Article

Development and Validation of the Certainty About Mental States Questionnaire (CAMSQ): A Self-Report Measure of Mentalizing Oneself and Others

Sascha Müller[1]* iD, Leon P. Wendt[1]* iD,
and Johannes Zimmermann[1] iD

Abstract

The Certainty About Mental States Questionnaire (CAMSQ) is a self-report measure of the perceived capacity to understand mental states of the self and others (i.e., mentalizing). In two studies (total $N = 1828$), we developed the CAMSQ in both English and German as a two-dimensional measure of Self- and Other-Certainty, investigated associations with other measures of mentalizing, and explored relationships to personality functioning and mental health. The CAMSQ performed well in terms of convergent and discriminant validity, internal consistency, test-retest reliability, and measurement invariance across the United States and Germany. The present research indicates that the CAMSQ assesses maladaptive forms of having too little or too much certainty about mental states (consistent with hypomentalizing and hypermentalizing). A psychologically adaptive profile of perceived mentalizing capacity appears to be characterized by high Self-Certainty that exceeds Other-Certainty, suggesting that imbalances between Self-Certainty and Other-Certainty (Other-Self-Discrepancy) play an important role within personality pathology.

How individuals perceive and interpret the mental states (e.g., thoughts, feelings, motives) of themselves and others plays a major role in how well individuals can adapt to their social environment and maintain psychological health. This concept can be referred to as *mentalizing* or *reflective functioning* (e.g., Allen et al., 2008; Fonagy et al., 2016). Mentalizing oneself fosters a narrative identity that integrates contradictory aspects into a coherent and stable representation of the self, thereby contributing to an accepting self-view (Fuchs, 2007). Mentalizing others is seen as crucial for adaptive social relations as it enables individuals to comprehend and predict behavior in terms of the intentions, beliefs, and desires of others (Bateman & Fonagy, 2019). Mentalizing has emerged as a cross-cutting theme that is of interest to psychological researchers in various fields. For example, clinical researchers pursue the question of how exactly mentalizing capacities relate to psychological functioning and the etiology of mental disorders (e.g., Fonagy et al., 2017), while specialized psychotherapeutic approaches are targeted at improving mentalizing capacities (Allen & Fonagy, 2006; Bateman & Fonagy, 2016). In frameworks such as interpersonal theory (e.g., Hopwood, 2018), perceptual distortions of mentalizing oneself and others are thought to contribute to recurring maladaptive interpersonal experiences that are observed in individuals

with personality pathology. However, what exactly characterizes a maladaptive way of perceiving oneself and others is not fully understood. In the present research, we aim to shed more light on this issue by developing a new self-report questionnaire of mentalizing that addresses the potential shortcomings of existing measures. Specifically, we introduce the first measure of mentalizing that is able to capture imbalances in the certainty associated with perceiving the mental states of oneself and others, and demonstrate the relevance of this construct for understanding personality pathology.

The term mentalizing is sometimes used synonymously with constructs in the realm of social cognition such as cognitive empathy, perspective-taking, and theory of mind (e.g., Olderbak & Wilhelm, 2020). As opposed to mentalizing that incorporates interpreting both the self and others (Fonagy & Target, 2006), however, these constructs are usually narrower in their definition, for example, by only

[1]University of Kassel, Germany

*S.M. and L.P.W. shared first authorship.

Corresponding Author:
Sascha Müller and Leon Wendt, Department of Psychology, University of Kassel, Holländische Straße 36-38, 34127 Kassel, Germany.
Emails: sascha.mueller@uni-kassel.de; l.wendt@uni-kassel.de

pertaining to the mental states of others and focusing on emotion recognition (e.g., cognitive empathy: "the ability to detect and understand emotional displays"; Ickes, 1993; Vachon & Lynam, 2016, p. 136). Furthermore, mentalizing is used to describe distinct concepts with the accuracy of inferring mental states as assessed by performance measures (e.g., Baron-Cohen et al., 2001) on the one hand and the subjective sense of one's mentalizing capacity as assessed by self-report (e.g., Fonagy et al., 2016) on the other. These assessments concern relatively independent constructs as indicated by their low convergence (e.g., Fonagy et al., 2016; Murphy & Lilienfeld, 2019; Realo et al., 2003).

According to mentalizing theory (e.g., Bateman & Fonagy, 2019; Luyten et al., 2020), how individuals *perceive* their own mentalizing capacities can be categorized into three qualitatively distinct forms of mentalizing. These variants are primarily distinguished by the degree of certainty that is involved in making mentalizing inferences. *Genuine mentalizing* is characterized by acknowledging the opaqueness of mental states (Fonagy et al., 2016), that is, being apprehensive of the circumstance that individuals are not always aware of their own inner workings and those of others. As Fonagy and colleagues put it, "a genuine mentalizing stance is characterized by modesty about knowing one's own mental states and humility in relation to knowing the mental states of others" (p. 3). The other two forms are variants of impaired mentalizing that reflect deviations from a presumed optimal level of certainty. *Hypermentalizing* (or pseudomentalizing) is characterized by excessive certainty about one's mentalizing capacity. Individuals prone to hypermentalizing tend to overinterpret mental states (Sharp et al., 2011; Sharp & Vanwoerden, 2015), perceive themselves as exceptionally capable in mentalizing, and engage in lengthy narratives about their own or others' mental states that lack connection to reality (e.g., Bateman & Fonagy, 2019). Such individuals might exhibit socially inadequate behaviors by jumping to rash and ill-advised conclusions about others, for example, by misinterpreting their motives as malignant (Bo et al., 2017). In this vein, hypermentalizers can be experienced as intrusive or offensive which might harm interpersonal relationships or prevent their formation. Conversely, *hypomentalizing* (or undermentalizing) is characterized by low certainty about mentalizing inferences and limited engagement in mentalizing in general (Bateman & Fonagy, 2019; Luyten et al., 2020). Individuals prone to hypomentalizing may exhibit an inability to consider complex explanatory models of mental states and instead provide simplistic narratives of human behavior.

It is commonly stated that deficiencies in mentalizing are a contributing factor to maladaptive patterns of feeling, thinking, and behaving. Although not termed as such, the concept of mentalizing was prominently incorporated in Criterion A (Bender et al., 2011) of the Alternative Model for Personality Disorders (AMPD) in the *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.; DSM-5; American Psychiatric Association [APA], 2013). Corresponding to genuine mentalizing, healthy personality functioning in Criterion A is (among other features) described as the ability to self-reflect productively and to comprehend and appreciate others' experiences and motivations. Generally, all personality disorders (PDs) are presumed to involve distorted views of the self and others including an impoverished or fragile self-concept and difficulties with self-other differentiation (Skodol et al., 2011). In contemporary interpersonal theory, perceptual distortions of the self and others drive recurring patterns of interpersonal conflict, disrupt relationships, prevent the fulfillment of individuals' communal and agentic motives, and inflict symptom distress on the individual (Hopwood, 2018; Pincus et al., 2020). Moreover, developmental changes in self- and other-perception may play a key role in personality maturation (Geukes et al., 2018), a concept that describes the average trend toward more socially adaptive personality configurations over the course of life (e.g., Bleidorn, 2015; Orth et al., 2010). Mentalizing capacity is thought to be shaped by parental interactions in early childhood (Fonagy & Luyten, 2016) and to develop with age during another critical developmental period in adolescence (Fonagy, 1991; Poznyak et al., 2019).

Mentalizing Assessment

The relevance of mentalizing impairments for psychological functioning has been researched extensively using a variety of assessment methods. Traditionally, the Reflective Functioning Scale (RFS; Fonagy et al., 1998; Taubner et al., 2013) is considered a gold standard of mentalizing assessment. This method involves clinician ratings that are applied to the Adult Attachment Interview (George et al., 1985) or the Parent Development Interview (Slade et al., 2004). Empirical findings support the notion that RFS scores are associated with psychopathology (e.g., Antonsen et al., 2016; Bouchard et al., 2008; Fonagy et al., 1996; Taubner et al., 2013). However, the RFS was developed to assess the capacity to mentalize attachment relationships with a focus on parental relationships and its assessment is very time-consuming. The Movie for the Assessment of Social Cognition (MASC; Dziobek et al., 2006) is a measure of mentalizing that tasks participants with watching a video and inferring the mental states of actors who are shown in social interactions. MASC scores distinguish between hypomentalizing and hypermentalizing and are associated with psychopathology (e.g., Fossati et al., 2017, 2018; for a meta-analysis, see McLaren et al., 2021) but the administration of the MASC is also rather time-consuming and limited

to the mentalizing of others. In addition, maladaptive forms of mentalizing may primarily manifest themselves in the context of interpersonal relationships (Bo et al., 2017) and might thus not be optimally measured under laboratory conditions when interpreting the mental states of actors with no possibility to interact.

