4-2021

Hidden framings and hidden asymmetries in the measurement of personality: A combined lens-model and frame-of-reference perspective

Julian SCHULZE
Freie Universitat Berlin

Stephen G. WEST
Arizona State University

Jan-Philipp FREUDENSTEIN
Freie Universitat Berlin

Philipp SCHAEPERS
Singapore Management University, pschapers@smu.edu.sg

Patrick MUSSEL
Freie Universitat Berlin

Follow this and additional works at: https://ink.library.smu.edu.sg/lkcsb_research

See next page for additional authors

Part of the Industrial and Organizational Psychology Commons, and the Organizational Behavior and Theory Commons

Citation

Author

Julian SCHULZE, Stephen G. WEST, Jan-Philipp FREUDENSTEIN, Philipp SCHAEPERS, Patrick MUSSEL, Michael EID, and Stefan KRUMM

ORIGINAL MANUSCRIPT

WILEY

Hidden framings and hidden asymmetries in the measurement of personality—A combined lens-model and frame-of-reference perspective

Julian Schulze[1] | Stephen G. West[2] | Jan-Philipp Freudenstein[1] |
Philipp Schäpers[3] | Patrick Mussel[4] | Michael Eid[5] | Stefan Krumm[1]

[1]Division Psychological Assessment and Differential and Personality Psychology, Department of Education and Psychology, Freie Universität Berlin, Berlin, Germany

[2]Department of Psychology, Arizona State University, Tempe, AZ, United States

[3]Lee Kong Chian School of Business, Singapore Management University, Singapore, Singapore

[4]Division for Personality Psychology and Psychological Assessment, Department of Education and Psychology, Freie Universität Berlin, Berlin, Germany

[5]Methods and Evaluation, Department of Education and Psychology, Freie Universität Berlin, Berlin, Germany

Correspondence
Julian Schulze, Division Psychological Assessment and Differential and Personality Psychology, Department of Education and Psychology, Freie Universität Berlin, Habelschwerdter Allee 45, Berlin 14195, Germany.
Email: julian.schulze@fu-berlin.de

Funding information
Alexander von Humboldt-Stiftung

Abstract

Objective: The symmetry principle and the frame-of-reference perspective have each made contributions to improving the measurement of personality. Although each perspective is valuable in its own right, we argue that even greater improvement can be achieved through the combination of both. Therefore, the goal of the current article was to show the value of a combined lens-model and frame-of-reference perspective.

Method: We conducted a literature review to summarize relevant research findings that shed light on the interplay of both perspectives and developed an integrative model.

Results: Based on the literature review and on theoretical grounds, we argue that a basic premise of the frame-of-reference literature—that personality items are open to interpretation and allow individuals to impose their own contextual framings—should be considered from a symmetry perspective. Unintended context-specificity in items may "spread" to personality facets and domains, and thus, impact the symmetry of personality measures with other criteria. As the individuals´ frames-of-reference and (a)symmetric relationships are not always apparent, we term them as "hidden."

Conclusions: The proposed combination of lens-model and frame-of-reference perspectives provides further insights into current issues in personality research and uncovers important avenues for future research.

KEYWORDS
Big Five, frame-of-reference, personality, personality measurement, symmetry principle, validity

1 | INTRODUCTION

The value of personality measures for the prediction of criteria such as work performance, academic success, or domains of life satisfaction, among others, is supported by several meta-analyses (e.g., Judge, Rodell, Klinger, Simon, & Crawford, 2013; Malouff, Thorsteinsson, Schutte, Bhullar, & Rooke, 2010; Poropat, 2009; Shaffer & Postlethwaite,

2012). Notwithstanding, critics of personality measures maintain that their predictive power is still relatively low (e.g., for predicting job performance; Morgeson et al., 2007) and equivocal across studies (as a recent example from the educational context, see Meyer, Fleckenstein, Retelsdorf, & Köller, 2019). From an assessment perspective, a lively debate surrounds the question of whether personality measures meet common model fit criteria in confirmatory factor analyses (Gignac, Bates, & Jang, 2007) and if the measures exhibit measurement invariance, for example, across age groups (e.g., Nye, Allemand, Gosling, Potter, & Roberts, 2016; Olaru, Schroeders, Wilhelm, & Ostendorf, 2019) or different kinds of raters (Mõttus, Allik, & Realo, 2020). Given the ongoing controversy about the predictive utility, quality, and functionality of personality measures, theoretical explanations are needed to move research in these fields forward.

Two broad approaches provide important opportunities for increasing the (construct- and criterion-related) validity of conclusions based on personality tests: The symmetry principle (Wittmann, 1988) as derived from the Brunswik (1956) lens model and frame-of-reference testing (e.g., Lievens, De Corte, & Schollaert, 2008; Schmit, Ryan, Stierwalt, & Powell, 1995). A key idea of the symmetry principle is based on the hierarchical organization of constructs (in personality, the domain, facet, and item levels, Costa & McCrae, 1995; McCrae, 2015). The symmetry principle posits that predictive power might be diminished if a predictor and a criterion are (a) related to each other at different levels of generality and/or (b) the predictor contains not all criterion-relevant components or contains criterion-irrelevant components (Wittmann & Klumb, 2006). The second approach, frame-of-reference, has shown that the predictive power of personality measures is higher when they are contextualized (e.g., through item tags such as "at work"—providing a *frame-of-reference*) to match the contextual domain of the criterion (e.g., work performance; Shaffer & Postlethwaite, 2012). The central idea is that contextualization reduces differences in item construal within and across persons and may increase conceptual overlap of the predictor with the criterion variable (Lievens et al., 2008).

To date, the two perspectives have been presented in largely separate literatures. We propose that integrating these approaches may yield important novel implications for personality psychology. Specifically, the frame-of-reference perspective suggests that individuals may infer contexts for even generic personality questionnaire items (Lievens et al., 2008). The inferred context may be the result of item characteristics, person characteristics, and their interaction (adapted from the situation construal model, Funder, 2016). If a material number of items share such a "hidden framing," context specificity may "spread" to facets or even domains of a personality test and can thus affect its contextual symmetry with criteria, as suggested by the symmetry principle (Wittmann,

1988). These (a)symmetries will often be unnoticed by researchers, and hence, are also referred to as "hidden" in the current article.

The present paper is organized as follows: First, we will present the two perspectives and introduce the concept of hidden framings. We will outline the mechanisms through which hidden framings may introduce context-specificity at the item, facet, and domain-level. Second, we will connect the concept of hidden framings to the symmetry principle. Third, we will outline the consequences of this reasoning for three important areas in personality research. Although we introduce both the symmetry principle and frame-of-reference with a focus on predictor-criterion overlap (predictive power), our integrated perspective will also highlight potential consequences for the (dis)agreement of self- and other reports on personality ratings as well as for the stability and change of personality traits over the life-course. Finally, we will briefly introduce a state-of-the-art statistical approach, the bifactor-(S-1) model (Eid, Geiser, Koch, & Heene, 2017), and qualitatively oriented approaches, which in combination may provide a framework for future empirical work.

2 | THE SYMMETRY PRINCIPLE

The Brunswik (1956) lens model has been applied in studies on interpersonal judgments (Borkenau & Liebler, 1992) and trait continuity (Asendorpf, 1992). Wittmann (1988) presented a *hierarchical* lens model and portrayed different cases of asymmetry. Ajzen (2005, p. 3) nominated the symmetry principle as a "law" in the psychological literature. Successful applications of the principle have been reported across diverse research areas, including personality psychology (Wittmann & Klumb, 2006) and intelligence (Zech, Bühner, Kröner, Heene, & Hilbert, 2017). Its key assumption is that the correlation between two constructs is affected by their symmetry (Wittmann, 1988). Figure 1 provides panels illustrating four types of asymmetry. In each panel the left side shows a hierarchically structured predictor and the right side a hierarchically organized criterion.

Panel (a) depicts a complete lack of symmetry. There is no intersection between the predictor and criterion leading to an expected zero correlation. Wittmann and Klumb (2006) describe an approximation of this case, in which they correlated self-reports of Extraversion with an aggregate of "repeatedly measured multiple-act criteria" (RMAC, p. 200) of the trait of Neuroticism. As expected, the correlation of this RMAC aggregate with the trait inventory rating of a noncorresponding personality dimension was low ($r = -.15$, *ns*). Asymmetry is not necessarily a problem as long as the researcher knows about the asymmetry. Full asymmetry can be seen as an interesting case of discriminant validity (see Wittmann & Klumb, 2006).
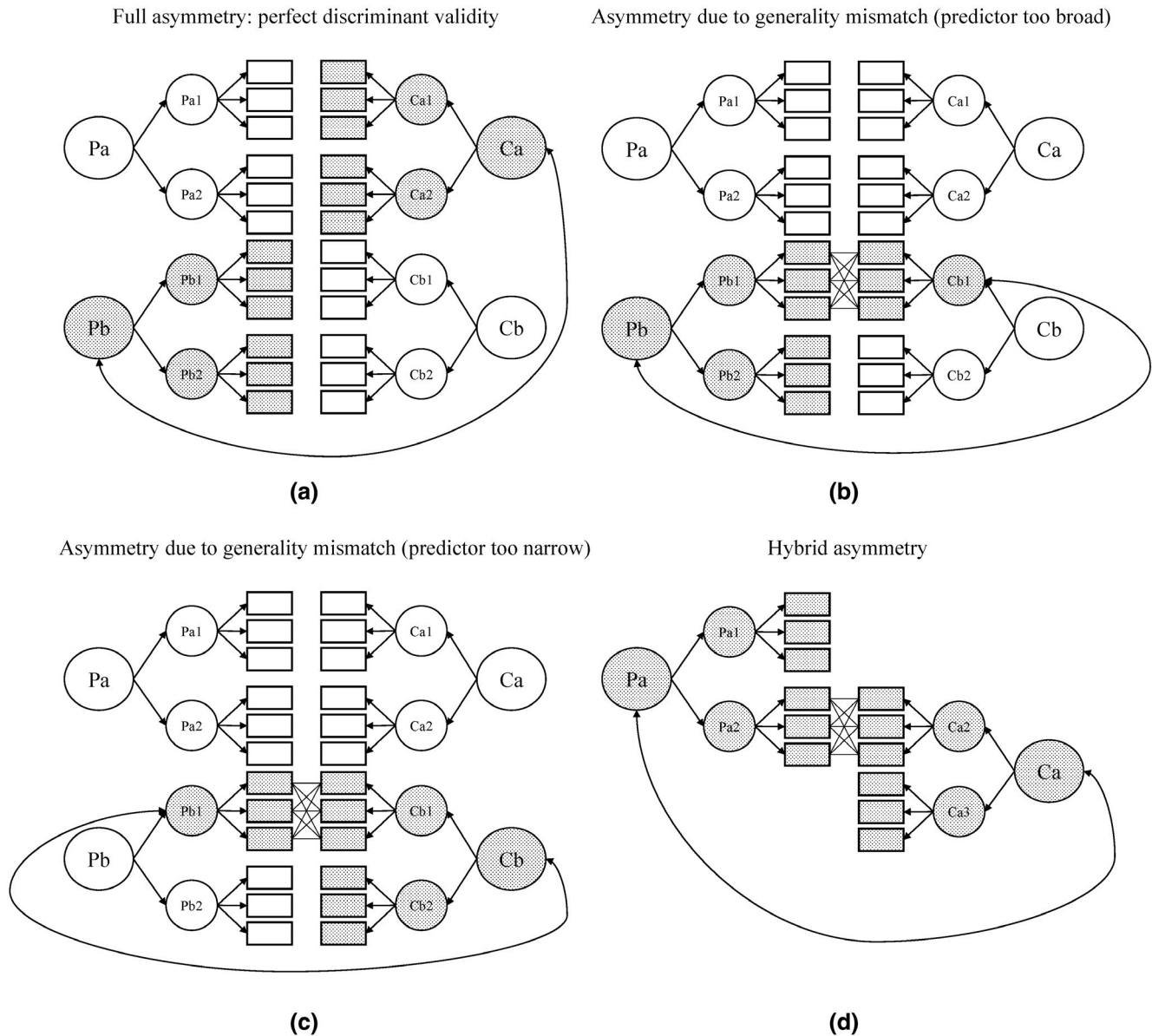
Full asymmetry: perfect discriminant validity

