on perceptions of personality and attractiveness. *Journal of Research in Personality*, 37(4), 339–347. [https://doi.org/10.1016/S0092-6566\(03\)00022-9](https://doi.org/10.1016/S0092-6566(03)00022-9)
- Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology*, 3(1), 1–18. [https://doi.org/10.1016/0022-2496\(66\)90002-2](https://doi.org/10.1016/0022-2496(66)90002-2)
- Oltmanns, J. R., Smith, G. T., Oltmanns, T. F., & Widiger, T. A. (2018). General factors of psychopathology, personality, and personality disorder: Across domain comparisons. *Clinical Psychological Science*, 6, 581–589. <https://doi.org/10.1177/2167702617750150>
- Orehek, E., Human, L. J., Sayette, M. A., Dimoff, J. D., Winograd, R. P., & Sher, K. J. (2020). Self-Expression While Drinking Alcohol: Alcohol Influences Personality Expression During First Impressions. *Personality and Social Psychology Bulletin*, 46(1), 109–123. <https://doi.org/10.1177/0146167219843933>
- Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. University of Illinois Press.
- Paunonen, S. V., & LeBel, E. P. (2012). Socially desirable responding and its elusive effects on the validity of personality assessments. *Journal of Personality and Social Psychology*, 103(1), 158–175. <https://doi.org/10.1037/a0028165>
- Peabody, D. (1967). Trait inferences: Evaluative and descriptive aspects. *Journal of Personality and Social Psychology*, 7, 1–18. <https://doi.org/10.1037/h0025230>
- Pelt, D. H., van der Linden, D., Dunkel, C. S., & Born, M. P. (2017). The general factor of personality and job performance: Revisiting previous meta-analyses. *International Journal of Selection and Assessment*, 25, 333–346. <https://doi.org/10.1111/ijsa.12188>
- Pettersson, E., Turkheimer, E., Horn, E. E., & Menatti, A. R. (2012). The general factor of personality and evaluation. *European Journal of Personality*, 26, 292–302. <https://doi.org/10.1002/per.839>
- Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88, 879–903. <https://doi.org/10.1037/0021-9010.88.5.879>
- Rau, R., Carlson, E. N., Back, M. D., Barranti, M., Gebauer, J. E., Human, L. J., Leising, D., & Nestler, S. (2021). What is the structure of perceiver effects? On the importance of global positivity and trait specificity across personality domains and judgment contexts. *Journal of Personality and Social Psychology*, 120(3), 745–764. <https://doi.org/10.1037/pspp0000278>
- Revelle, W., & Wilt, J. (2013). The general factor of personality: A general critique. *Journal of Research in Personality*, 47, 493–504. <https://doi.org/10.1016/j.jrp.2013.04.012>
- 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, 258–277. <https://doi.org/10.1002/per.7>
- Robinaugh, D. J., Hoekstra, R. H., Toner, E. R., & Borsboom, D. (2020). The network approach to psychopathology: a review of the literature 2008–2018 and an agenda for future research. *Psychological Medicine*, 50(3), 353–366. <https://doi.org/10.1017/S0033291719003404>
- Rohrer, J. M. (2018). Thinking clearly about correlations and causation: graphical causal models for observational data. *Advances in Methods and Practices in Psychological Science*, 1(1), 27–42. <https://doi.org/10.1177/2515245917745629>
- Schmittmann, V. D., Cramer, A. O. J., Waldorp, L. J., Epskamp, S., Kievit, R. A., & Borsboom, D. (2013). Deconstructing the construct: A network perspective on psychological phenomena. *New Ideas in Psychology*, 31, 43–53. <https://doi.org/10.1016/j.newideapsych.2011.02.007>
- Sharp, C., Wright, A. G., Fowler, J. C., Frueh, B. C., Allen, J. G., Oldham, J., & Clark, L. A. (2015). The structure of personality pathology: Both general ('g') and specific ('s') factors? *Journal of Abnormal Psychology*, 124, 387. <https://doi.org/10.1037/abn0000033>
- Sikström, S., & Garcia, D. (2019). *Statistical semantics –4 methods and applications*. Springer.
- 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. <https://doi.org/10.1037/pspp0000096>
- Spearman, C. (1904). "General Intelligence" objectively determined and measured. *The American Journal of Psychology*, 15, 201–292. <https://doi.org/10.1037/11491-006>