Due to the limitations of the assessment methods described above, self-report measures can be an efficient tool to study mentalizing impairments. The Reflective Functioning Questionnaire (RFQ; Fonagy et al., 2016) has emerged as a popular tool for assessing mentalizing by self-report. The RFQ is intended to measure both hypomentalizing and hypermentalizing but critical concerns have been raised with regard to its validity for assessing hypermentalizing in particular (Müller et al., 2021). Although there is ample evidence on its associations with psychopathology and personality pathology, it is mostly limited to low certainty about mental states as an indicator of hypomentalizing (e.g., de Meulemeester et al., 2017; Fonagy et al., 2016; Müller et al., 2021; Spitzer, Zimmermann, et al., 2021). Further mentalizing questionnaires such as the Mentalization Questionnaire (MZQ; Hausberg et al., 2012) and the Mentalization Scale (MentS; Dimitrijević et al., 2018) are available, but comparatively little is known about their validity and they are not intended to assess hypermentalizing but rather mentalizing problems in general. Still other measures such as the RFQ-Youth (Sharp et al., 2021) or the Parental RFQ (Luyten et al., 2017) are targeted at specific populations.

In our view, a self-report measure of mentalizing should first and foremost cover a core definition of the construct as, for example, provided by Fonagy et al. (2016, p. 1), where mentalizing is described as "the capacity to interpret both the self and others in terms of internal mental states, such as feelings, wishes, goals, desires, and attitudes." A shared feature of existing mentalizing questionnaires (see Supplemental Table S1 for an interpretation of the item content from the authors' point of view), however, is that their item content might conflate the given definition with assumed psychopathological consequences of mentalizing impairments such as problems in interpersonal relationships, impulsive behavior, or emotion dysregulation (e.g., Bo et al., 2017; Müller et al., 2021). This issue impedes investigating the associations between mentalizing and psychopathology because correlation coefficients may be inflated merely due to shared item content. Another shortcoming of existing questionnaires might be that none of these measures except for the MentS clearly distinguish between inferring the mental states of the self and others. For example, both the RFQ and the MZQ mainly pertain to mental states of the self. Finally, none of the questionnaires offers a balanced representation of different mental states such as feelings, thoughts/attitudes, and motives/goals.

Developing the Certainty About Mental States Questionnaire (CAMSQ)

Given the clinical relevance of mentalizing for understanding personality pathology, the convenience of administering self-report measures, and the potential shortcomings of existing questionnaires, new test developments could contribute to moving the field forward. We propose the Certainty About Mental States Questionnaire (CAMSQ) for assessing the self-reported certainty involved in making inferences about the mental states of the self and others (Self-Certainty and Other-Certainty), thereby following the conceptualization of mentalizing that aligns with perceived mentalizing capacity (e.g., Fonagy et al., 2016). It has been pointed out that mentalizing can involve the self or others as a target, affective (feelings) or cognitive (thoughts or motives) content, inferences based on external (observable) or internal (needs to be inferred from behavior) cues, and automatic or controlled processes (e.g., Luyten et al., 2020). To account for aspects that can be assessed via self-report, we constructed items reflecting different mental states content (i.e., feelings, thoughts and attitudes, goals and motives) of both the self and others in a balanced manner. An important strategic decision was to focus on a narrow definition pertaining to inferring mental states and to refrain from including potential consequences of impaired mentalizing (e.g., emotion dysregulation). To promote the generalizability of the measure, we simultaneously created two parallel sets of items in both English and German. Items were created by the authors based on consensus and the back-translation method was used to ensure comparability of the meaning conveyed by the items in the two languages. The final items were checked by a professional bilingual translator.

To assess individuals' certainty about mental states, we opted to present respondents with statements about mentalizing inferences (e.g., Item 2: "I can tell whether another person is at peace with themselves") and employ a response scale in which respondents indicate how frequently they mentalize accurately. We assume that reporting more frequent instances of successful mentalizing conveys a higher subjective certainty about one's mentalizing inferences. Furthermore, using a frequency scale ranging from *never* to *always* offers clear and definitive endpoints of behavioral frequencies (Bocklisch et al., 2012) as compared to more commonly used endorsement-based response formats (as used in other mentalizing questionnaires). Apart from that, endorsement-based scales often include frequency quantifiers such as "sometimes" or "often" as part of the items (e.g., Item 6 of the RFQ: "Sometimes I do things without really knowing why"). However, using frequency quantifiers in conjunction with an endorsement scale may be ambiguous with regard to the elicited response processes. For

example, rejecting the above statement may indicate that individuals actually *never* or *always* "do things without really knowing why." Such items may hence mask the distinction between extremely low and extremely high levels of subjective certainty. Indeed, the central aim of the new test development was to derive a measure that is capable of capturing and distinguishing between both maladaptive variants of impaired mentalizing (hypo- and hypermentalizing) as well as genuine mentalizing. As yet, there is no clear evidence that any of the existing self-report questionnaires of mentalizing can assess a maladaptive form of having too much certainty about mental states (hypermentalizing). For instance, the certainty pole of the RFQ is rather consistently associated with mental health rather than psychopathology (e.g., Fonagy et al., 2016; Müller et al., 2021).

By addressing the shortcomings of previous self-report measures of mentalizing, we considered that hypermentalizing might manifest itself in the CAMSQ in various ways. First, it was conceivable that hypomentalizing and hypermentalizing may be associated with low and high levels of certainty, respectively. Such a pattern would be evident in U-shaped associations between certainty about mental states and indicators of personality pathology or symptom distress (and inverse U-shaped associations with indicators of mental health). For example, when predicting personality dysfunction by means of certainty about mental states, both low and high levels of certainty would be associated with high dysfunction, whereas middle levels of certainty would exhibit low dysfunction. Previous research, however, found no evidence for such associations using the RFQ (Müller et al., 2021; Spitzer, Zimmermann, et al., 2021). Second, irrespective of the fallible nature of knowing oneself and others (e.g., Vazire, 2010), we argue that individuals should, in general, find it easier to perceive and interpret their own mental states as compared to others' mental states (e.g., Thornton et al., 2019) because others' mental states need to be inferred from external cues, whereas individuals can directly experience their own mental states consciously. This would translate to higher levels of Self-Certainty as compared to Other-Certainty for the average individual. It might thus be important to also examine different configurations of Self- and Other-Certainty by exploring their unique associations with psychopathology. Although not explicitly stated in mentalizing theory, one could surmise that hypermentalizing is primarily salient in inferring others' mental states because it is often portrayed in this way in case vignettes that are based on clinical experience (e.g., Bateman & Fonagy, 2019; Bo et al., 2017). Considering some of the more general descriptions of PDs in DSM-5 (i.e., unstable self-image vs. suspiciousness toward others; APA, 2013), it is conceivable that hypermentalizing could also be reflected in having greater certainty about knowing others than about knowing oneself.

The Present Research

This research aims to test the CAMSQ as a self-report measure of perceived mentalizing capacity. We investigate its internal structure and reliability, associations with convergent and discriminant constructs, and its capability of assessing maladaptive forms of having too much and too little certainty about mental states. We report on four data collections from two countries, namely, the United States and Germany. Although impaired mentalizing may be a particularly prominent feature of psychologically disturbed populations, we deliberately rely on samples drawn from the general population to delineate the construct across the full range of personality and psychological functioning. We use modern techniques of item selection to optimize the psychometric performance of the questionnaire. By conducting a first exploratory study (Study 1) and replicating the central findings in preregistered independent data collections (Study 2), we follow recommendations to more clearly separate exploratory from confirmatory analyses (Wagenmakers et al., 2012).

Study 1

In Study 1, we first aimed to explore the dimensionality of the initial CAMSQ item pool. As our construction rationale was focused on creating items that assess the perceived capacity to interpret mental states of the self and others, we expected the emergence of two prominent factors reflecting certainty about one's own mental states and certainty about others' mental states. However, it was also conceivable that content-specific factors might emerge reflecting different types of mental states (e.g., feelings, thoughts and attitudes, and goals and motives). Furthermore, we aimed to derive a final version of the CAMSQ that is equivalent in English and German and psychometrically optimized in both languages in terms of providing good fit to simple structure, high internal consistency, and strong measurement invariance. To achieve this, we relied on a meta-heuristic that enables the optimization of multiple statistical criteria simultaneously.

Using the final CAMSQ, we then explored its nomological network in terms of convergent and discriminant correlations. We expected the CAMSQ to exhibit strong associations with constructs that tap into perceived mentalizing capacity. Specifically, we used the RFQ (Fonagy et al., 2016) as a measure mainly concerning mentalizing the self while using measures of cognitive empathy and perspective-taking as measures concerning mentalizing others (e.g., Ickes, 1993). *Affective empathy* refers to emotional or behavioral reactions to others' mental states such as empathic concern, sympathy, and compassion (e.g., Vachon & Lynam, 2016). As inferring others' mental states and reacting to these are conceptually different, we expected

rather low correlations with affective empathy. Considering that the CAMSQ is thought to assess the perceived success of one's mentalizing activities and might thus be influenced by the positivity of self-views, we also examined correlations with constructs that are related to positive self-evaluation (i.e., self-esteem and subclinical narcissism).