Asymmetry due to generality mismatch (predictor too broad)

(a)

(b)

Asymmetry due to generality mismatch (predictor too narrow)

Hybrid asymmetry

(c)

(d)

FIGURE 1   Four cases of asymmetry as described in Wittmann and Klumb (2006, see also Wittmann, 1988). (a) Full asymmetry (zero correlation between predictor and criterion); (b) Asymmetry due to generality mismatch (predictor too broad); (c) Asymmetry due to generality mismatch (predictor too narrow); (d) Hybrid case of asymmetry (same level of generality, but predictor and criterion only partially overlap). Ca, criterion a; Cb, criterion b; Pa, predictor a; Pb, predictor b, lower order facets are numbered. Please note that the two sides of the hierarchical lens model do not necessarily need to take the form of predictor and criterion but could represent a variety of constructs measured by different methods. We use this terminology as most empirical applications have done. Adapted from Wittmann and Klumb (2006). Copyright 2006 by the American Psychological Association. Adapted with permission

Panels (b) and (c) portray asymmetries due to a mismatch of generality. The predictor is at a level of generality that is too broad or too narrow in scope, because either (1) unwanted variance with respect to the criterion has been introduced or (2) substantial criterion-relevant variance has been excluded. In the cases represented by Panels (b) and (c), the correlation with the criterion will be attenuated (Wittmann & Klumb, 2006). For example, Epstein (1979) suggested that nonaggregate daily reports of emotions and behaviors cannot be expected to correlate highly with trait-based measures of personality because they are too specific. If the goal is to increase the symmetry of such measures, their level of generality must be matched. Epstein (1979) could show that self-reported personality traits showed meaningful relationships with daily records of feelings and behaviors, if such indicators were aggregated over multiple occasions, thereby matching the level of generality more closely.

Finally, Panel (d) portrays a case of partial symmetry between predictor and criterion (hybrid case), albeit at a similar
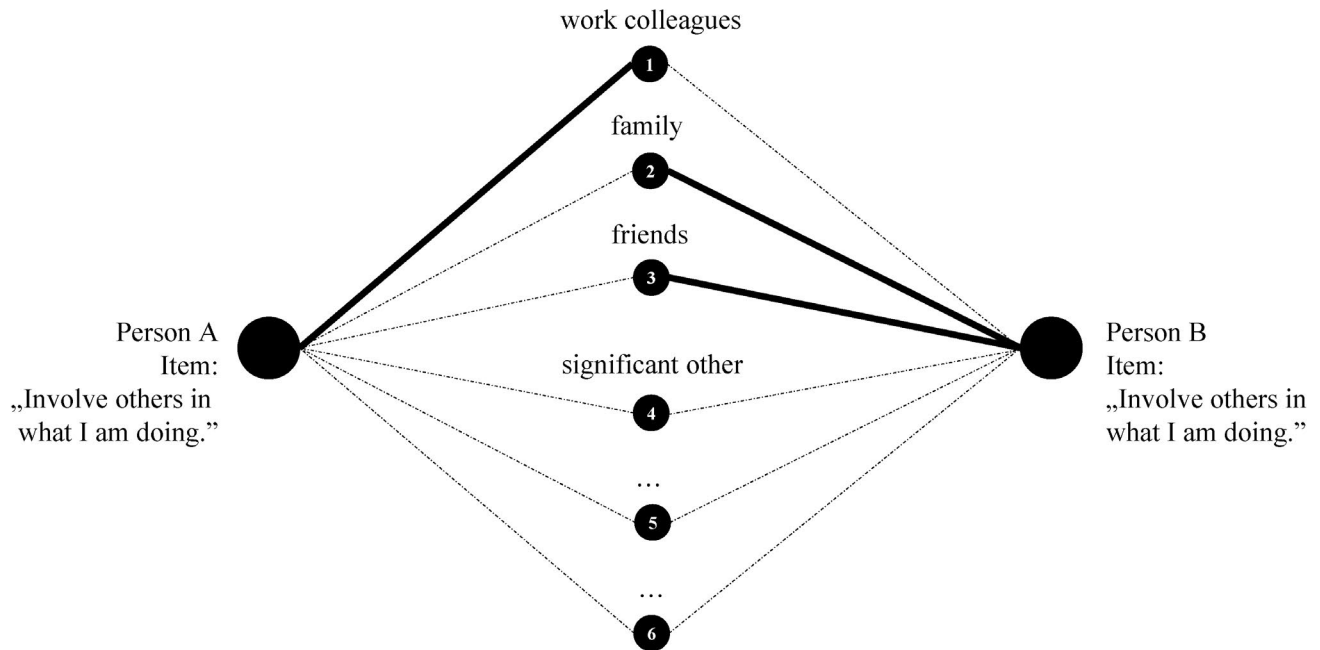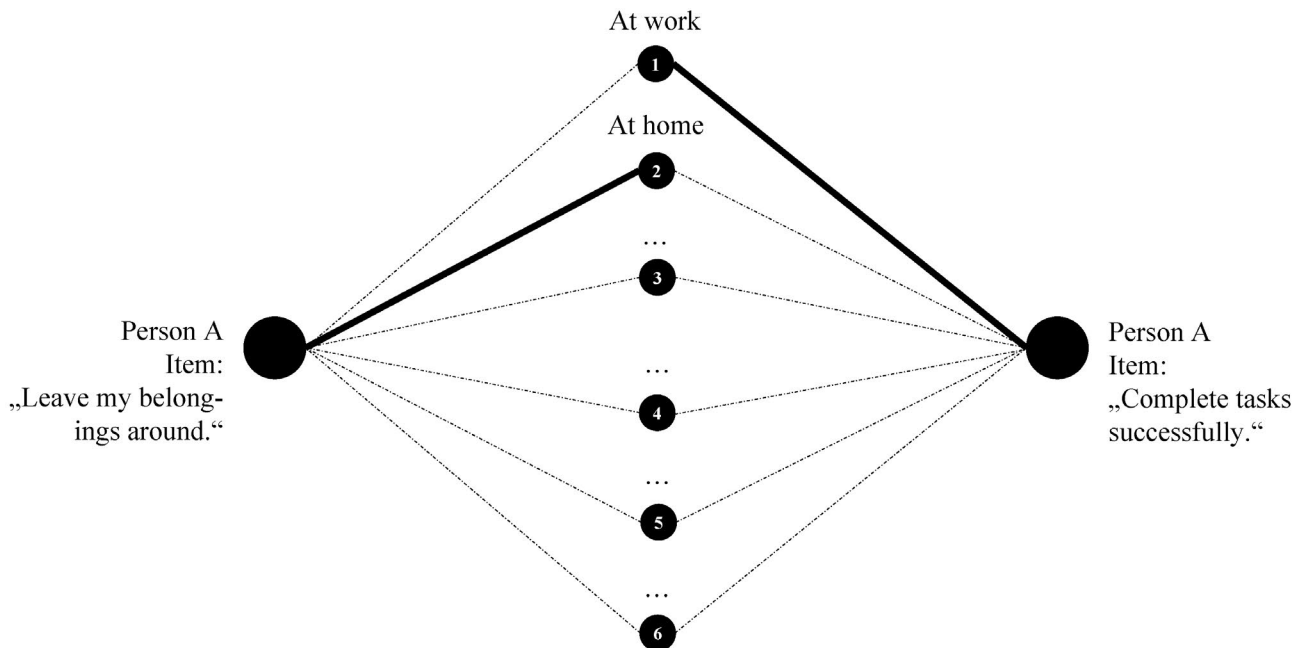
Between-person variability

Frame-of-reference

work colleagues

family

friends

significant other

...

...

Person A
Item:
„Involve others in
what I am doing."

Person B
Item:
„Involve others in
what I am doing."

Within-person inconsistency

Frame-of-reference

At work

At home

...

...

...

...

Person A
Item:
„Leave my belong-
ings around."

Person A
Item:
„Complete tasks
successfully."

FIGURE 2 Frame-of-reference, between-person variability and within-person inconsistency. Panel (a): Illustration of between-person variability; Panel (b): Illustration of within-person inconsistency (see main text for details; Items derived from IPIP: https://ipip.ori.org/Alphabetic alItemList.htm)

level of generality. The predictor and criterion overlap, but their correlation is attenuated due to a mismatch of components on both sides of the lens model (Wittmann & Klumb, 2006). Comprehensive assessments of all relevant components of the construct on both the predictor and the criterion sides are needed to identify reasons for (a)symmetry (see Wittmann, 1988, for examples).

In sum, the literature on the symmetry principle highlights (1) the importance of matching the generality level of predictor and criterion, and (2) the inclusion of all

criterion-relevant components and avoidance of criterion-irrelevant components in aggregate scores. To achieve higher levels of symmetry, time, target, action, and context elements must be matched between predictors and criteria (Wittmann & Klumb, 2006; see also Ajzen, 2005). The frame-of-reference literature, to which we now turn, also implicitly incorporates ideas from the symmetry principle by focusing on context-specificity in item ratings.

3 | FRAME-OF-REFERENCE TESTING

The frame-of-reference testing literature has identified some methods of enhancing the predictive power of personality tests (e.g., Bing, Whanger, Davison, & VanHook, 2004; Lievens et al., 2008; Schmit et al., 1995). Based on the idea of conditional dispositions (Wright & Mischel, 1987), researchers provided personality tests with a *frame-of-reference* (context) by using item tags (e.g., "at work," "at school") or contextualized instructions. These adapted tests were more predictive of criteria than the original inventories when the contextualization matched the domain of the outcome (e.g., "at school" provided a context for item-ratings of the trait of Conscientiousness when the criterion was grade point average; Schmit et al., 1995). Meta-analytic evidence suggests that tagged measures of the Big Five personality traits predicted work performance ratings better than their untagged counterparts (Shaffer & Postlethwaite, 2012). In addition to a contextual match between predictor and criterion, Lievens et al. (2008) suggested that increases in predictive power resulted from the reduction of (1): "between-person variability" and (2) "within-person inconsistency" in item responding (see Figure 2).

3.1 | Between-person variability

When responding to generic personality questionnaires, Holtz, Ployhart, and Dominguez (2005) proposed that "some test-takers may respond in accordance with how they perceive their personality across situations while others may respond specifically to how they view themselves at work, home, or elsewhere" (p. 76). This first source of variability has been termed between-person variability (Lievens et al., 2008; see Figure 2a). Whereas Person A may focus on work colleagues while responding to the item "Involve others in what I am doing" (International Personality Item Pool (IPIP), Goldberg et al., 2006; http://ipip.ori.org/), Person B may focus on friends and family while responding to the identical item. Between-person variability will reduce the predictive power of personality inventory scores if different individuals use frames-of-reference that vary in their relevance to the criterion when responding to items (Lievens et al., 2008). If an individual were to focus on leisure time behaviors that did not match the context of the criterion of work performance, this context-mismatch would be expected to attenuate the predictor-criterion correlation. Matching the frame-of-reference to the context of the criterion enhances the likelihood that the same (relevant) context will be used by all respondents, thereby reducing between-person variability (e.g., Davison & Bing, 2009; Lievens et al., 2008).