- Srivastava, S., Guglielmo, S., & Beer, J. S. (2010). - Perceiving others' personalities: Examining the dimensionality, assumed similarity to the self, and stability of perceiver effects. *Journal of Personality and Social Psychology*, 98, 520–534. <https://doi.org/10.1037/a0017057>
- Stachl, C., Au, Q., Schoedel, R., Gosling, S. D., Harari, G. M., Buschek, D., Völkel, S. T., Schuwerk, T., Oldemeier, M., Ullmann, T., Hussmann, H., Bischl, B., & Bühner, M. (2020). Predicting personality from patterns of behavior collected with smartphones. *Proceedings of the National Academy of Science*, 117(30), 17680–17687. <https://doi.org/10.1073/pnas.1920484117>
- Steyer, R., Schmitt, M., & Eid, M. (1999). Latent state-trait theory and research in personality and individual differences. *European Journal of Personality*, 13(5), 389–408. [https://doi.org/10.1002/\(SICI\)1099-0984\(199909/10\)13:5%3C389::AID-PER361%3E3.0.CO;2-A](https://doi.org/10.1002/(SICI)1099-0984(199909/10)13:5%3C389::AID-PER361%3E3.0.CO;2-A)
- Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology*, 5, 25–29. <https://doi.org/10.1037/h0071663>
- van der Linden, D., te Nijenhuis, J., & Bakker, A. B. (2010). The general factor of personality: A meta-analysis of Big Five intercorrelations and a criterion-related validity study. *Journal of Research in Personality*, 44, 315–327. <https://doi.org/10.1016/j.jrp.2010.03.003>
- van der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: the positive manifold of intelligence by mutualism. *Psychological Review*, 113(4), 842–861. <https://doi.org/10.1037/0033-295X.113.4.842>
- Vazire, S. (2010). Who knows what about a person? The self–other knowledge asymmetry (SOKA) model. *Journal of Personality and Social Psychology*, 98(2), 281–300. <https://doi.org/10.1037/a0017908>
- Viswesvaran, C., Schmidt, F. L., & Ones, D. S. (2005). Is there a general factor in ratings of job performance? A meta-analytic framework for disentangling substantive and error influences. *Journal of Applied Psychology*, 90(1), 108–131. <https://doi.org/10.1037/0021-9010.90.1.108>
- Watts, A. L., Greene, A. L., Bonifay, W., & Fried, E. I. (2024). A critical evaluation of the p-factor literature. *Nature Reviews Psychology*, 3, 108–122. <https://doi.org/10.1038/s44159-023-00260-2>
- Wessels, N. M., Zimmermann, J., Biesanz, J. C., & Leising, D. (2020). Differential associations of knowing and liking with accuracy and positivity bias in person perception. *Journal of Personality and Social Psychology*, 118(1), 149–171. <https://doi.org/10.1037/pspp0000218>
- West, T. V., & Kenny, D. A. (2011). The truth and bias model of judgment. *Psychological Review*, 118, 357–378. <https://doi.org/10.1037/a0022936>
- Wetzel, E., Lüdtke, O., Zettler, I., & Böhnke, J. R. (2016). The stability of extreme response style and acquiescence over 8 years. *Assessment*, 23(3), 279–291. <https://doi.org/10.1177/1073191115583714>
- Wiedenroth, A., & Leising, D. (2020a). The more the better—but more of which? Information quantity and shared meaning as predictors of consistency and accuracy in person judgment. *Journal of Research in Personality*, 87, Article 103968. <https://doi.org/10.1016/j.jrp.2020.103968>
- Wiedenroth, A., & Leising, D. (2020b). What's in an adjective? An exploratory study of meaning systems. *Journal of Individual Differences*, 41, 152–159. <https://doi.org/10.1027/1614-0001/a000316>
- Wood, D., Harms, P., & Vazire, S. (2010). Perceiver effects as projective tests: what your perceptions of others say about you. *Journal of Personality and Social Psychology*, 99, 174–190. <https://doi.org/10.1037/a0019390>
- Wood, D., Hensler-Gardner, M., & Harms, P. D. (2015). How functionalist and process approaches to behavior can explain trait covariation. *Psychological Review*, 122, 84–111. <https://doi.org/10.1037/a0038423>
- Wood, D., Lowman, G. H., Armstrong, B. F., III, & Harms, P. D. (2023). Using retest-adjusted correlations as indicators of the semantic similarity of items. *Journal of Personality and Social Psychology*, 125(2), 437–454. <https://doi.org/10.1037/pspp0000441>
- Zimmermann, J., Schindler, S., Klaus, G., & Leising, D. (2018). The effect of dislike on accuracy and bias in person perception. *Social Psychological and Personality Science*, 9(1), 80–88. <https://doi.org/10.1177/1948550617703167>