We further included performance measures of emotion recognition and general cognitive ability (i.e., crystallized intelligence). As noted above, performance in emotion recognition tasks exhibits low convergence with self-reported mentalizing capacity (e.g., Fonagy et al., 2016; Murphy & Lilienfeld, 2019) which is why we did not expect to find substantial associations between the CAMSQ and such tasks. However, employing performance measures offered the possibility to operationalize perceived mentalizing capacity by asking respondents how they perceived their own performance in a task. On the premise that the CAMSQ measures perceived mentalizing capacity, we thus suspected that the CAMSQ would be related to *perceived* performance in the emotion recognition tasks but not necessarily to perceived performance in cognitive ability tasks.

Finally, we investigated the associations between the CAMSQ and indicators of personality pathology and symptom distress to examine the (mal)adaptivity of different levels and configurations of certainty about mental states. To operationalize personality dysfunction, we relied on the Personality Inventory for DSM-5 Brief Form + Modified (PID5BF+M; Bach et al., 2020) because it not only captures stylistic features of personality pathology with its domains (i.e., Criterion B of the AMPD) but it can also be used to extract a broad indicator of severity (i.e., similar to Criterion A) from the common variance of these domains (e.g., Zimmermann et al., 2020).

Method

Samples in Study 1. For the first study, we recruited participants online via panel providers from the United States (i.e., Amazon Mechanical Turk [Mturk]) and Germany (i.e., Respondi). Data quality was ensured by a series of attention and validity checks (for details on exclusion criteria, see Supplemental Note S1). Participants received minimum wage as compensation or comparable rewards. There were no missing data as individuals were not able to proceed without answering each item. We aimed to collect samples of about $N = 500$ to achieve stable estimation of latent associations (Kretzschmar & Gignac, 2019).

U.S. Sample 1. The sample consisted of two separate but subsequent data collections on Mturk of which the first comprised 256 valid cases and the second comprised 263 valid cases after data cleaning. Thus, the combined sample was $N = 519$. Participants' (56% female) age ranged from 18 to 74 years ($M = 37.9$, $SD = 12.5$). With 70% holding a bachelor's degree or higher, participants had on average a higher level of education than the general population. The majority of participants was employed for wages (64%).

German Sample 1. The sample was recruited via the panel provider Respondi. After excluding participants who failed validity checks, the sample comprised 505 valid cases. Individuals were roughly representative of the German general population in terms of age and gender. Participants' (51% female) age ranged from 18 to 81 years ($M = 46.4$, $SD = 15.4$). With regard to the educational level (e.g., 28% with a bachelor's degree or higher) and occupational status (e.g., 54% employed for wages), there was broad variation.

Measures and Procedure in U.S. Sample 1. The first round of data collection was conducted as part of another project that will be reported elsewhere. Participants completed a range of measures of which only the CAMSQ item pool and the PID5BF+M were relevant for the present research. For the second round of data collection, all measures that were administered pertained to this study and are listed below. In the study, participants first completed the self-report measures of mentalizing including the CAMSQ item pool before taking the task-based measures. Further details on the measures' internal consistencies and fit indices of their measurement models are provided in the supplement (Supplemental Table S2).

CAMSQ Item Pool. The full CAMSQ item pool entailed 40 self-descriptive statements that were answered on a 7-point frequency scale ranging from *never* (1) to *always* (7) with intermediate response options of *almost never, sometimes, half of the time, often*, and *almost always*. The response options were chosen to provide roughly equidistant frequency quantifiers across the middle categories and finer graduations at the extremes (i.e., *almost never/almost always, never/always*; Bocklisch et al., 2012). The statements are geared toward successful mentalization processes and are thus assumed to capture variation in perceived mentalizing capacity (i.e., subjective certainty about mental states). Roughly half of the items pertained to interpreting one's own mental states (e.g., Item 1: "I know my innermost wishes and desires"), while the other half pertained to interpreting others' mental states (e.g., Item 2: "I can tell whether another person is at peace with themselves"). An initial set of 31 CAMSQ items was administered in the first round of data collection on Mturk, after which 9 additional items were added and used in all subsequent data collections. We expanded the item pool to increase the likelihood of finding a psychometrically optimal item selection. To substantiate our interpretation of the frequency scale as measuring certainty about mental states, we further included two additional response formats using the same core CAMSQ

items and examined the convergence between response formats (see Supplemental Note S2 for details, analyses, and results).

Personality Inventory for DSM-5 *Brief Form + Modified (PID5BF+M).* The PID5BF+M (Bach et al., 2020) is a 36-item measure that assesses six broad dimensions of personality pathology (i.e., negative affectivity, antagonism, psychoticism, disinhibition, detachment, and anankastia) as outlined in DSM-5 (APA, 2013) and in the *International Classification of Diseases 11th Revision* (ICD-11; World Health Organization, 2018). The items (e.g., "I get emotional easily, often for very little reason") are rated on a 4-point scale ranging from *very false or often false* (0) to *very true or often true* (3). For the use in latent variable models, we used parcels based on observed PID5BF+M facet scores to model the PID-5 domains (e.g., manipulativeness, grandiosity, and deceitfulness as indicators of antagonism). In addition, we modeled a higher-order factor (higher-order confirmatory factor analysis [CFA]) explaining the positive correlations between first-order factors (i.e., domains). This higher-order factor reflects the severity of personality dysfunction (e.g., Gomez et al., 2020; Zimmermann et al., 2020).

Affective and Cognitive Measure of Empathy (ACME). The ACME (Vachon & Lynam, 2016) is a 36-item self-report measure of empathy. Items (e.g., "I have a hard time figuring out what someone else is feeling") are answered on a 5-point scale ranging from *disagree strongly* (1) to *agree strongly* (5). In this study, only the 12-item subscale assessing cognitive empathy was administered.

Basic Empathy Scale (BES). The BES (Jolliffe & Farrington, 2006) is a 20-item self-report measure of empathy. Items (e.g., "I can usually work out when my friends are scared") are rated on a 5-point scale ranging from *strongly disagree* (1) to *strongly agree* (5). In this study, only the 8-item subscale assessing cognitive empathy was administered.

Interpersonal Reactivity Index (IRI). The IRI (Davis, 1983) is a 28-item self-report measure of empathy. This study used only the 7-item perspective-taking scale. Items (e.g., "I believe that there are two sides to every question and try to look at them both") are answered on a 5-point scale ranging from *does not describe me well* (0) to *describes me very well* (4).

Reflective Functioning Questionnaire (RFQ). The RFQ (Fonagy et al., 2016) is an 8-item self-report measure of mentalizing that is rated on a 7-point scale ranging from *strongly disagree* (1) to *strongly agree* (7). Items primarily reflect states of uncertainty about one's own mental states (e.g., "I don't always know why I do what I do"). In accordance with recent recommendations (Müller et al., 2021), we refrained

from applying the originally proposed scoring procedure by Fonagy et al. (2016) because it results in statistically redundant scales (RFQ_C and RFQ_U). Instead, we scored the mean of a psychometrically optimized 6-item version of the scale (RFQ-6). This unidimensional construct can be interpreted as reflecting a continuum from low to high uncertainty about mental states (i.e., hypomentalizing).

Rosenberg Self-Esteem Scale (RSES). The RSES (Rosenberg, 1965) is a 10-item self-report measure of self-esteem. Items (e.g., "On the whole, I am satisfied with myself") are answered on a 4-point scale ranging from *strongly disagree* (1) to *strongly agree* (4).

Reading the Mind in the Eyes Test (RMET). The RMET (Baron-Cohen et al., 2001) is a 36-item task-based measure assessing emotion recognition based on visual cues. In the task, participants view pictures of persons' eye regions and have to decide which of the four presented terms best describes the mental state of the target (e.g., Item 12: "skeptical", "indifferent", "embarrassed", and "dispirited"). In this study, a psychometrically optimized 10-item version of the RMET was used (Olderbak et al., 2015). The sum of correct answers was used as an indicator of an individual's performance.

Geneva Emotion Knowledge Test—Blends Brief Form (GEMOK-Blends). The brief form of the GEMOK-Blends (Schlegel & Scherer, 2018) is a 10-item task-based measure of emotion recognition based on text descriptions. In the task, participants read scenarios involving two emotional experiences of a target person and decide which of the five pairs of terms best describes the mental states that the target has experienced (e.g., Item 2: "joy and pride", "happiness and joy", "surprise and pleasure", "joy and pleasure", and "pleasure and happiness"). The sum of correct answers was used as an indicator of an individual's performance.