3.2 | Within-person inconsistency

Within-person inconsistency occurs if individuals vary their frame-of-reference while responding to several generic personality questionnaire items (see Figure 2b). A respondent may use "… one frame of reference to answer one generic item and another frame of reference to answer another item" (Lievens et al, 2008, p. 270). Although an individual may adopt an "at home" frame-of-reference while responding to the Conscientiousness item "Leave my belongings around," this frame may change to an "at work" frame-of-reference while responding to the item "Complete tasks successfully" (both items from IPIP; http://ipip.ori.org/). Such shifts in frames-of-reference introduce within-person inconsistency in item responding. Contextualizing a personality test is thought to reduce this inconsistency; the same frame-of-reference (e.g., "at work") is now implied by all items, minimizing re-framing by respondents (Lievens et al., 2008).

4 | "HIDDEN FRAMINGS"

The situation construal model (SCM; Funder, 2016) provides a theoretical underpinning for frames-of-reference effects (see Schäpers et al., 2020, for a similar adaptation of situation construal models to Situational Judgment Test items). The SCM posits that both personality and objective situational characteristics influence an individual's subjective construal of a situation, which in turn influences the individual's response to the situation. Adapted to frames-of-reference, the SCM posits that personality item characteristics and a respondent's personal characteristics affect an individual's subjective contextual construal of the item. This construal—along with unmediated person and item effects—influences the respondent's reply to that item. Because researchers may not intend to measure and not even be aware of the respondent's construal (framing), we label this effect *hidden framing*. One consequence of hidden framings is that the context-generality of personality judgments may vary across individuals or across items: The more distinct and abstract frames-of-reference the individual imagines when responding to an item,

TABLE 1 References to the views that (a) personality items are open to (contextual) interpretation, (b) item characteristics (i.e., the wording) may contribute to the emergence of hidden framings, and (c) person characteristics may contribute to the emergence of hidden framings

	Study methodology	Relevant quote/finding
(a) References supporting the view that individuals construe items in different ways		
Eisenberg (1941)	After completion of 25 selected items from a personality inventory, individuals were asked to indicate why they answered to a question (in a yes/no/? format) as they did	"It will be observed from these tables that questions have different meanings to different individuals, though their responses may be the same. In fact, there was no question in which we did not find different interpretations of the item" (p. 26)
Lundmann and Villadsen (2016)	After completion of the BFI-10 (Rammstedt & John, 2007), individuals were asked to indicate why they chose a particular item response	"What is obvious from the qualitative analyses above is the large qualitative variation in respondents' understandings of the items and the ways in which respondents generate a subjective meaningful answer" (p. 177)
McCrae, Costa, and Martin (1998)	After completion of the revised NEO Personality Inventory (Costa & McCrae, 1992) by married couples to investigate agreement between self- and corresponding spouse-ratings, couples were interviewed to explore causes for their disagreement in ratings	"Different interpretation of items was followed in frequency by a diverse set of other reasons, including differences in the specific behaviors, time frames, or roles considered when responding to items…" (p. 303)
McCune (2010)	The author used thinking-aloud interviews to explore if respondents spontaneously mention contexts when responding to a standard IPIP-50 questionnaire	"Every single participant indicated that they were considering context when choosing their answers to the personality inventory. In addition, the nature of the contexts used when responding varied within individuals, indicating that a single individual was using multiple FORs [frames-of-reference]" (p. 65)
		"… participants appeared to be thinking of themselves as they generally behave, as well as how they behave in certain contexts, when responding to items" (p. 77)
Uher (2018)	Raters were asked to give interpretations for a fictitious individual that has a high score on the item "tends to be lazy" (item derived from the BFI-10; Rammstedt & John, 2007, p. 210). Answers were then clustered in a "field of meaning constructed by raters" (Uher, 2018, p. 18; based on Uher and Dharyial, unpublished)	"… single raters did not consider the item's broad field of meaning that it may generally have in their sociolinguistic community. Instead, when judging the target person, different raters thought of very different behaviors and contexts…" (p. 18)
(b) References supporting the view that the item wording may contribute to hidden framings		
De Raad (1993)	Theoretical article	"Trait descriptive expressions… are not to be considered as concepts liberated from context. Quite the contrary, the majority of trait words, used in personality taxonomies, not only capture important personality characteristics, but also represent contextual information" (p. 271)
Jackson, Hill, and Roberts (2012)	Theoretical article	"While not all personality trait items include contextual cues, many do and almost all major questionnaires employ such items (e.g., Revised NEO Personality Inventory [NEO-PI-R], …, and International Personality Item Pool [IPIP-Big Five])." (p. 747)
McCune (2010)	After completion of an IPIP-50 questionnaire, respondents indicated for each item if they were thinking of predefined frames-of-reference (e.g., at work, at university, with friends, in general; multiple categories could be chosen)	"… it is interesting to note that the 'In General' response option was endorsed by more than half of the participants for only 22 of the 50 items." (p. 100)
		Items such as "Am always prepared" or "Pay attention to details" triggered predominantly frames-of-reference to the "at school" and "at work" context besides the "in general" category

(Continues)

TABLE 1 (Continued)

	Study methodology	Relevant quote/finding
Saucier and Conley (2015)	Theoretical article	"Many commonly used Big Five measures feature items that reference situational contingencies, thus perhaps confounding personality and situation." (p. 411)
Werner and Pervin (1986)	The authors examined the content of items from several personality inventories using three judges. Among several aspects, the authors studied the inclusion of situational information	"On the average, over half (55.7%) of the items on the inventories included reference to situation. The percentage of items rated as referring to situation ranged from a low of 40% … to a high of 72.7% …" (p. 624/625)
(c) References supporting the view that person characteristics may contribute to hidden framings		
Heller, Perunovic, and Reichman (2009)	Theoretical article	The authors present a "bottom-up" theoretical model of personality change. It is argued that the life experiences that individuals make in different social roles (e.g., as a spouse, employee, student) may (through several mediators) shape broad personality traits
McCune (2010)	After completion of an IPIP-50 questionnaire, respondents indicated for each item if they were thinking of predefined frames-of-reference (e.g., at work, at university, with friends, in general; multiple categories could be chosen). Demographic variables were correlated with context categories	The dissertation showed initial evidence that person characteristics were correlated with context endorsement rates. For example, older individuals endorsed less often the "with friends" context-category than younger students and more conscientious individuals endorsed more often the "at work" context-category
Wrzus, Wagner, and Riediger (2016)	The authors used an experience-sampling methodology to get an impression of the situations that individuals are experiencing and how these situations differ over age. Subsequently, personality-situation transactions were investigated and if they are generalizable over age	The study showed that being with different interaction partners (such as family, work colleagues) and activities pursued (such as work or school activities, doing chores) varied over age

the more that response will have context-generality and lack context-specificity.

Table 1 lists references supporting the view that individuals construe personality items differently and that item and person characteristics may relate to these construals. Many personality items feature contextual cues in their wording, which may contribute to the emergence of hidden framings. For example, frequently used words in standard Conscientiousness items such as "job" or "work" may prime work contexts (see also Shaffer & Postlethwaite, 2012). Aspects of the trait itself (as formulated in the item) may also be a contributor to context construal. Specifically, there might be a limited universe of contexts, in which the trait is of relevance (see also the trait activation theory by Tett & Guterman, 2000). "Working hard" may be less relevant in the context of leisure activities compared to the workplace, limiting the potential universe of contexts available for the individual's construal (for a similar argument, see McCune, 2010). Finally, person characteristics related to the individual's life experiences may also influence hidden framings. For example, unemployed individuals may construe standard Conscientiousness items with work-related wording (e.g., "I see myself as someone who does a thorough

job"; BFI-10, Rammstedt & John, 2007, p. 210) differently than currently working employees.

Although the literature suggests that imposing a frame-of-reference makes the context of items in personality tests less ambiguous, thus, decreasing the variability of the individuals' construals (Lievens et al., 2008), some variability may remain. This may be because individuals still focus on different information in the specified context (see also Fisher, Cunningham, Kerr, & Allscheid, 2017). Individuals who work from home may still focus on work-related tasks when answering "at home" contextualized Conscientiousness items, whereas others may focus on domestic chores. Consequently, between-person variability and within-person inconsistency due to hidden framings are expected to be reduced, but not completely eliminated when items are presented with a context tag.

4.1 | Consequences of hidden framings for the context-specificity of personality traits

Item-level hidden framings potentially introduce context-specificity at the facet level when items that share a

framing are aggregated to a facet score. If individuals adopt an "at work" context for items that constitute the *achievement striving* facet (see Costa, McCrae, & Dye, 1991 for facet definitions), this "work" context will become part of the achievement striving aggregate score. Figure 3 illustrates this reasoning using a hierarchical personality model. Hidden framings are denoted by lines connecting each particular personality item with "framing ovals" A, B, and C. Each oval represents a context, for example "at home" or "at work," or other imaginable context framings. Although researchers would like to assume that these framings are randomly distributed across the personality inventory (illustrated in facet 2), they will commonly be somewhat systematically linked to underlying facets of the personality domain (illustrated in facet 1). Roberts, Lejuez, Krueger, Richards, and Hill (2014) reasoned that "… orderliness is typically manifest in homes and workplaces, not in public spaces or social interactions.
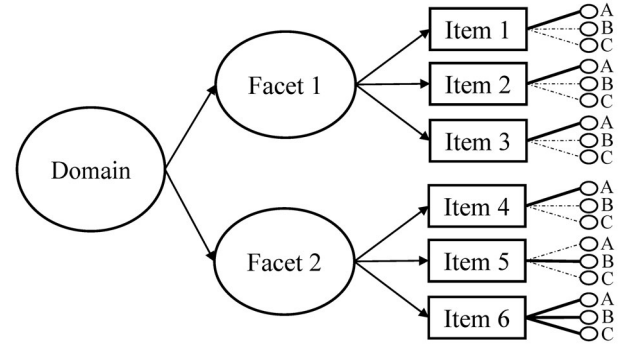

FIGURE 3 Integration of hidden framings into a hierarchical model of personality. Different hidden framings may exist at the item level (ovals A, B, and C; bold lines indicate dominant item-level hidden framing). However, within facets, items may share a hidden framing (Facet 1) or less so (Facet 2). Aggregating over items that share a hidden framing may introduce context-specificity into facet and domain scores

TABLE 2 Example facets illustrating the potential "spread" of context-specificity from the item- to the facet-level

Example facets and items	Potential contexts in facet
IPIP–NEO–120 facet items (see Johnson, 2014, p. 82–83)	
Achievement-Striving (Conscientiousness)	
Do more than what's expected of me	e.g., hidden framings to work and university contexts
Work hard	
Put little time and effort into my work	
Do just enough work to get by	
Orderliness (Conscientiousness)	
Like to tidy up	e.g., hidden framings to home and work contexts
Often forget to put things back in their proper place	
Leave a mess in my room	
Leave my belongings around	
HEXACO-60 facet items (see Ashton, & Lee, 2009, p. 345)	
Social Boldness (Extraversion)	
When I'm in a group of people, I'm often the one who speaks on behalf of the group	e.g., hidden framings to interactions with strangers or group settings
In social situations, I'm usually the one who makes the first move	
I rarely express my opinions in group meetings	
Organization (Conscientiousness)	
I plan ahead and organize things, to avoid scrambling at the last minute	e.g., hidden framings to work and university contexts or leisure time activities
When working, I sometimes have difficulties due to being disorganized	
Aesthetic appreciation (Openness to Experience)	
I would be quite bored by a visit to an art gallery	Hidden framings to leisure time activities
If I had the opportunity, I would like to attend a classical music concert	

Likewise, industriousness is manifest in achievement settings, which in modern society means school and work" (p. 1318). The context-specificity of a facet will be lower to the degree that (1) its items do no not prime similar hidden framings or (2) individuals utilize a cross-contextual average of the behaviors in question (facet 2, item 6).