## Appendix2

The following simplifying assumptions are made:-

- 1.- Information-overlap-(Kenny,-1994)-is-100%.-That-is,- judgments-may-not-differ-from-one-another-because-one-perceiver-knows-something-about-a-target-that-the-other-perceiver-does-not-know.-Accordingly,-self-and-other-judgments-are-also-formed-in-exactly-the-same-fashion.-This-explicitly-includes-a-neglect-of-the-fact-that,-in-reality,-people-have-‘privileged-access’-to-certain-kinds-of-substantive-information,such-as-their-own-interoceptions-or-thoughts-(Funder-&-Dobroth,-1987;-Vazire,-2010).-We-also-neglect-the-fact-that-attitudes-contribute-more-variance-to-other-judgments-than-to-self-judgments-(Leising-et-al,-2021).-
- 2.- Observation-time-(i.e.,-the-length-of-the-observation-interval)-is-ignored.-That-is,-when-perceivers-have-access-to-proximal-substantive-information-about-a-target,they-immediately-know-all-there-is-to-know-about-that-target.-
- 3.- In-forming-their-judgments,-perceivers-only-use-the-intensity-and/or-frequency-of-the-targets’-behavior-and-compare-this-with-the-possible-range-of-intensities-for-that-behavior.-This-means-that-they-ignore-situational-context-as-well-as-the-group-to-which-a-target-belongs.-The-implicit-range-of-behavioral-intensity-with-which-perceivers-compare-what-they-observed-is-the-same-for-all-perceivers.-
- 4.- Perceivers-derive-their-judgments-independently.-There-is-no-communication-between-perceivers-(Kenny,-1994)-about-their-judgments,-or-if-there-is,-it-has-no-effect-on-the-perceivers’-judgments.-
- 5.- Causal-effects-only-work-in-one-direction.-For-example,-judgments-may-reflect-attitudes-and-substance,-but-not-affect-them.-
- 6.- We-do-not-distinguish-between-attitude-components-that-are-under-the-perceiver’s-control-(e.g.,-willful-attempts-to-make-a-target-look-good,-by-rating-them-in-a-certain-way)-and-ones-that-are-not-(e.g.,-actually-believing-the-target-is-good).-
- 7.- We-only-account-for-a-single-attitude-variable,-neglecting-the-possibility-that-the-same-substance-may-be-evaluated-differently-in-regard-to-different-goals.-
- 8.- We-only-consider-linear-effects-of-one-variable-on-another.-
- 9.- We-neglect-the-influence-of-stereotypes-tied-to-“categorical-information”-such-as-gender-or-age-(Kenny,-1994)-and-the-possible-existence-of-illusory-correlations-between-substance-variables.-
- 10.- We-only-consider-target-variance-in-the-substance-variables,-and-ignore-systematic-target-by-situation-interactions,-as-well-as-situation-main-effects-(Chmielewski-&-Watson,-2009;-Leising-&-Schilling,-2025;-Steyer-et-al,-1999;-Wood-et-al,-2023).-



## Supplementary Materials2

### Peer Review Communication2

Download: [https://collabra.scholasticahq.com/article/133683-why-do-judgments-on-different-person-descriptive-attributes-correlate-with-one-another-a-conceptual-analysis-with-relevance-for-most-psychometric-res/attachment/274441.docx?auth\\_token=dIDqUArG791YvP9LWJkC](https://collabra.scholasticahq.com/article/133683-why-do-judgments-on-different-person-descriptive-attributes-correlate-with-one-another-a-conceptual-analysis-with-relevance-for-most-psychometric-res/attachment/274441.docx?auth_token=dIDqUArG791YvP9LWJkC)

---