Vocabulary Test. We created a 12-item task-based measure of vocabulary knowledge which we derived from 45 synonym items of the Vocabulary IQ Test (Open-Source Psychometrics Project, 2020). In each item, participants have to identify the two synonyms (e.g., "alone" and "solo") among a list comprising five English words in total including three distractors (e.g., "inverted", "drunk", and "worldly"). For the short 12-item version, we aimed to derive a set of items with balanced difficulty to optimize discriminatory power across the latent trait spectrum (i.e., vocabulary knowledge). Because we assumed that the difficulty of the items strongly depends on how frequently the presented target words appear in the English language and given that no empirical data on the 45-item Vocabulary IQ Test were available at the time, we based our selection on word frequency data (Davies, 2011) from the Corpus of Contemporary American English (Davies, 2008). Items

were presented in consecutive random order with a time limit of 15 seconds per item to prevent participants from looking up the correct answer. The sum of correct answers was used as an indicator of vocabulary knowledge. The test provided adequate fit to a unidimensional model and showed very good internal consistency for a brief task-based measure with ω = .82.

Perceived Performance in the Tasks. After each completed task-based test, participants were asked to indicate how many items they thought they had solved correctly. Participants responded on a point scale ranging from believing they had solved none of the items correctly (0) to believing they had solved every item correctly (10 or 12, respectively).

Measures and Procedure in German Sample 1. In the German sample, the CAMSQ item pool, the PID5BF+M, the RMET, and the GEMOK-Blends were administered as well. Measures that were not used in the U.S. sample are described below. As in the U.S. sample, mentalizing questionnaires were administered before the tasks. Details on measures' internal consistencies and fit indices of their measurement models are provided in the supplement (Supplemental Table S2).

Empathy Quotient (EQ). The EQ (Baron-Cohen & Wheelwright, 2004) is a 40-item self-report measure of empathy. This study used only the 9-item cognitive empathy scale and the 8-item emotional reactivity scale as measures of cognitive and affective empathy, respectively. Items (e.g., "I often find it difficult to judge if something is rude or polite") are rated on a 4-point scale ranging from *strongly disagree* (1) to *strongly agree* (4).

Narcissistic Admiration and Rivalry Questionnaire (NARQ). The NARQ (Back et al., 2013) is an 18-item self-report measure assessing two dimensions of subclinical grandiose narcissism (i.e., narcissistic admiration and rivalry) that are conceptualized as two distinct processes to achieve self-promotion. The items (e.g., "I deserve to be seen as a great personality") are rated on a 6-point scale ranging from *strongly disagree* (1) to *strongly agree* (6).

Berlin Test for the Assessment of Fluid and Crystallized Intelligence—Short Form Crystallized Intelligence (BEFKI GC-K). The BEFKI GC-K (Schipolowski et al., 2013) is a task-based measure for assessing general knowledge with 12 questions. Participants are presented with knowledge questions on various topics (e.g., "What is amber made of?") and are instructed to choose the correct answer from a list of four response options (e.g., "from volcanic magma", "from fossil resin", "from silicates", and "from crystals"). Items were presented in consecutive random order with a time limit of 20 seconds per item to prevent participants from

looking up the correct answer. The sum of correct answers was used as an indicator of crystallized intelligence.

Symptom Checklist K9 (SCL-K9). The SCL-K9 (Petrowski et al., 2019) is a 9-item self-report measure of symptom distress. The scale is a short form of the SCL (Derogatis, 1977; Franke, 2000). On a 5-point scale ranging from *not at all* (0) to *extremely* (4), respondents indicate whether they were bothered by a list of psychological symptoms (e.g., "Feeling uptight or agitated") during the past week.

Satisfaction With Life Scale (SWLS). The SWLS (Diener et al., 1985; Janke & Glöckner-Rist, 2014) is a 5-item self-report measure of life satisfaction. Items (e.g., "I am satisfied with my life") are answered on a 7-point scale ranging from *strongly disagree* (1) to *strongly agree* (7).

Statistical Analyses. Analyses were performed using R version 4.0.3 (R Core Team, 2020), the packages of lavaan (Rosseel, 2012) and stuart (Schultze, 2020), and Mplus 8.4 (Muthén & Muthén, 1998–2019). We provide open data and scripts for reproducing the analyses at https://osf.io/ea5nx/.

Results

Dimensionality of the CAMSQ. We first explored the dimensionality of the 40 CAMSQ items. Parallel analysis using maximum likelihood factor extraction indicated three factors in the U.S. sample and four factors in the German sample. However, the scree plot suggested only two strong factors in both samples in terms of magnitude of eigenvalues as well as the elbow criterion (Supplemental Figure S1). To better understand the substantive nature of factor solutions, we performed exploratory factor analyses (EFA) in both samples. We estimated correlated factors using geomin rotation and full information maximum likelihood with robust test statistics (MLR). In solutions comprising more than two factors, third and fourth factors mostly exhibited low factor loadings, were not robust across samples, and were not meaningfully interpretable. We thus did not consider Factors 3 and 4 as relevant for the intended instrument as they were not sufficiently represented by the item pool at hand. With respect to the two-dimensional factor solution, the factor loading pattern was highly similar across languages/samples (Supplemental Table S3). The rotated factor loadings clearly indicated that the intended constructs of Self-Certainty and Other-Certainty were reflected by the two factors. Model fit in the U.S. sample was $\chi^2(701) = 1,679.77$, comparative fit index (CFI) = .90, root mean square error of approximation (RMSEA) = .05, standardized root mean square residual (SRMR) = .04. Model fit in the German sample was $\chi^2(701) = 1,700.46$, CFI = .89, RMSEA = .05, SRMR = .04.

Deriving the Final CAMSQ Using Ant Colony Optimization (ACO). After having determined the dimensionality of the

CAMSQ items, we aimed to derive a psychometrically optimized item selection for assessing the two most-prominent factors identified as Self- and Other-Certainty. Specifically, our goal was to obtain equivalent English and German versions of the CAMSQ with (a) a manageable number of 10 items per scale, (b) high internal consistencies for Self- and Other-Certainty, (c) adequate fit to a simple structure, and (d) strong measurement invariance between the two languages. We relied on a meta-heuristic (i.e., ACO; Leite et al., 2008) to find a combination of items approaching these objectives because traditional item selection procedures (e.g., selecting items based on main loadings) are less suitable for such complex combinatorial problems. Conversely, meta-heuristics such as ACO can optimize several statistical criteria simultaneously and are capable of finding close-to-optimal solutions (Janssen et al., 2017; Olaru et al., 2019) without needing to evaluate every possible item combination (in this case, there are approximately 32.5 billion possible item combinations). Analogous to how ants communicate the most efficient path to sources of food via pheromones, the algorithm repeatedly draws combinations of items and, based on how well specific combinations performed in terms of predefined criteria, adjusts the probability of items to be drawn again in later iterations. When the algorithm does not find a new best model for a given number of iterations, the process converges on the best item combination.

We used the ACO algorithm as implemented in the *stuart* package (Schultze, 2020). Twenty items were selected with 10 for each scale. The optimization function was set to maximize the internal consistency of the two scales (high values of ω) and model fit (high values of CFI, low values of RMSEA) of a multigroup CFA with equality constraints for intercepts and loadings across the U.S. and the German samples (i.e., scalar measurement invariance). For further details about ACO settings including the exact optimization function, see Supplemental Note S3. The algorithm selected the best combination of items found in terms of the optimization function. This solution approached the targeted criteria in the multigroup CFA, CFI = .96, RMSEA = .04, $\omega_{\text{Self-Certainty}}$ = .89, $\omega_{\text{Other-Certainty}}$ = .88. Both metric (ΔCFI = −.000) and scalar measurement invariance (ΔCFI = −.006) between the two samples and languages were supported for the optimized solution with respect to the recommendations by Cheung and Rensvold (2002). This selection forms the final 20-item CAMSQ which is displayed in the Appendix. The structural model and parameter estimates based on the optimization process are shown in Supplemental Figure S2. We scored the mean for each CAMSQ scale to operationalize individual scores on Self- and Other-Certainty. In both samples, individuals were on average more certain about inferring their own mental states (U.S. sample [US]: $M_{\text{Self-Certainty}}$ = 5.34, German sample [GER]: $M_{\text{Self-Certainty}}$ = 5.36) than about inferring others' mental states (US: $M_{\text{Other-Certainty}}$ = 4.73, GER: $M_{\text{Other-Certainty}}$ = 4.65). The mean

difference amounted to M = −0.61/−0.71 (US/GER) and was significant with $p < .001$ in both samples. Self- and Other-Certainty scores were positively interrelated (US: r = .54; GER: r = .53).