We have reviewed several personality inventories and have identified contexts that may systematically affect the facet scores. Table 2 (top panel) presents eight items from the IPIP-NEO-120 (Johnson, 2014) organized into two Conscientiousness facets (orderliness and achievement striving). The bottom panel presents seven items from the HEXACO-60 questionnaire (Ashton, & Lee, 2009), organized into three facets. These examples illustrate, across two different personality inventories, that context-specificity may exist at the facet-level if items share a common hidden framing.

Hidden framings may even spread to the domain level. At the level of the Big Five domains, the domain score would represent in part context-specific variance because the effect of hidden framings would not be fully eliminated through aggregation of item or facet level scores. In their meta-analysis, Shaffer and Postlethwaite (2012) showed that the magnitude of the effect of contextualization for the prediction of work performance varied across personality traits. The smallest increase in predictive power was obtained for the Conscientiousness factor. One reviewer of the study suggested that "…the nature of Conscientiousness causes noncontextualized measures of Conscientiousness to be contextualized to some extent. This occurs because noncontextualized measures of Conscientiousness tend to include at least some items that could be construed as being specifically linked to work behaviors" (p. 465). Shaffer and Postlethwaite (2012, p. 465) also stated "… that when completing items that referred to "jobs," "work," or "tasks," respondents [mostly workers] were more inclined to describe their workplace behavior for these items even though items were not explicitly contextualized for selection purposes." This reasoning maps well onto our concept of hidden framings.

To probe this reasoning, we did a search of the IPIP-NEO-120 (Johnson, 2014), the HEXACO-60 (Ashton, & Lee, 2009), and BFI-10 (Rammstedt & John, 2007) Conscientiousness items. We searched for three keywords that may prime the work context: job, work(ing), and task(s), permitting slight variations. From the 24 Conscientiousness items in the IPIP-NEO-120, six (25%) featured one of the keywords. From the ten HEXACO-60 Conscientiousness items, four items (40%) featured the word "work(ing)." Finally, from the two BFI-10 Conscientiousness items, one (50%) featured the word "job." This word count suggests that many items in inventories measuring Conscientiousness could be construed as relating to job-related behaviors (or to university work-related behaviors in a student sample).

The percentage of such "trigger words" varies considerably from inventory to inventory--trait aggregates may differ in their contextual specificity across inventories and samples. Similarly, a considerable number of Openness items in the NEO-PI-3 (McCrae et al., 2005) feature content that is situated in leisure time contexts (e.g., pursuing new hobbies). Saucier and Conley (2015) noted that many Extraversion items in the IPIP-50 (Goldberg et al., 2006; see https://ipip.ori.org/New_IPIP-50-item-scale.htm) feature party situations or could be construed of as being related to interactions with unfamiliar people.

5 | FROM HIDDEN FRAMINGS TO HIDDEN (A)SYMMETRIES

The implications of hidden framings become evident when combined with the principles of symmetry (Wittmann, 1988). Figure 4 portrays the linkage of the two perspectives. The left side of the figure represents a hierarchically structured personality construct and the right side a hierarchically organized criterion. Note that "criterion" serves an umbrella term for constructs measured using different methods. In line with the application of lens model ideas (Brunswik, 1956), the criterion side of Figure 4 may consist of, for example, other-reports (Borkenau & Liebler, 1992) or another measurement time point (Asendorpf, 1992). Ovals in the middle of the figure represent different hidden framings that people adopt when responding to items. The four cases of asymmetry are described below.

The first case of asymmetry (see Figure 4, Facet 1) portrays a complete lack of context-symmetry. All individuals in a sample adopt a hidden framing for all items of a predictor facet that does not match to the context of the criterion facet as indicated by nonconverging bold lines connecting the context ovals. In contrast to Wittmann's (1988) classical full asymmetry, we should not necessarily expect a zero correlation as contextualized personality items tend to be correlated across contexts: Individuals often show cross-contextual consistency in their behaviors. For example, Robinson (2009) showed that Extraversion scores that were contextualized to interactions with friends, with parents, and with work colleagues correlated in the range of .18 to .41.

Figure 4 depicts predictors that are contextually too broad (facet 2) or too narrow (facet 3). If the criterion is obtained from a specific context, but the indicators of the predictor are subject to a variety of hidden framings, this mismatch in *context generality* may attenuate the predictor-criterion correlation. Irrelevant variance components with regard to context may be introduced into the predictor facet scores. Conversely, if a predictor facet has a narrow context, but the criterion has a broader context, relevant
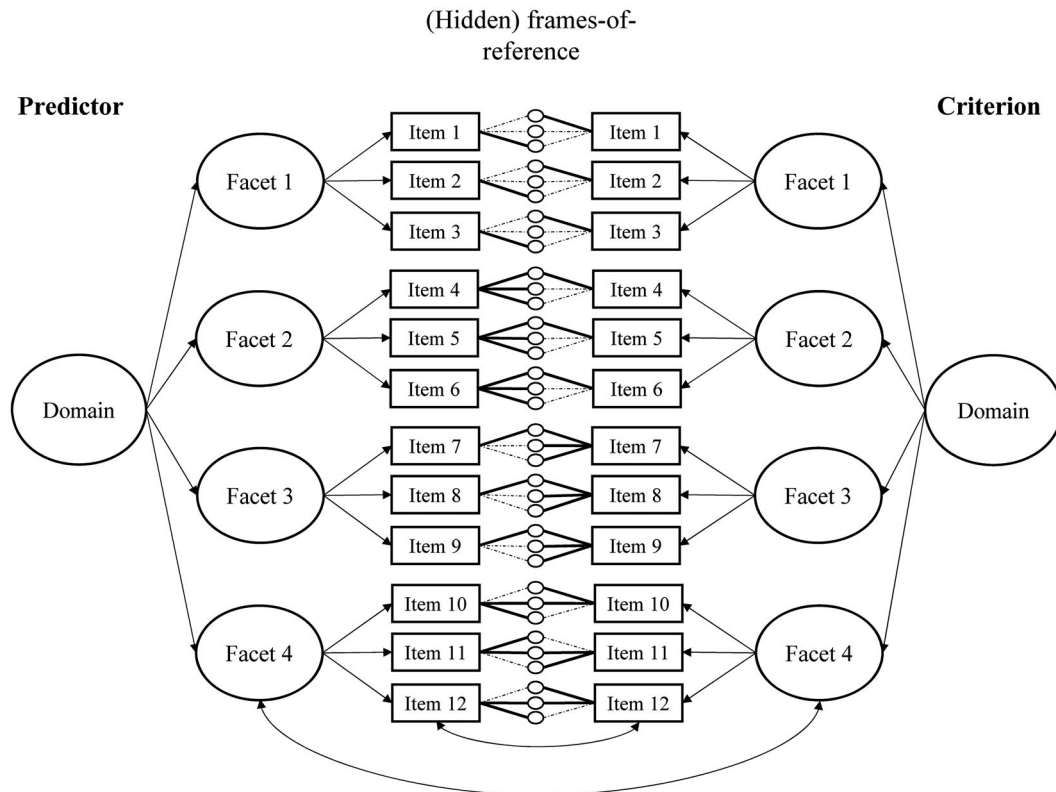
(Hidden) frames-of-
reference

Predictor **Criterion**

FIGURE 4 Integration of the hierarchical model of personality, hidden framings, and the symmetry principle. Different hidden framings may exist at the item level on both sides of the hierarchical lens model (centrally arranged ovals; bold lines indicate hidden framings). Predictor-criterion relationships can be examined at different levels of the hierarchy (examples depicted by bidirectional arrows). (A)symmetries may be present at multiple levels of the hierarchies

contextual variance components will be missing in the predictor facet attenuating the predictor-criterion correlation as well.

Figure 4, facet 4 depicts a hybrid case in which both the predictor and the criterion have similar context generality, but the constructs on both sides of the hierarchical lens model include irrelevant as well as relevant variance components with regard to context. To identify reasons for contextual (a)symmetry, a researcher would ideally administer differentially contextualized measures along with the generic inventories (see Murtha, Kanfer, & Ackerman, 1996 for an empirical example). In the following paragraphs, we will highlight potential consequences of hidden framings and hidden (a)symmetries for three mainstream personality research areas.

5.1 | Predictive power of personality tests

Aggregate scores of personality items (facets; domains) have been frequently related to criteria such as work performance (Judge et al., 2013), academic achievement (Poropat, 2009), or desired leisure time experiences (Barnett, 2013). Recently, item-based models have also been used to predict criteria (Seeboth & Mõttus, 2018). The measured criterion variables

have often had a focus on a specific context (e.g., performance in the *work* context; Judge et al., 2013). In Figure 4, the lines connecting context ovals with the items of such a criterion would point in the same direction (i.e., addressing similar contexts). Based on our reasoning, differential effects of personality might be in part the result of differences in hidden contextual symmetry between the predictors and the criteria. Judge et al. (2013) conducted a meta-analysis of the relationship between personality traits with work performance criteria. The predictive power varied considerably from trait to trait. Personality facets that showed higher predictive power (e.g., achievement striving, self-discipline as facets from Conscientiousness) feature items with particular relevance for work performance. At the same time, some of the measured facets might have been composed of items that primed a hidden work framing, thus, additionally boosting predictive power. Barnett (2013) predicted participant's desired leisure time experiences from facets of the NEO-PI-R (Costa & McCrae, 1992). Factors such as "desiring new experiences in free time" (Barnett, 2013, p. 167) were strongly predicted by Openness facets (among other facets). Some of them include behaviors that are primarily performed in the leisure time context (e.g., watching ballet, developing hobbies). These examples illustrate that predictive power may also be influenced by hidden contextual (a)symmetry with

the criterion. Our approach suggests that the degree of contextual symmetry will depend on the items of the chosen personality inventory, the sample of respondents, the form of aggregation, and the criterion of interest.

5.2 | Convergence of self- and other-reports

Identifying causes for (dis)agreement between self- and other-ratings of personality has been a fundamental challenge for researchers (e.g., Funder & West, 1993). Hidden contextual (a)symmetry between self- and other-ratings may be one cause for (dis)agreement. Connelly and Ones (2010) hypothesized that informants such as classmates or work colleagues may have context-specific knowledge about the target person's personality, whereas self-reports would be made up of knowledge from a more diverse set of contexts. Studies by Kim, Dar-Nimrod, and MacCann (2018), Kurtz and Palfrey (2016) as well as Small and Diefendorff (2006) collected both generic and contextualized personality measurements for targets and generic measurements for informants. These studies did not show generally higher agreement between contextualized self- and the generic other-reports that came from specific contexts on the domain level. De Raad, Sullot, and Barelds (2008) chose another design. These authors collected data from students and well-acquainted informants (mostly friends and parents) and gave them trait adjectives with and without situational specification (different situational information rather than a uniform contextualization). Results showed that the change in agreement varied from domain to domain and from item to item. So, a much more differentiated view on contextual knowledge might be needed that focuses on different levels of the personality hierarchy. The possibility of frequently observing behaviors of the target in trait-relevant situations in one specific context may contribute to hidden framings in items and facets of context-specific informants (for the importance of these components for accurate judgments, see Funder, 2012; Hirschmüller, Egloff, Schmukle, Nestler, & Back, 2015). The level of acquaintance may have an influence. Informants who are well-acquainted with the target may integrate more contexts into their judgment (see also Kurtz & Palfrey, 2016). Finally, simply adding a *context tag* to personality items (e.g., I am lazy *at work*) still allows for differences in interpretation (e.g., offload work on other people, avoidance of work, low quality of work, see Uher, 2018). This variability in interpretation may attenuate self-other agreement (see Hayes & Dunning, 1997 for preliminary evidence; but also see De Raad et al., 2008 who argued that specifications can be at times too specific, lowering agreement). Future research may benefit from focusing on (1) contextualizing both the self- *and* other reports (e.g., as in Bongard, Martin, Seip, & al'Absi, 2011) and (b) considering narrower levels of the personality hierarchy (facets, items).

5.3 | Personality stability and change

Figure 4 also helps visualize the potential impact of hidden framings in longitudinal studies. The criterion side of Figure 4 now represents another measurement time point for the same individuals. Individuals may adopt the same or a different framing at the two time points, illustrated by (non) converging lines connecting context ovals at two time points. There is controversy over whether personality inventories can achieve measurement invariance over age (e.g., Nye et al., 2016; Olaru et al., 2019). Different item interpretation and life experiences have been discussed as potential reasons for age-related differential item functioning (Nye et al.). If life experiences affect hidden framings of personality items, important life transitions (e.g., from work to retirement) may be a useful starting point to further explore this issue. Researchers could collect personality data as well as information about critical life transitions to contrast item construal before and after the event (e.g., first job, parenthood, divorce, retirement; see Bleidorn, Hopwood, & Lucas, 2018 for a review of personality development and life events). Bleidorn et al. suggested that a late life decrease in Conscientiousness could be "… related to the shift out of work-related roles" (p. 91). Mean-level changes and rank-order inconsistency may be the result of true personality change, but the role of hidden framings related to the individual's life transitions in these changes needs systematic investigation.

6 | METHODOLOGICAL CONSIDERATIONS

In this section, we offer quantitative and qualitative approaches that appear to be promising vehicles for the exploration of hidden framings in personality inventories.

6.1 | Quantitative methods

Brunner, Nagy, and Wilhelm (2012) reviewed a variety of confirmatory factor models that address different questions about hierarchical measurement structures. The models included single factor, correlated factor, bifactor, and higher order factor models, each of which hypothesizes a different structure. Two of these initially appealing models, the classic bifactor and higher order G-factor models, have been shown to be problematic under certain conditions. Specifically, bifactor and higher order G-factor models assume that the facets comprising a trait are interchangeable (randomly sampled from a universe of identical facets; Eid et al., 2017). Personality facets, however, are likely structurally different rather than interchangeable. Applied to structurally different facets, these models may yield anomalous solutions, in which

one or more of the specific factors collapse or have irregular loading patterns. Eid et al. (2017) proposed the bifactor-(S-1) model which does not make this assumption and yields a clear solution. In this model, one facet is chosen as a reference facet and is contrasted against the other measured facets of interest. In measuring achievement, the generic, unaltered achievement striving facet could be chosen as the reference and other measured facets (e.g., the same facet contextualized to work and home contexts) could be contrasted with it. Figure 5 shows an illustration of the bifactor-(S-1) model applied to this example. The bifactor-(S-1) model provides estimates of consistency and specificity. The consistency index provides information about the proportion of variance in the contextualized items and facets that is determined by the reference facet (e.g., here, generic achievement striving). The specificity index estimates the proportion of variance that is unique to each context (home, work) (Eid et al., 2017). The two indices provide information about potential context-specificity and hidden framings in items and facets.

More generally, Eid, Geiser, and Koch's (2016) design-oriented approach to specifying models permits researchers to address a variety of questions related to hidden framings. Self-other agreement regarding contextualized questionnaires can be addressed using the correlated trait-correlated method minus one (CTC(M − 1)) model (see Eid, Lischetzke, Nussbeck, & Trierweiler, 2003). In a CTC(M − 1) model, self-reported generic or contextualized facets may serve as the reference that is contrasted against reports from context-specific informants (e.g., work colleagues, cohabitants). In this design, consistency and specificity coefficients could inform about hidden framings of informants. For instance, if targets' work-contextualized self-discipline would show higher consistency with generic self-discipline reported by work colleagues compared to cohabitants, this would provide support for a more pronounced work-specific hidden framing in work-colleague reports. Other designs are possible. Comparisons of consistency and specificity coefficients across different groups (e.g., age groups) can be performed and the reference and context-specific facets can serve as predictors of criteria.

6.2 | Qualitative methods and method combinations

Jobe (2003) presented an overview of methods used in cognitive psychology to test and improve self-report questionnaires. Cognitive interviews, including the think-aloud technique and probing questions, would be particularly useful in exploring the role of context in personality judgments. In think-aloud studies, respondents are asked to think-aloud while responding to a questionnaire, providing a better understanding of the thought process of respondents (Jobe,
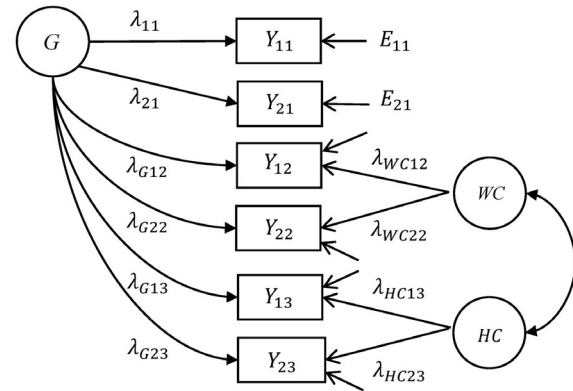
FIGURE 5 Example bifactor-(S-1) model. The "G"-factor in this model could represent the generic, unaltered achievement striving facet of the Conscientiousness domain. This reference facet can then be contrasted against contextualized versions of the same facet, e.g., a work- and a home-contextualized achievement striving facet (i.e., WC and HC)

2003). Probe questions can be used to collect topic-specific information. Following McCune (2010), researchers could use the think-aloud technique to examine if respondents spontaneously mention contexts while responding to personality items. Follow-up probe questions with a specific focus on context could supplement this approach: To infer the context-ladenness of self-reports, researchers could ask respondents if they had considered contexts in their judgments or ask for behavioral exemplars. Variants of cognitive interviewing exist in the literature. For instance, Lundmann and Villadsen (2016) provided online open text fields to respondents and asked for elaboration of their item responses. McCrae, Stone, Fagan, and Costa (1998) had interviewers explore reasons for disagreement on personality ratings of couples. These methods could be tailored to focus on the context-specificity of personality judgments. Finally, researchers could also present individuals prespecified context-categories after completing a personality inventory (Fisher et al., 2017; McCune, 2010).

Qualitative and quantitative approaches can be combined in meaningful ways. Meitinger (2017) conducted measurement invariance tests of national identity items across several countries. Online probing helped to understand the reasons for a failed test of scalar invariance. Similarly, the results of our proposed latent variable models can be supplemented with qualitative data. Qualitative responses regarding context construal from targets and informants may help to understand differences in consistency coefficients (= inter-rater agreement) in CTC(M − 1) models. Researchers could examine whether higher (lower) agreement is associated with (dis)similar context construal in self- and other reports. Qualitative data could also supplement information on predictive power of Big Five traits in extended bifactor-(S-1) models. Researchers could examine if higher predictive power of personality traits can

be observed for those individuals who qualitatively reported that they framed their personality judgments to the context of the criterion.

7 | DISCUSSION

The major contribution of the current article was to bring together two theoretical approaches, the symmetry principle (Wittmann, 1988) and the frame-of-reference effect (e.g., Lievens et al., 2008), and to discuss their joint implications for personality research. The symmetry principle highlights the need to match the generality of measures and elements such as time, action, target, and context to maximize the correlation between assessments (Ajzen, 2005; Wittmann & Klumb, 2006). The frame-of-reference literature empirically underlines the importance of one of these elements, *context-specificity* (Shaffer & Postlethwaite, 2012). Building on the basic premise of frame-of-reference researchers—that individuals may consider context while responding to generic personality items (Lievens et al., 2008)—we introduced the concept of "hidden" framings and provided a foundation for their emergence by drawing parallels to the SCM (Funder, 2016). We argued that these framings may introduce context-specificity at different levels of the personality hierarchy, which may influence the contextual symmetry of personality traits to other constructs. We highlighted potential consequences of these hidden (a)symmetries for three research areas in personality psychology and outlined methodological approaches for testing our hypotheses.

7.1 | Theoretical implications and future research opportunities

Several implications and future research opportunities arise from our review. By integrating these two previously separate streams of research, we were able to shed light on the intersection of trait and context and the need to further explore hidden framings in personality inventories. Several researchers have remarked that our knowledge about context in personality measurements is still insufficient (De Raad, 2005; De Raad et al., 2008; Roberts, 2007). De Raad (2005) noted that "the majority of instruments are rather systematic where traits or behaviors are concerned, but ad hoc and unsystematic where situations are concerned" (p. 201). Likewise, Roberts (2007) stated that "despite the intrinsic contextualized nature of personality traits, the measurement of situations is typically tossed unceremoniously into the black box of the personality trait inventory" (p. 1077). Our review of the literature supports these views and additionally suggests explorations of this black box not only from an item perspective, but also from a person perspective.

Grounded in the frame-of-reference literature (Lievens et al., 2008) and drawing on the SCM (Funder, 2016), we hypothesized that item *and* person characteristics as well as their interaction may influence hidden framings. Although hidden framings and the concepts of between-person variability and within-person inconsistency offer considerable potential to strengthen our understanding of context in personality measurement, Lievens et al. (2008) noted that "… we have no information about the prevalence of these inconsistencies" (p. 278). Indeed, researchers differ materially in their assumptions regarding how individuals arrive at a personality item rating and how much context-specificity remains in standard trait judgments. Lievens et al. (2008) stressed the potentially contextualized nature of personality judgments and suggested that individuals may respond to personality items by drawing on cognitive schemata and autobiographical memories. Other researchers have emphasized the decontextualized nature of generic personality traits (e.g., Bandura, 2015). Given enough experience with a particular trait (Klein, Babey, & Sherman, 1997), Klein, Cosmides, Tooby, and Chance (2002) assume that people will have formed abstract summary representations of their personality traits (so called trait summaries) that are retrieved from *semantic* memory when judging a trait. Only trait-*in*consistent episodic memories are then retrieved along with a trait summary (Klein et al., 2002), supporting a more decontextualized view. Klein et al.'s findings, however, are primarily based on studies using trait *adjectives* that include minimal contextual information compared to the behavioral indicators of inventories such as the IPIP. Clearly, more research is needed to clarify these positions and to identify the item and person characteristics that may lead to a larger or smaller contribution of context-specificity in personality measurement. Cognitive processing models may help in identifying features that may contribute to hidden framings. Based on a literature review, Angleitner, John, and Löhr (1986, p. 84) proposed a three-phase processing model, which consists of the stages "encoding" (reading the item and deducing its meaning), "item-self comparison" (deciding to agree or disagree by comparing deduced meaning with self-information), and "utility control" (final utility check before marking the item). This model may aid in the understanding of the potential effect of item and person characteristics on each phase. Researchers might ask how the ambiguity of an item (see Angleitner et al., 1986) impacts context construal at the encoding stage. They might also explore how many contexts individuals integrate during the item-self comparison stage.