Convergent and Discriminant Correlations. For the full correlation matrices of measures included in Study 1, see Supplemental Table S4 (U.S. sample) and Supplemental Table S5 (German sample). Across both samples, the correlational patterns of Self- and Other-Certainty indicated high convergent correlations and comparatively lower discriminant correlations (see Table 1). Specifically, the strongest correlates of Other-Certainty were the cognitive empathy scales of ACME, EQ, and BES ($.54 \leq r \leq .72$), whereas associations with EQ affective empathy, RSES self-esteem, RFQ-6 uncertainty about mental states (with five of six items pertaining to mental states of the self), and NARQ subclinical narcissism were less pronounced. Self-Certainty correlated most strongly with RFQ-6 (r = −.55) and RSES self-esteem (r = .54). Consistent with our expectations, certainty about mental states as reflected in the CAMSQ scales was virtually independent of *actual* performance in both the emotion recognition tasks (i.e., RMET and GEMOK-Blends) and the knowledge tasks (i.e., vocabulary test and BEFKI crystallized intelligence). In line with meta-analytic findings, the other self-report measures of social-cognitive abilities also explained little variance in the tasks. Importantly, higher certainty about mental states was linked to higher *perceived* performance in the emotion recognition tasks ($.07 \leq r \leq .33$) but not in the knowledge tasks. However, we did not find a robust pattern regarding the specificity of Self- or Other-Certainty as a correlate of perceived accuracy in the emotion recognition tasks.

Associations With Psychopathology. To investigate the associations between the CAMSQ scores and indicators of personality dysfunction and symptom distress, we first inspected linear associations. Whereas low Self-Certainty was consistently associated with poorer mental health (i.e., high scores on PID5BF+M and SCL-K9, low scores on SWLS), scores on Other-Certainty exhibited scattered but overall weak linear relationships with these indicators (Table 2). For example, Other-Certainty exhibited positive associations with antagonism in the U.S. sample and with life satisfaction in the German sample. Conversely, there were negative associations with detachment and disinhibition in both samples.

Second, we examined U-shaped relationships between the CAMSQ scales and psychopathology measures to test whether both low and high (but not medium) levels of certainty about mental states were maladaptive. The two-lines test (Simonsohn, 2018) estimates two regression lines for high and low values of a criterion to detect a sign change of the regression slope. Using this test, we found no evidence for this hypothesis. We provide smoothed scatterplots in

Table 1. Convergent and Discriminant Correlations of the CAMSQ in the U.S. Sample and the German Sample in Study 1.

Measure	Self-Certainty		Other-Certainty	
	US	GER	US	GER
ACME cognitive empathy	.48*	—	.72*	—
BES cognitive empathy	.47*	—	.54*	—
RFQ-6	−.55*	—	−.28*	—
RSES	.54*	—	.34*	—
IRI perspective-taking	.24*	—	.24*	—
EQ cognitive empathy	—	.33*	—	.69*
EQ affective empathy	—	.06	—	.14*
NARQ admiration	—	.22*	—	.23*
NARQ rivalry	—	.22*	—	.25*
Task performance				
RMET	.00	−.05	.01	.05
GEMOK-Blends	.02	−.02	.00	.03
Vocabulary test	−.05	—	−.09	—
BEFKI GC-K	—	.02	—	−.04
Perceived task performance				
RMET	.09	.21*	.33*	.23*
GEMOK-Blends	.07	.19*	.24*	.17*
Vocabulary test	.02	—	.06	—
BEFKI GC-K	—	.07	—	.02

Note. Results from the U.S. sample (US) are displayed on the left ($N = 263$) and results from the German sample (GER) are on the right ($N = 505$). CAMSQ = Certainty About Mental States Questionnaire; ACME = Affective and Cognitive Measure of Empathy; BES = Basic Empathy Scale; RFQ-6 = Reflective Functioning Questionnaire (6-item version); RSES = Rosenberg Self-Esteem Scale; IRI = Interpersonal Reactivity Index; EQ = Empathy Quotient; NARQ = Narcissistic Admiration and Rivalry Questionnaire; RMET = Reading the Mind in the Eyes Test; GEMOK-Blends = Geneva Emotion Knowledge Test—Blends Brief Form; BEFKI GC-K = Berlin Test for the Assessment of Fluid and Crystallized Intelligence—Short Form Crystallized Intelligence.
*$p < .05$.

the supplement that further illustrate the absence of U-shaped associations (Supplemental Figures S3–S6). In sum, the bivariate analyses indicated a maladaptivity of low Self-Certainty, but no clear picture emerged for Other-Certainty. It was evident, however, that the correlational patterns of Self- and Other-Certainty with respect to psychopathology were strongly divergent, which is noteworthy considering the substantial positive correlation between the CAMSQ scales.

Third, we computed semipartial correlations between each CAMSQ scale and indicators of psychopathology while statistically controlling for the effect of the other CAMSQ scale, respectively (Table 2). Whereas the associations between Self-Certainty and criteria remained negative and were virtually unchanged, Other-Certainty now exhibited consistently positive associations with personality dysfunction and symptom distress across most criteria. The finding that associations between the two CAMSQ scales with psychopathology have opposite signs when controlling for the respective other may hint at possible discrepancy effects. When Self- and Other-Certainty were entered as predictors in joint regression analyses, plotting the predicted values of maladaptive criteria for observed configurations of the CAMSQ scales suggested two effects. While (a)

lower Self-Certainty generally predicted personality dysfunction and symptom distress, (b) the more Other-Certainty approached or exceeded the level of Self-Certainty, the more personality dysfunction and symptom distress were reported. As an example, Figure 1 shows the predicted values of the PID5BF+M total score for combinations of Self- and Other-Certainty observed in the data. Higher personality dysfunction was predicted in the area below the diagonal line of congruence (LOC) where Other-Certainty exceeds Self-Certainty, and lower personality dysfunction was predicted above the LOC where Self-Certainty exceeds Other-Certainty. This pattern would indeed correspond to a discrepancy effect.[1]

Individual scores of such an Other-Self-Discrepancy can be operationalized as the algebraic difference of the scale scores of Other-Certainty minus Self-Certainty. Positive scores (i.e., exceeding Other-Certainty) indicate that an individual is more certain about others' mental states and negative scores indicate that they are more certain about their own mental states. Scores on Other-Self-Discrepancy exhibited consistent positive associations with psychopathology outcomes (Table 2). Although calculating difference scores is a straightforward approach to quantify a discrepancy between two variables at the individual level,

Table 2. Associations of the CAMSQ With Psychopathology in Study 1.

| | U.S. sample (N = 519) | | | | | German sample (N = 505) | | | | |
| | Bivariate (semipartial) correlations | | | Latent CRA | | Bivariate (semipartial) correlations | | | Latent CRA | |
Outcome	Self-Certainty	Other-Certainty	Other – Self	abs	p value	Self-Certainty	Other-Certainty	Other – Self	abs	p value
PID5BF+M total score	−.29* (−.33*)	−.01 (.17*)	.28*	0.67*	<.001	−.32* (−.33*)	−.09* (.10*)	.25*	0.48*	<.001
Negative affectivity	−.29* (−.30*)	−.08 (.10*)	.22*	0.42*	.004	−.33* (−.36*)	−.06 (.15*)	.29*	0.63*	<.001
Antagonism	−.13* (−.21*)	.10* (.20*)	.24*	0.65*	<.001	−.15* (−.17*)	.00 (.09*)	.15*	0.27*	.041
Psychoticism	−.17* (−.23*)	.05 (.16*)	.22*	0.54*	<.001	−.25* (−.29*)	.00 (.15*)	.26*	0.62*	<.001
Detachment	−.32* (−.32*)	−.10* (.08)	.22*	0.39*	.011	−.29* (−.25*)	−.15* (.00)	.15*	0.19	.159
Disinhibition	−.44* (−.44*)	−.14* (.12*)	.30*	0.63*	<.001	−.32* (−.28*)	−.15* (.03)	.18*	0.20	.136
Anankastia	.10* (.06)	.10* (.05)	.00	−0.14	.849	−.03 (−.02)	−.01 (.00)	.02	0.02	.459
SCL-K9	—	—	—	—	—	−.34* (−.38*)	−.04 (.18*)	.32*	0.69*	<.001
SWLS	—	—	—	—	—	.33* (.25*)	.21* (.04)	−.13*	−0.04	.594