Identifying factors that potentially lead to hidden framings will be key to understanding the context-ladenness of personality ratings. Our primary focus in the current article was on the wording and understanding of personality items potentially introducing contextual framings. However, additional factors may also be worthy of consideration. Extending

our current framework, researchers might further investigate whether factors like test situation (e.g., employee selection situation; see Fisher et al., 2017), survey instructions, item composition, or keying of items affect hidden framings (see also McCune, 2010). For example, negatively keyed items ("Waste my time") may stimulate different hidden framings compared to positively keyed items ("Start tasks right away"; both items from IPIP; http://ipip.ori.org/). Taking a more person-oriented view, the literature on situation prototypes (Cantor, Mischel, & Schwartz, 1982) suggests that contexts may be differentially weighted while responding to personality items. Rather than equally weighting contexts in judgments, individuals may have formed prototypical word-context linkages (e.g., "task" and job-related context) that may lead to stronger weighting of one context over another.

Our approach also provides research opportunities for further exploring the interplay of states and traits. Fleeson and Gallagher (2009) showed that the averages of individual personality state distributions correlate with the means of corresponding classical trait self-reports. Recently, researchers have shown the value of measuring situational information along with personality states and traits in experience sampling studies (e.g., Sherman, Rauthmann, Brown, Serfass, & Jones, 2015) and in considering context-specific aggregates of states (Geukes, Nestler, Hutteman, Küfner, & Back, 2017). Future studies could collect state measures along with context information (e.g., location, interaction partner) as well as generic and contextualized personality trait questionnaires. For instance, the mean of achievement striving states that have been collected in a nonwork context (e.g., at home) may show a lower correlation with a generic achievement striving trait compared to the mean of achievement striving states that have been collected in a work context.

7.2 | Measurement implications

Our framework provides some guidance on the construction of personality tests. We propose (and thank an anonymous reviewer for this idea) that the control of framings in assessment follows a continuum ranging from no explicit control to maximum control of context (similar to Lievens & Sackett, 2017, who described levels of contextualization for different personnel selection methods). Different realizations of this continuum may be proposed based on different theoretical assumptions. The "no control" end of the continuum may be realized by giving no explicit context-information in the items. An example is the traditional list of trait adjectives. A researcher following this realization believes that the free choice of framing is essential for personality assessment and trusts in the functional equivalence of framings as originally proposed by Allport (1961).

Contextual cues may be added to limit uncontrolled variation in framing. A researcher may casually rather than systematically include context information in the items as is done in many current personality inventories (see De Raad, 2005; Roberts, 2007). We take the position that trait inventories should build in contextual cues systematically (De Raad et al., 2008; Ten Berge & De Raad, 1999; Van Heck, Perugini, Caprara, & Fröger, 1994). Facet-theory could serve as a guide in the development of such inventories (see Guttman & Greenbaum, 1998). For example, researchers could develop a mapping sentence consisting of a trait facet and a context facet that may help to combine these facets in a more structured way (for an empirical application of a mapping sentence in the context of sensitivity to reward and punishment, see Beauducel, Kersting, & Liepmann, 2005). The inclusion of context could be realized with different methods that vary in the degree of contextual specification. For instance, short context tags ("job," "leisure," "friends," etc.) could be systematically added to personality items. An even higher degree of contextualization may be realized by systematically constructing an inventory of fully contextualized items (see Robie, Risavy, Holtrop, & Born, 2017 for item examples) or situational vignettes (see Ziegler, Horstmann, & Ziegler, 2019 for an empirical example). Realizations that include contexts in a representative way are consistent with the perspective of Brunswik's (1956) representative design and emphasize the importance of measuring the person in context (Ten Berge & De Raad, 1999).

Following Anastasi (1983) and Ten Berge and De Raad (1999), we believe that a personality construct is best defined by measuring the person-in-context. As a consequence, we take the position that the intended contexts in measurement instruments should be made more explicit and obvious for respondents to better control for hidden framings. If the goal is to measure broader personality traits, integrating measurements over multiple relevant representatively chosen contexts offers the best option to capture the full breadth of the trait. Our position follows that of Brunswik (1956) who argued for the use of a design, in which representative situations/contexts were included. In contrast, if the goal is to maximize predictive power, a researcher may focus on the narrower personality trait as expressed in the specific context of the criterion to increase contextual symmetry between the constructs. This approach necessarily will decrease the breadth of the personality construct. Researchers need to be clear about their goals: There is a tradeoff between attempting to assess a trait in all its generality versus attempting to maximize the predictor-criterion correlation by matching the contexts of interest.

It would be informative to further study the impact of contextual cues on the factorial structure of personality

traits (see Robie, Schmit, Ryan, & Zickar, 2000; Schmit et al., 1995 for examples). Different hidden framings and differences in context-generality may contribute to lower intercorrelations of generic personality traits. It can be hypothesized that adding common contextual cues to items may at times increase intercorrelations of personality constructs. For instance, Johnson (2014, p. 86) reported that the NEO-IPIP Neuroticism facet "Self-Consciousness shows a stronger (negative) loading on the Extraversion factor than the Neuroticism factor." The self-consciousness items of the NEO-IPIP-120 primarily feature interactional context cues—cues that are similar to those included in Extraversion items (e.g., behaviors and feelings in social situations). This cue similarity might impact the discriminant validity of the facet. Including representative context information would allow for the systematic investigation of the factorial structure of personality traits in different contexts. The results from such analyses could help to refine item pools and measurement models of personality. It may also help to develop theoretical understanding of the correlational structure of personality traits in certain contexts, thus, contributing to the construct validity of personality measures.

From the perspective that individuals, when responding to trait adjectives, choose individual framings that are functionally equivalent (see also Allport, 1961), one might question any inclusion of contextual information. Based on our framework of hidden (a)symmetries, we argue that correlations of unframed personality assessments with other constructs (e.g., a contextual criterion or other-reports) will often be difficult to interpret. One approach to minimizing this problem would be to identify the framings used by individuals without controlling contextual cues. We might be able to cluster persons into types who construe contexts in items in similar ways (see also Lievens et al., 2008), yielding an improved prediction and interpretation, but one that is conditional on each person's assessed type. As an alternative approach, Rauthmann, Sherman, Nave, and Funder (2015) found that personality affects situation construal. Researchers could experimentally manipulate item wording (contextualization) and examine how these variations impact context construal in interaction with personality. For example, people scoring high in Conscientiousness may more readily use work-related frames-of-reference based on minimal or even no explicit contextual cues in comparison to less conscientious individuals (see also McCune, 2010).

With an increasing degree of contextual specification, a reduction of interpretational differences could be achieved: Whereas adding a simple context-tag may still allow for substantial variability in interpretation, fully contextualized items or even situational vignettes define a clearer domain of behavior in contexts that need to be considered by respondents. This greater clarity may improve our conclusions in the context of studies on predictive power, self-other agreement, and longitudinal research. However, adding more context can sometimes come with its own problems and costs. Increasingly specific indicators may have no or little relevance for some individuals (Olaru et al., 2019). With increasing contextualization, more items are likely needed to fully implement a representative design and to derive a trait score, thereby extending testing time. Although progress has been made in providing taxonomies for contextual information (e.g., Rauthmann et al., 2014; Saucier, Bel-Bahar, & Fernandez, 2007), a *fully* representative design might not be achievable at present.

There might be an optimal level of contextualization, depending on the nature and purpose of the assessment: A level that is specified enough to warrant adequate conclusions, but open enough to be of relevance for most individuals. There might be no single personality test that is optimal for every research design. We encourage researchers to identify the optimal level of contextualization for their specific research questions.

7.3 | Limitations

Several limitations of our approach should be noted. First, we have discussed only one of the elements, context, that may underlie (a)symmetric relationships. Space limitations precluded giving equal attention to other elements such as time frames or reference groups. Debus, Greulich, König, and Kleinmann (2019) explored which time frames participants used when responding to time-unframed job insecurity items. These authors stated that individuals considered different time frames when answering the noncontextualized (unframed) items (see also McCrae et al., 1998). We believe that the frames-of-reference concept could be broadened to not only cover contextual framings, but also time or reference group framings (see Debus et al., 2019, for temporal framings). The lens model in Figure 4 can serve as a starting point to illustrate possible effects of these components. For example, the context ovals could also represent "time frame ovals" and signify differential temporal symmetry between constructs.

Second, the exact nature of contextual information that make up hidden frames-of-reference has been little researched. Traditionally, frames-of-reference have been conceptualized in objective terms, such as locations or interaction partners (e.g., "at school," Schmit et al., 1995; "with friends," Robinson, 2009). We adopted this view in the current article. However, Rauthmann et al. (2014, p. 681; see also Mischel & Shoda, 1995) suggested that besides the more objective situational information, "*psychological characteristics of situations*" are also of great importance. There is a growing body of evidence showing that these characteristics

can also be taxonomized (see Rauthmann et al., 2014). We propose that broadening the view of the frames-of-reference literature to also include psychological characteristics in the equation may be an important avenue for future research. For example, exploring the contextual features in items that drive hidden framings may benefit from consideration of taxonomies such as the Situational Eight DIAMONDS (Rauthmann et al., 2014) in combination with more objective contextual information (e.g., Saucier et al., 2007).

Finally, although we advocate for research using contextualized inventories, studies by Baird and Lucas (2011) as well as Baird, Lucas, and Donnellan (2017) caution us that the *repeated* administration of a contextualized personality inventory may affect the response process or within-person variability indices. Additional methods for the contextual assessment of personality are needed to investigate the consistency of findings. For example, ambulatory assessment methodology (e.g., Sherman et al., 2015) may be a valuable alternative method, because it allows more direct measurement of personality in context. However, it should be borne in mind that these methods also rely on self-report and could, therefore, be prone to problems similar to those reported by Baird et al. (2017).

8 | CONCLUSION

In this article, two largely separate research streams, the literature on the symmetry principle and the frame-of-reference effect, were integrated. Our goal was to raise awareness of a source of variance in personality inventories, context-specificity and generality, that may impact research findings in an often "hidden" way. In accordance with the frame-of-reference literature, we highlight that generic personality items are open to contextual interpretation, contributing to hidden framings. These hidden contextual framings, in turn, may have implications for the contextual symmetry of personality ratings to a variety of measurements such as other-reports and real-life criteria. We hope our review will stimulate new research on hidden framings and hidden (a)symmetries in personality measurement.

CONFLICT OF INTEREST
The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ORCID
Julian Schulze https://orcid.org/0000-0001-8420-493X
Jan-Philipp Freudenstein https://orcid.org/0000-0002-9029-5003
Philipp Schäpers https://orcid.org/0000-0002-8270-5105
Patrick Mussel https://orcid.org/0000-0001-5010-5677
Stefan Krumm https://orcid.org/0000-0002-0840-0864

REFERENCES
Ajzen, I. (2005). Laws of human behavior: Symmetry, compatibility, and attitude-behavior correspondence. In A. Beauducel, B. Biehl, M. Bosnjak, W. Conrad, G. Schönberger, & D. Wagener (Eds.), *Multivariate research strategies: Festschrift in honor of Werner W. Wittmann* (pp. 3–19). Aachen, Germany: Shaker.