Note. Estimates are standardized. Semipartial correlations controlling for Self-Certainty or Other-Certainty, respectively. The sign of all discrepancy effects is positive (*a3* parameter). CAMSQ = Certainty About Mental States Questionnaire; Other − Self = Other-Self-Discrepancy operationalized by the algebraic difference between Other-Certainty and Self-Certainty; CRA = condition-based regression analysis; *abs* = |Other − Self| − |Other+Self|; PID5BF+M = Personality Inventory for DSM-5 Brief Form + Modified; SCL-K9 = Symptom Checklist Short Scale; SWLS = Satisfaction with Life Scale.
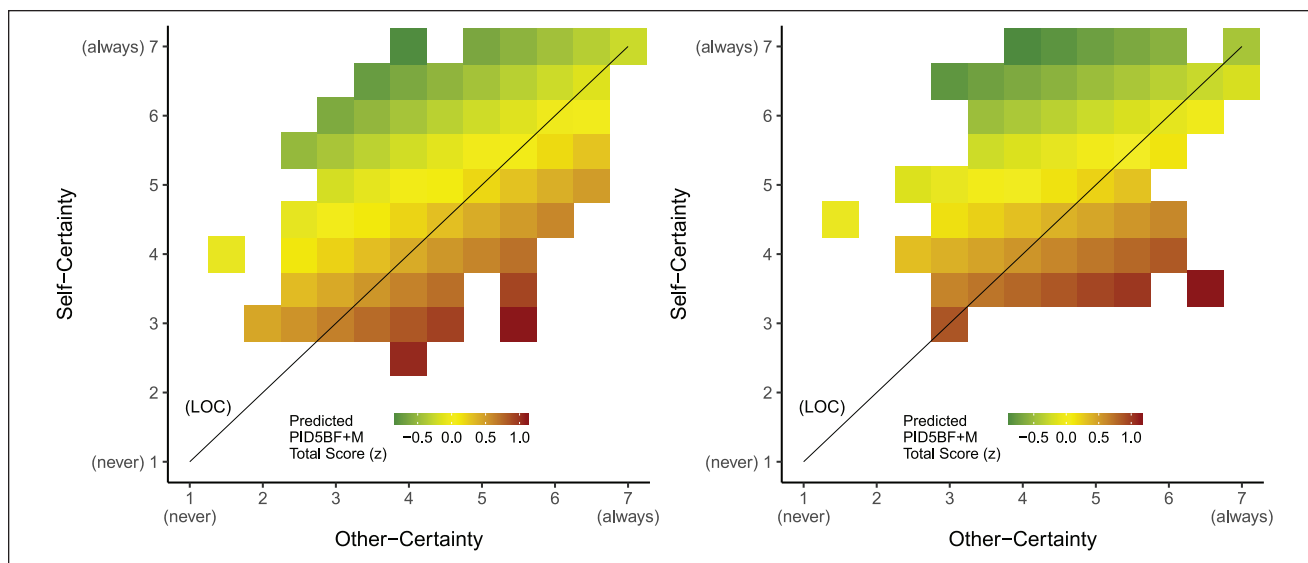*p < .05.

Figure 1. Predicted Values of PID5BF+M Total Score (z) for Observed Configurations of Self- and Other-Certainty in Study 1.
Note. Regression model estimates of the U.S. sample are displayed on the left ($N = 519$) and estimates of the German sample are on the right ($N = 505$). The LOC indicates equal levels of Self- and Other-Certainty. Below the LOC, Other-Certainty exceeds Self-Certainty. Above the LOC, Self-Certainty exceeds Other-Certainty. We only display the predicted values of PID5BF+M Total Score (z) for combinations of Self- and Other-Certainty that were observed in the data. PID5BF+M = Personality Inventory for DSM-5 Brief Form + Modified; LOC = line of congruence.

they are typically less reliable than the scores they were derived from and, more importantly, they are confounded with the latter. Condition-based regression analysis (CRA) has been specifically proposed as a means to address this issue and to test for discrepancy effects more rigorously (Humberg et al., 2018). In CRA, a criterion is regressed on two predictors for which discrepancy effects are to be examined, providing an omnibus test of whether both regression slopes exhibit significantly opposite effects. Discrepancy effects are tested for significance by defining the auxiliary parameter $abs = |b_{\text{Other-Certainty}} - b_{\text{Self-Certainty}}| - |b_{\text{Other-Certainty}} + b_{\text{Self-Certainty}}|$. A significantly positive *abs* parameter in conjunction with a positive (negative) difference between regression coefficients ($b_{\text{Other-Certainty}} - b_{\text{Self-Certainty}}$) indicates that higher Other-Self-Discrepancies are associated with higher (lower) scores on an outcome. A nonsignificant and/or negative *abs* parameter indicates the absence of a discrepancy effect. Performing CRA using latent variables indicated significant discrepancy effects across almost all criteria of psychopathology in both the U.S. and the German sample including the PID5BF+M general factor, its domains, and symptom distress as assessed by the SCL-K9 (Table 2). The exact same pattern of significant results was obtained when using manifest variables in CRA (see Supplemental Table S6).

Discussion

In Study 1, we derived the CAMSQ that aims to address potential shortcomings of previous attempts to assess mentalizing by self-report. The structure of the 20-item measure

indicated two empirically related but distinguishable factors of Self-Certainty and Other-Certainty. The CAMSQ scales were located in a nomological net that indicated plausible convergent and discriminant correlations. Specifically, the differential patterns of association indicated that Other-Certainty aligned with measures of inferring others' mental states (e.g., cognitive empathy, perceived performance in emotion recognition tasks), whereas Self-Certainty aligned with measures primarily pertaining to the self (e.g., mentalizing as measured by the RFQ-6, self-esteem). The rather strong association between Self-Certainty and self-esteem suggests that individuals reporting a good understanding of themselves also evaluate themselves more positively. The discriminant correlations of the CAMSQ with affective empathy, subclinical narcissism, and performance measures of emotion recognition and cognitive ability demonstrated that the CAMSQ scales cannot be subsumed under these constructs.

Different configurations of Self- and Other-Certainty exhibited pronounced associations with psychological dysfunction. In line with most studies using the RFQ, high levels of certainty about mental states were not indicative of psychopathology per se, but low levels of Self-Certainty were consistently maladaptive. Interestingly, the maladaptivity of Other-Certainty was dependent on holding the respective levels of Self-Certainty constant, thus giving rise to a discrepancy effect. More specifically, while the average individual was more certain about their own mental states as compared to others' mental states, a maladaptive profile of the CAMSQ was characterized by low Self-Certainty combined with similar or exceeding Other-Certainty. On the one

hand, it appears that low Self-Certainty corresponds to a maladaptive form of having too little certainty about mental states that may correspond to hypomentalizing. On the other hand, higher Other-Certainty was consistently maladaptive when adjusting for levels of Self-Certainty. Remarkably, this points to the possibility that hypermentalizing manifests itself in a form of imbalance between Self- and Other-Certainty corresponding to a maladaptive profile that comprises a level of Other-Certainty exceeding what would be expected given the level of Self-Certainty. Our results may thus suggest a refined conceptualization of hypermentalizing that hinges on a specific configuration of Self- and Other-Certainty.

Study 2

Building upon the initial findings from Study 1, we aimed to further validate the CAMSQ by replicating and extending the central results in a second study. In particular, we aimed to test whether our interpretations of hypomentalizing and hypermentalizing as reflected in configurations of the CAMSQ scales would find support in new and independent samples. For Study 2, we preregistered our analytic plan including a confirmatory test of the two-dimensional factor structure (H1: two correlated factors of Self- and Other-Certainty) and measurement invariance of the CAMSQ (H2: scalar measurement invariance between the English and the German version), the mean difference between Self- and Other-Certainty (H3: higher Self-Certainty than Other-Certainty on average), as well as their discrepancy effect in predicting general personality dysfunction (H4: greater discrepancies between Other- and Self-Certainty are associated with more severe impairment). The test-retest reliability of the CAMSQ scales was evaluated over a 2-week interval in the U.S. sample.

After solely relying on the PID5BF+M domains as an indicator of Criterion B of the AMPD and their general factor as an indicator of Criterion A in Study 1, we expanded the nomological net of the CAMSQ by including the Level of Personality Functioning Scale—Brief Form 2.0 (LPFS-BF) as a direct measure of Criterion A (Weekers et al., 2019). Moreover, we considered additional tests of convergent and discriminant validity using further self-report questionnaires of mentalizing. Similar to the CAMSQ, the MentS (Dimitrijević et al., 2018) comprises Self and Other scales of mentalizing and also includes a third scale measuring motivation to mentalize which should tap into a different aspect of mentalizing. The MZQ (Hausberg et al., 2012) mainly pertains to mentalizing the self and is thus similar to the RFQ. To move beyond investigating the associations between the CAMSQ and broader dimensions of personality pathology, in Study 2, we also examined further maladaptive traits that specifically relate

to socially aversive styles associated with interpersonal dysfunction. First, the dark core of personality (D16; Moshagen et al., 2018) is conceptualized as the dispositional tendency to "maximize one's individual utility—disregarding, accepting, or malevolently provoking disutility for others—accompanied by beliefs that serve as justifications" (Moshagen et al., 2018, p. 1) and is a common characteristic of PDs such as narcissistic, antisocial, paranoid, and borderline PD (Hilbig et al., 2021). Second, victim sensitivity refers to a dispositional pattern of perceiving oneself as a victim that is being treated unfairly by others (Schmitt et al., 2010). Victim sensitivity is associated with paranoia, jealousy, suspiciousness, and vengeance-seeking (Schmitt et al., 2005, 2010) and may contribute to interpersonal problems that are characteristic of borderline PD (Lis et al., 2018). Third, suspiciousness is defined by expectations and heightened sensitivity toward interpersonal ill-intent and is characteristic of paranoid and schizotypal PD (APA, 2013). Empirically, suspiciousness is an interstitial facet that does not clearly map onto a single PID-5 domain (Somma et al., 2019).

Method

Samples in Study 2. For the second study, we recruited participants online via panel providers from the United States (i.e., Prolific) and Germany (i.e., Bilendi) that differed from those used in the first study. Data quality was again ensured by attention and validity checks (for details on exclusion criteria, see Supplemental Note S1). Participants received minimum wage as compensation or comparable rewards. There were no missing data as individuals were not able to proceed without answering each item. To determine the required sample size, we performed a power analysis using a Monte Carlo simulation of the discrepancy effect found in the German sample of Study 1 (H4 of the preregistration) because it has the highest requirements regarding sample size for achieving sufficient power. The power for rejecting the null hypothesis was estimated at 87% for $N = 400$.

U.S. Sample 2. The sample was recruited via the panel provider Prolific. After data exclusions, the sample comprised 403 valid cases. Individuals were selected to approximate representativeness of the general U.S. population in terms of age, gender, and ethnicity. Participants (52% female) ranged in age from 18 to 89 years ($M = 45.2$, $SD = 16.6$). Participants had on average a higher level of education than the general population with 62% holding a bachelor's degree or higher and varied broadly with regard to occupational status (e.g., 42% employed for wages). Two weeks after the main study, 100 participants (41% female, $M_{age} = 53.2$, $SD_{age} = 15.0$, range = 18–79) completed a retest of the CAMSQ.[2]

German Sample 2. The sample was recruited via the panel provider Bilendi. The retained sample comprising 401 valid cases was approximately representative of the general German population in terms of age, gender, and region. Participants' (49% female) age ranged from 18 to 86 years (*M* = 45.9, *SD* = 16.6). Participants varied broadly with regard to the educational level (e.g., 30% with a bachelor's degree or higher) and occupational status (e.g., 52% employed for wages).

Measures. In Study 2, the CAMSQ was presented in its final 20-item form as derived in Study 1. On a separate page of the study, we also included the remaining items from the original item pool that were not selected for the CAMSQ in case the optimized solution performed poorly. The PID5BF+M, the SCL-K9, the RFQ, and the RSES were administered in the same form as in Study 1. Further measures that were used for the second study are described below. The same set of measures was administered in both the U.S. and the German sample. Measures were freely available in both languages or made available by the test creators. The study also included a range of additional variables (including political attitudes) that were measured for another project and will be reported elsewhere. As for Study 1, details on measures' internal consistencies and fit indices of their measurement models are provided in the supplement (Supplemental Table S7).

Mentalization Scale (MentS). The MentS (Dimitrijević et al., 2018) is a 28-item self-report measure assessing three dimensions of mentalizing (i.e., self, other, and motivation to mentalize). The items (e.g., "I find it important to understand reasons for my behavior") are rated on a 5-point scale ranging from *completely incorrect* (1) to *completely correct* (5).

Mentalization Questionnaire (MZQ). The MZQ (Hausberg et al., 2012) is a 15-item self-report measure assessing four aspects associated with mentalizing (i.e., emotional awareness, regulation of affect, psychic equivalence mode, and refusing self-reflection). The items (e.g., "Sometimes I only become aware of my feelings in retrospect") are rated on a 5-point scale ranging from *don't agree at all* (1) to *fully agree* (5).

Level of Personality Functioning Scale - Brief Form 2.0 (LPFS-BF). The LPFS-BF (Spitzer, Müller, et al., 2021; Weekers et al., 2019) is a 12-item self-report measure assessing impairments in the domains of self-functioning (6 items) and interpersonal functioning (6 items). The items (e.g., "I often make unrealistic demands on myself") are answered on a 4-point scale ranging from *completely untrue* (1) to *completely true* (4). High scores on the respective scales indicate self-dysfunction or interpersonal dysfunction.

Personality Inventory for DSM-5 Faceted Brief Form (PID-5-FBF). We used the four items included in the PID-5-FBF (Maples et al., 2015) to assess pathological suspiciousness according to DSM-5 (APA, 2013). Items (e.g., "Plenty of people are out to get me") are rated on a 4-point scale ranging from *very false or often false* (0) to *very true or often true* (3).

Dark Core of Personality (D16). The D16 (Moshagen et al., 2020) is a 16-item self-report measure assessing the dark core of personality. The items (e.g., "Payback needs to be quick and nasty") are answered on a 5-point scale ranging from *strong rejection* (1) to *strong endorsement* (5).

Justice Sensitivity Inventory (JSI). The JSI (Schmitt et al., 2010) is a 40-item self-report measure assessing four components of justice sensitivity (e.g., victim sensitivity, observer sensitivity, beneficiary sensitivity, and perpetrator sensitivity). This study used only the 10-item victim sensitivity scale. The items (e.g., "It makes me angry when others receive a reward that I have earned") are answered on a 6-point scale ranging from *not at all* (0) to *exactly* (5).

Statistical Analyses. The analyses were conducted in complete accordance with the preregistration using R version 4.0.3 (R Core Team, 2020), the package *lavaan* (Rosseel, 2012), and Mplus 8.4 (Muthén & Muthén, 1998–2019). We provide the preregistration, open data, and scripts for reproducing the analyses at https://osf.io/ea5nx/.

Results

CFA and Measurement Invariance. We hypothesized that the psychometrically optimized 20-item CAMSQ adheres to a simple structure of two correlated factors (H1). We thus estimated a two-dimensional CFA model in both samples, respectively. Model fit was evaluated using dynamic cutoffs (McNeish & Wolf, 2021) as implemented in the Dynamic Fit Index Shiny application (Wolf & McNeish, 2020). Dynamic cutoffs were proposed to provide more reliable cutoffs for model fit evaluation by taking into account various aspects of a factor solution (e.g., the magnitude of factor loadings, factor correlation) that are neglected when using rigid cutoffs (e.g., Hu & Bentler, 1999). However, dynamic cutoffs adhere to the same rigor in evaluating misfit as implemented by Hu and Bentler (i.e., corresponding to a 95% confidence interval to detect a misspecification equivalent to one unmodeled cross-loading). The calculation of dynamic cutoffs is based on the estimated model parameters; hence, they are calculated post hoc. Model estimates of the two separate CFAs per sample are displayed in Figure 2. For the U.S. sample, the dynamic cutoffs for good fit were CFI ≥ .903, RMSEA ≤ .079, and SRMR ≤ .072. The fit indices were in favor of the tested
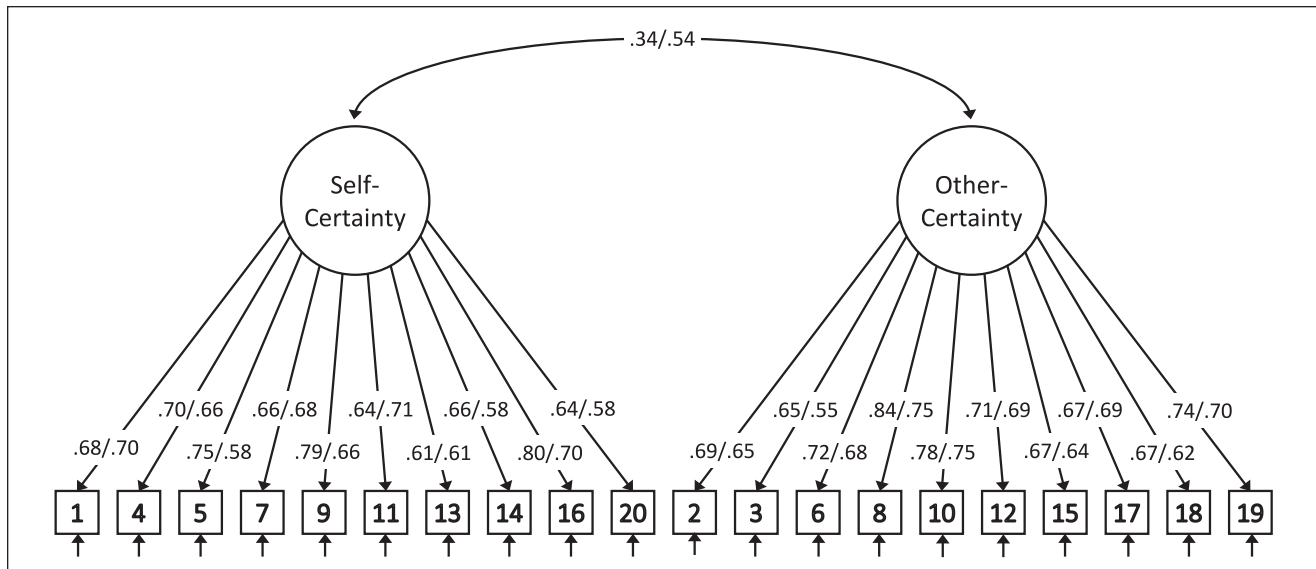
Figure 2. Confirmatory Factor Models of the CAMSQ in the U.S. Sample and the German Sample in Study 2.
Note. Estimates are from two separate CFAs. Standardized estimates are displayed. Item numbers of the final CAMSQ are displayed (see Appendix). Model estimates of the U.S. sample are displayed on the left ($N = 403$) and estimates of the German sample are on the right ($N = 401$). Residual variances are not displayed. Model fit in the U.S. sample was $\chi^2(169) = 371.21$, CFI = .93, RMSEA = .05, SRMR = .06. Model fit in the German sample was $\chi^2(169) = 330.64$, CFI = .94, RMSEA = .05, SRMR = .05. CFA = confirmatory factor analysis; CAMSQ = Certainty About Mental States Questionnaire; CFI = comparative fit index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual.