Allport, G. W. (1961). *Pattern and growth in personality*, New York, NY: Holt, Rinehart, and Winston.

Anastasi, A. (1983). Traits, states, and situations: A comprehensive view. In H. Wainer & S. Messick (Eds.), *Principals of modern psychological measurement. A Festschrift for Frederic M. Lord* (pp. 345–356). Hillsdale, NJ: Erlbaum.

Angleitner, A., John, O. P., & Löhr, F.-J. (1986). It's *what* you ask and *how* you ask it: An itemmetric analysis of personality questionnaires. In A. Angleitner & J. S. Wiggins (Eds.), *Personality assessment via questionnaires. Current issues in theory and measurement* (pp. 61–108). Berlin, Germany: Springer-Verlag.

Asendorpf, J. B. (1992). A Brunswikean approach to trait continuity: Application to shyness. *Journal of Personality*, *60*, 53–77. https://doi.org/10.1111/j.1467-6494.1992.tb00265.x

Ashton, M. C., & Lee, K. (2009). The HEXACO–60: A short measure of the major dimensions of personality. *Journal of Personality Assessment*, *91*, 340–345. https://doi.org/10.1080/00223890902935878

Baird, B. M., & Lucas, R. E. (2011). "… And how about now?": Effects of item redundancy on contextualized self-reports of personality. *Journal of Personality*, *79*, 1081–1112. https://doi.org/10.1111/j.1467-6494.2011.00716.x

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

Bandura, A. (2015). On deconstructing commentaries regarding alternative theories of self-regulation. *Journal of Management*, *41*, 1025–1044. https://doi.org/10.1177/0149206315572826

Barnett, L. A. (2013). What people want from their leisure: The contributions of personality facets in differentially predicting desired leisure outcomes. *Journal of Leisure Research*, *45*, 150–191. https://doi.org/10.18666/jlr-2013-v45-i2-3010

Beauducel, A., Kersting, M., & Liepmann, D. (2005). A multitrait-multimethod model for the measurement of sensitivity to reward and sensitivity to punishment. *Journal of Individual Differences*, *26*, 168–175. https://doi.org/10.1027/1614-0001.26.4.168

Bing, M. N., Whanger, J. C., Davison, H. K., & VanHook, J. B. (2004). Incremental validity of the frame-of-reference effect in personality scale scores: A replication and extension. *Journal of Applied Psychology*, *89*, 150–157. https://doi.org/10.1037/0021-9010.89.1.150

Bleidorn, W., Hopwood, C. J., & Lucas, R. E. (2018). Life events and personality trait change. *Journal of Personality*, *86*, 83–96. https://doi.org/10.1111/jopy.12286

Bongard, S., Martin, N. M., Seip, M., & al'Absi, M. (2011). Evaluation of a domain-specific anger expression assessment strategy. *Journal of Personality Assessment*, *93*, 56–61. https://doi.org/10.1080/00223891.2010.513705

Borkenau, P., & Liebler, A. (1992). Trait inferences: Sources of validity at zero acquaintance. *Journal of Personality and Social Psychology*, *62*, 645–657. https://doi.org/10.1037/0022-3514.62.4.645

Brunner, M., Nagy, G., & Wilhelm, O. (2012). A tutorial on hierarchically structured constructs. *Journal of Personality*, *80*, 796–846. https://doi.org/10.1111/j.1467-6494.2011.00749.x

Brunswik, E. (1956). *Perception and the representative design of psychological experiments*, Berkeley, CA: University of California Press.

Cantor, N., Mischel, W., & Schwartz, J. C. (1982). A prototype analysis of psychological situations. *Cognitive Psychology*, *14*, 45–77. https://doi.org/10.1016/0010-0285(82)90004-4

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

Costa, P. T. Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*, Odessa, FL: Psychological Assessment Resources.

Costa, P. T. Jr., & McCrae, R. R. (1995). Domains and facets: Hierarchical personality assessment using the Revised NEO Personality Inventory. *Journal of Personality Assessment*, *64*, 21–50. https://doi.org/10.1207/s15327752jpa6401_2

Costa, P. T. Jr., McCrae, R. R., & Dye, D. A. (1991). Facet scales for Agreeableness and Conscientiousness: A revision of the NEO Personality Inventory. *Personality and Individual Differences*, *12*, 887–898. https://doi.org/10.1016/0191-8869(91)90177-d

Davison, H. K., & Bing, M. N. (2009). Content validity does matter for the criterion-related validity of personality tests. *Industrial and Organizational Psychology*, *2*, 501–503. https://doi.org/10.1111/j.1754-9434.2009.01182.x

Debus, M. E., Greulich, B., König, C. J., & Kleinmann, M. (2019). Insecure about how to rate your job insecurity? A two-study investigation into time frames applied to job insecurity measures. *Occupational Health Science*, *3*, 421–435. https://doi.org/10.1007/s41542-019-00049-x

De Raad, B. (1993). Have trait psychologists been ignorant of context? Commentary on the chapter by N.S. Endler. In J. Hettema & I. J. Deary (Eds.), *Foundations of personality* (pp. 269–273). Dordrecht, the Netherlands: Kluwer. https://doi.org/10.1007/978-94-011-1660-2_22

De Raad, B. (2005). Situations that matter to personality. In A. Eliasz, S. E. Hampson, & B. De Raad (Eds.), *Advances in personality psychology* (Vol. *2*, pp. 179–204). Hove, UK: Psychology Press.

De Raad, B., Sullot, E., & Barelds, D. P. H. (2008). Which of the Big Five factors are in need of situational specification? *European Journal of Personality*, *22*, 269–289. https://doi.org/10.1002/per.668

Eid, M., Geiser, C., & Koch, T. (2016). Measuring method effects: From traditional to design-oriented approaches. *Current Directions in Psychological Science*, *25*, 275–280. https://doi.org/10.1177/0963721416649624

Eid, M., Geiser, C., Koch, T., & Heene, M. (2017). Anomalous results in G-factor models: Explanations and alternatives. *Psychological Methods*, *22*, 541–562. https://doi.org/10.1037/met0000083

Eid, M., Lischetzke, T., Nussbeck, F. W., & Trierweiler, L. I. (2003). Separating trait effects from trait-specific method effects in multitrait-multimethod models: A multiple indicator CT-C (M-1) model. *Psychological Methods*, *8*, 38–60. https://doi.org/10.1037/1082-989x.8.1.38

Eisenberg, P. (1941). Individual interpretation of psychoneurotic inventory items. *The Journal of General Psychology*, *25*, 19–40. https://doi.org/10.1080/00221309.1941.10544702

Epstein, S. (1979). The stability of behavior: I. On predicting most of the people much of the time. *Journal of Personality and Social Psychology*, *37*, 1097–1126. https://doi.org/10.1037//0022-3514.37.7.1097

Fisher, D. M., Cunningham, S., Kerr, A. J., & Allscheid, S. P. (2017). Contextualized personality measures in employee selection: Extending frame-of-reference research with job applicant samples. *International Journal of Selection and Assessment*, *25*, 18–35. https://doi.org/10.1111/ijsa.12156

Fleeson, W., & Gallagher, P. (2009). The implications of Big Five standing for the distribution of trait manifestation in behavior: Fifteen experience-sampling studies and a meta-analysis. *Journal of Personality and Social Psychology*, *97*, 1097–1114. https://doi.org/10.1037/a0016786

Funder, D. C. (2012). Accurate personality judgment. *Current Directions in Psychological Science*, *21*, 177–182. https://doi.org/10.1177/0963721412445309

Funder, D. C. (2016). Taking situations seriously: The situation construal model and the Riverside Situational Q-Sort. *Current Directions in Psychological Science*, *25*, 203–208. https://doi.org/10.1177/0963721416635552

Funder, D. C., & West, S. G. (1993). Consensus, self-other agreement, and accuracy in personality judgment: An introduction. *Journal of Personality*, *61*, 457–476. https://doi.org/10.1111/j.1467-6494.1993.tb00778.x

Geukes, K., Nestler, S., Hutteman, R., Küfner, A. C. P., & Back, M. D. (2017). Trait personality and state variability: Predicting individual differences in within-and cross-context fluctuations in affect, self-evaluations, and behavior in everyday life. *Journal of Research in Personality*, *69*, 124–138. https://doi.org/10.1016/j.jrp.2016.06.003

Gignac, G. E., Bates, T. C., & Jang, K. L. (2007). Implications relevant to CFA model misfit, reliability, and the five-factor model as measured by the NEO-FFI. *Personality and Individual Differences*, *43*, 1051–1062. https://doi.org/10.1016/j.paid.2007.02.024

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, *40*, 84–96. https://doi.org/10.1016/j.jrp.2005.08.007

Guttman, R., & Greenbaum, C. W. (1998). Facet theory: Its development and current status. *European Psychologist*, *3*, 13–36. https://doi.org/10.1027/1016-9040.3.1.13

Hayes, A. F., & Dunning, D. (1997). Construal processes and trait ambiguity: Implications for self-peer agreement in personality judgment. *Journal of Personality and Social Psychology*, *72*, 664–677. https://doi.org/10.1037/0022-3514.72.3.664

Heller, D., Perunovic, W. Q. E., & Reichman, D. (2009). The future of person–situation integration in the interface between traits and goals: A bottom-up framework. *Journal of Research in Personality*, *43*, 171–178. https://doi.org/10.1016/j.jrp.2008.12.011

Hirschmüller, S., Egloff, B., Schmukle, S. C., Nestler, S., & Back, M. D. (2015). Accurate judgments of Neuroticism at zero acquaintance: A question of relevance. *Journal of Personality*, *83*, 221–228. https://doi.org/10.1111/jopy.12097

Holtz, B. C., Ployhart, R. E., & Dominguez, A. (2005). Testing the rules of justice: The effects of frame-of-reference and pre-test validity information on personality test responses and test perceptions. *International Journal of Selection and Assessment*, *13*, 75–86. https://doi.org/10.1111/j.0965-075x.2005.00301.x

Jackson, J. J., Hill, P. L., & Roberts, B. W. (2012). Misconceptions of traits continue to persist: A response to Bandura. *Journal of Management*, *38*, 745–752. https://doi.org/10.1177/0149206312438775

Jobe, J. B. (2003). Cognitive psychology and self-reports: Models and methods. *Quality of Life Research*, *12*, 219–227. https://doi.org/10.1023/A:1023279029852

Johnson, J. A. (2014). Measuring thirty facets of the Five Factor Model with a 120-item public domain inventory: Development of the IPIP-NEO-120. *Journal of Research in Personality*, *51*, 78–89. https://doi.org/10.1016/j.jrp.2014.05.003

Judge, T. A., Rodell, J. B., Klinger, R. L., Simon, L. S., & Crawford, E. R. (2013). Hierarchical representations of the five-factor model of personality in predicting job performance: Integrating three organizing frameworks with two theoretical perspectives. *Journal of Applied Psychology*, *98*, 875–925. https://doi.org/10.1037/a0033901

Kim, L. E., Dar-Nimrod, I., & MacCann, C. (2018). Teacher personality and teacher effectiveness in secondary school: Personality predicts teacher support and student self-efficacy but not academic achievement. *Journal of Educational Psychology*, *110*, 309–323. https://doi.org/10.1037/edu0000217

Klein, S. B., Babey, S. H., & Sherman, J. W. (1997). The functional independence of trait and behavioral self-knowledge: Methodological considerations and new empirical findings. *Social Cognition*, *15*, 183–203. https://doi.org/10.1521/soco.1997.15.3.183

Klein, S. B., Cosmides, L., Tooby, J., & Chance, S. (2002). Decisions and the evolution of memory: Multiple systems, multiple functions. *Psychological Review*, *109*, 306–329. https://doi.org/10.1037/0033-295X.109.2.306

Kurtz, J. E., & Palfrey, A. P. (2016). On being the same in different places: Evaluating frame-of-reference effects across two social contexts. *Journal of Personality*, *84*, 113–120. https://doi.org/10.1111/jopy.12144

Lievens, F., De Corte, W., & Schollaert, E. (2008). A closer look at the frame-of-reference effect in personality scale scores and validity. *Journal of Applied Psychology*, *93*, 268–279. https://doi.org/10.1037/0021-9010.93.2.268

Lievens, F., & Sackett, P. R. (2017). The effects of predictor method factors on selection outcomes: A modular approach to personnel selection procedures. *Journal of Applied Psychology*, *102*, 43–66. https://doi.org/10.1037/apl0000160

Lundmann, L., & Villadsen, J. W. (2016). Qualitative variations in personality inventories: Subjective understandings of items in a personality inventory. *Qualitative Research in Psychology*, *13*, 166–187. https://doi.org/10.1080/14780887.2015.1134737

Malouff, J. M., Thorsteinsson, E. B., Schutte, N. S., Bhullar, N., & Rooke, S. E. (2010). The five-factor model of personality and relationship satisfaction of intimate partners: A meta-analysis. *Journal of Research in Personality*, *44*, 124–127. https://doi.org/10.1016/j.jrp.2009.09.004

McCrae, R. R. (2015). A more nuanced view of reliability: Specificity in the trait hierarchy. *Personality and Social Psychology Review*, *19*, 97–112. https://doi.org/10.1177/1088868314541857

McCrae, R. R., Costa, P. T. Jr., & Martin, T. A. (2005). The NEO–PI–3: A more readable revised NEO personality inventory. *Journal of Personality Assessment*, *84*, 261–270. https://doi.org/10.1207/s15327752jpa8403_05

McCrae, R. R., Stone, S. V., Fagan, P. J., & Costa, P. T., Jr. (1998). Identifying causes of disagreement between self-reports and spouse ratings of personality. *Journal of Personality*, *66*, 285–313. https://doi.org/10.1111/1467-6494.00013

McCune, E. A. (2010). *Who am I? Well, it depends: How frame-of-reference imposes context in non-contextualized personality inventories*, Portland, OR: [Doctoral dissertation, Portland State University]. PDXScholar. https://doi.org/10.15760/etd.725

Meitinger, K. (2017). Necessary but insufficient: Why measurement invariance tests need online probing as a complementary tool. *Public Opinion Quarterly*, *81*, 447–472. https://doi.org/10.1093/poq/nfx009

Meyer, J., Fleckenstein, J., Retelsdorf, J., & Köller, O. (2019). The relationship of personality traits and different measures of domain-specific achievement in upper secondary education. *Learning and Individual Differences*, *69*, 45–59. https://doi.org/10.1016/j.lindif.2018.11.005

Mischel, W., & Shoda, Y. (1995). A cognitive-affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review*, *102*, 246–268. https://doi.org/10.1037/0033-295X.102.2.246

Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology*, *60*, 683–729. https://doi.org/10.1111/j.1744-6570.2007.00089.x

Mõttus, R., Allik, J., & Realo, A. (2020). Do self-reports and informant-ratings measure the same personality constructs? *European Journal of Psychological Assessment*, *36*, 289–295. https://doi.org/10.1027/1015-5759/a000516

Murtha, T. C., Kanfer, R., & Ackerman, P. L. (1996). Toward an interactionist taxonomy of personality and situations: An integrative situational—Dispositional representation of personality traits. *Journal of Personality and Social Psychology*, *71*, 193–207. https://doi.org/10.1037//0022-3514.71.1.193

Nye, C. D., Allemand, M., Gosling, S. D., Potter, J., & Roberts, B. W. (2016). Personality trait differences between young and middle-aged adults: Measurement artifacts or actual trends? *Journal of Personality*, *84*, 473–492. https://doi.org/10.1111/jopy.12173

Olaru, G., Schroeders, U., Wilhelm, O., & Ostendorf, F. (2019). "Grandpa, do you like roller coasters?": Identifying age-appropriate personality indicators. *European Journal of Personality*, *33*, 264–278. https://doi.org/10.1002/per.2185

Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin*, *135*, 322–338. https://doi.org/10.1037/a0014996

Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, *41*, 203–212. https://doi.org/10.1016/j.jrp.2006.02.001

Rauthmann, J. F., Gallardo-Pujol, D., Guillaume, E. M., Todd, E., Nave, C. S., Sherman, R. A., … Funder, D. C. (2014). The Situational Eight DIAMONDS: A taxonomy of major dimensions of situation characteristics. *Journal of Personality and Social Psychology*, *107*, 677–718. https://doi.org/10.1037/a0037250

Rauthmann, J. F., Sherman, R. A., Nave, C. S., & Funder, D. C. (2015). Personality-driven situation experience, contact, and construal: How people's personality traits predict characteristics of their situations in daily life. *Journal of Research in Personality*, *55*, 98–111. https://doi.org/10.1016/j.jrp.2015.02.003.

Roberts, B. W. (2007). Contextualizing personality psychology. *Journal of Personality*, *75*, 1071–1082. https://doi.org/10.1111/j.1467-6494.2007.00467.x

Roberts, B. W., Lejuez, C., Krueger, R. F., Richards, J. M., & Hill, P. L. (2014). What is Conscientiousness and how can it be assessed? *Developmental Psychology*, *50*, 1315–1330. https://doi.org/10.1037/a0031109

Robie, C., Risavy, S. D., Holtrop, D., & Born, M. P. (2017). Fully contextualized, frequency-based personality measurement: A replication and extension. *Journal of Research in Personality*, *70*, 56–65. https://doi.org/10.1016/j.jrp.2017.05.005

Robie, C., Schmit, M. J., Ryan, A. M., & Zickar, M. J. (2000). Effects of item context specificity on the measurement equivalence of a personality inventory. *Organizational Research Methods*, *3*, 348–365. https://doi.org/10.1177/109442810034003

Robinson, O. C. (2009). On the social malleability of traits: Variability and consistency in Big 5 trait expression across three interpersonal contexts. *Journal of Individual Differences*, *30*, 201–208. https://doi.org/10.1027/1614-0001.30.4.201

Saucier, G., Bel-Bahar, T., & Fernandez, C. (2007). What modifies the expression of personality tendencies? Defining basic domains of situation variables. *Journal of Personality*, *75*, 479–504. https://doi.org/10.1111/j.1467-6494.2007.00446.x

Saucier, G., & Conley, M. (2015). Separating personality and situation. *European Journal of Personality*, *29*, 410–411. https://doi.org/10.1002/per.2005

Schäpers, P., Mussel, P., Lievens, F., König, C. J., Freudenstein, J.-P., & Krumm, S. (2020). The role of situations in situational judgment tests: Effects on construct saturation, predictive validity, and applicant perceptions. *Journal of Applied Psychology*, *105*, 800–818. https://doi.org/10.1037/apl0000457

Schmit, M. J., Ryan, A. M., Stierwalt, S. L., & Powell, A. B. (1995). Frame-of-reference effects on personality scale scores and criterion-related validity. *Journal of Applied Psychology*, *80*, 607–620. https://doi.org/10.1037//0021-9010.80.5.607

Seeboth, A., & Mõttus, R. (2018). Successful explanations start with accurate descriptions: Questionnaire items as personality markers for more accurate predictions. *European Journal of Personality*, *32*, 186–201. https://doi.org/10.1002/per.2147

Shaffer, J. A., & Postlethwaite, B. E. (2012). A matter of context: A meta-analytic investigation of the relative validity of contextualized and noncontextualized personality measures. *Personnel Psychology*, *65*, 445–494. https://doi.org/10.1111/j.1744-6570.2012.01250.x

Sherman, R. A., Rauthmann, J. F., Brown, N. A., Serfass, D. G., & Jones, A. B. (2015). The independent effects of personality and situations on real-time expressions of behavior and emotion. *Journal of Personality and Social Psychology*, *109*, 872–888. https://doi.org/10.1037/pspp0000036

Small, E. E., & Diefendorff, J. M. (2006). The impact of contextual self-ratings and observer ratings of personality on the personality–performance relationship. *Journal of Applied Social Psychology*, *36*, 297–320. https://doi.org/10.1111/j.0021-9029.2006.00009.x

Ten Berge, M. A., & De Raad, B. (1999). Taxonomies of situations from a trait psychological perspective. A review. *European Journal of Personality*, *13*, 337–360. https://doi.org/10.1002/(SICI)1099-0984(199909/10)13:5<337:AID-PER363>3.0.CO;2-F

Tett, R. P., & Guterman, H. A. (2000). Situation trait relevance, trait expression, and cross-situational consistency: Testing a principle of trait activation. *Journal of Research in Personality*, *34*, 397–423. https://doi.org/10.1006/jrpe.2000.2292

Uher, J. (2018). Quantitative data from rating scales: An epistemological and methodological enquiry. *Frontiers in Psychology*, *9*, 2599. https://doi.org/10.3389/fpsyg.2018.02599

Van Heck, G. L., Perugini, M., Caprara, G.-V., & Fröger, J. (1994). The Big Five as tendencies in situations. *Personality and Individual Differences*, *16*, 715–731. https://doi.org/10.1016/0191-8869(94)90213-5

Werner, P. D., & Pervin, L. A. (1986). The content of personality inventory items. *Journal of Personality and Social Psychology*, *51*, 622–628. https://doi.org/10.1037//0022-3514.51.3.622

Wittmann, W. W. (1988). Multivariate reliability theory. Principles of symmetry and successful validation strategies. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (pp. 505–560). New York, NY: Plenum. https://doi.org/10.1007/978-1-4613-0893-5_16

Wittmann, W. W., & Klumb, P. L. (2006). How to fool yourself with experiments in testing theories in psychological research. In R. R. Bootzin & P. E. McKnight (Eds.), *Strengthening research methodology: Psychological measurement and evaluation* (pp. 185–211). Washington, DC: American Psychological Association. https://doi.org/10.1037/11384-010

Wright, J. C., & Mischel, W. (1987). A conditional approach to dispositional constructs: The local predictability of social behavior. *Journal of Personality and Social Psychology*, *53*, 1159–1177. https://doi.org/10.1037/0022-3514.53.6.1159

Wrzus, C., Wagner, G. G., & Riediger, M. (2016). Personality-situation transactions from adolescence to old age. *Journal of Personality and Social Psychology*, *110*, 782–799. https://doi.org/10.1037/pspp0000054

Zech, A., Bühner, M., Kröner, S., Heene, M., & Hilbert, S. (2017). The impact of symmetry: Explaining contradictory results concerning working memory, reasoning, and complex problem solving. *Journal of Intelligence*, *5*, 22. https://doi.org/10.3390/jintelligence5020022

Ziegler, M., Horstmann, K. T., & Ziegler, J. (2019). Personality in situations: Going beyond the OCEAN and introducing the Situation Five. *Psychological Assessment*, *31*, 567–580. https://doi.org/10.1037/pas0000654