measurement model, CFI = .934, RMSEA = .054, SRMR = .055. For the German sample, the dynamic cutoff criteria were determined at CFI ≥ .934, RMSEA ≤ .059, and SRMR ≤ .052. The fit indices supported the model, CFI = .935, RMSEA = .049, and SRMR = .052.

In the next step, we investigated measurement invariance across the English and the German version of the CAMSQ using multigroup CFA following our preregistration plan (H2). As per our preregistration, we relied on a CFI difference of ≥−.010 between adjacent levels of measurement invariance as a cutoff criterion[3] (Cheung & Rensvold, 2002). We specified multigroup CFA models for testing configural, metric, and scalar measurement invariance by gradually constraining factor structure, factor loadings, and intercepts to equality between groups. Both metric (ΔCFI = −.002) and scalar measurement invariance (ΔCFI = −.010) between the two samples and languages were supported.

Empirical Distributions, Reliability, and Associations With Age and Gender. The empirical distributions of the CAMSQ scales observed for U.S. and German participants are displayed in Figure 3. The CAMSQ scales produced variation that covered large parts of the available scale; for example, percentiles of Other-Certainty for the U.S. sample were $P_{2.5} = 2.80$ ($\approx$ *sometimes* being the average response) and $P_{97.5} = 6.20$ ($\approx$ *almost always*). Responses of "never" or "almost never" were rather rarely observed. There were

no indications of floor or ceiling effects. In both samples, the mean difference between Self-Certainty and Other-Certainty was replicated (H3). As in Study 1, the average individual was more certain about inferring their own mental states as compared to inferring others' mental states, US: $t(402) = −13.98$, $p < .001$, GER: $t(400) = −17.61$, $p < .001$. Consistent with the findings from Study 1, the average Other-Self-Discrepancy amounted to a difference of $M = −0.73/−0.71$ (US/GER). There was broad variation in Other-Self-Discrepancy, for example, in the U.S. sample: $P_{2.5} = −2.80$ (higher Self-Certainty) and $P_{97.5} = 1.30$ (higher Other-Certainty).

Internal consistency was high for both Self-Certainty, $\omega = .90/.88$ (US/GER), and Other-Certainty, $\omega = .91/.89$. High test-retest correlations over a 2-week interval ($N = 100$) indicated the consistency of all three CAMSQ scores, namely, Self-Certainty ($r_{tt} = .85$), Other-Certainty ($r_{tt} = .78$), and Other-Self-Discrepancy ($r_{tt} = .82$). Participants' age was positively related to Self-Certainty ($r = .30/.23$ [US/GER]) and negatively related to Other-Certainty ($−13/−.17$) with younger individuals receiving lower scores on Self-Certainty and higher scores on Other-Certainty. Specifically, at the age of 20 years, the average difference between Other- and Self-Certainty was around 0, whereas it was around −1 for individuals at the age of 70 years (see Figure 4). Accordingly, a discrepancy effect was observed for age (*abs* = 0.52/0.76 [US/GER], both $p < .001$). Note that all mentalizing questionnaires as well
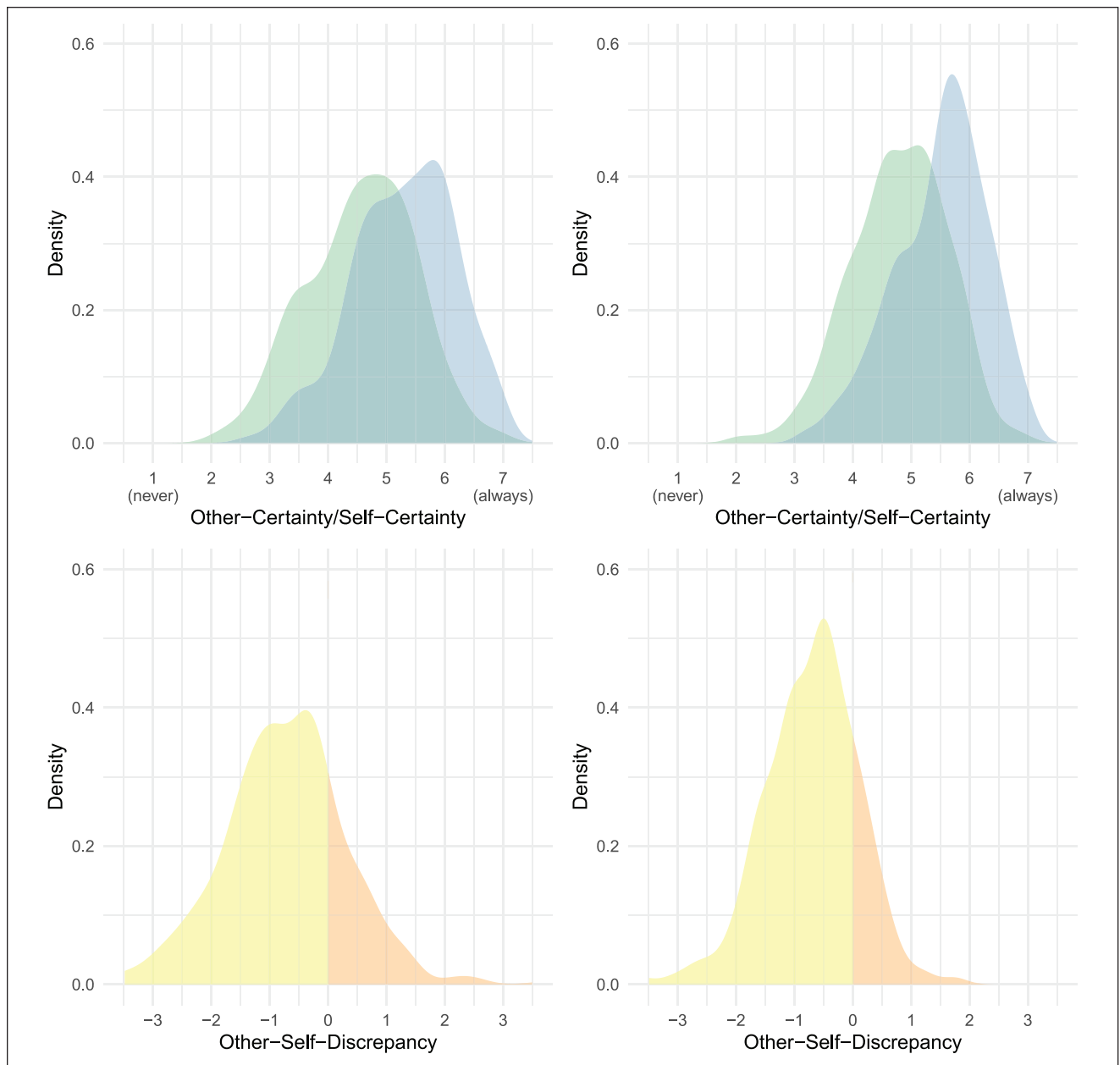
Figure 3. Density Distributions of CAMSQ Test Scores in the U.S. Sample and the German Sample in Study 2.
Note. Test score distributions from the U.S. sample are on the left (N = 403) and distributions from the German sample are on the right (N = 401). Self-Certainty is shown in blue and Other-Certainty is shown in green. $M_{Self\text{-}Certainty}$ = 4.57/4.80 (US/GER); $M_{Other\text{-}Certainty}$ = 5.30/5.51; $M_{Other\text{-}Self\text{-}Discrepancy}$ = −0.73/−0.71. Other-Self-Discrepancy is operationalized by the algebraic difference of Other-Certainty − Self-Certainty. Negative values of Other-Self-Discrepancy are highlighted in yellow and positive values are highlighted in orange. CAMSQ = Certainty About Mental States Questionnaire.

as the PID5BF+M exhibited substantial associations with age in the present data (see, for example, the association with PID5BF+M total score in Figure 4). Neither of the CAMSQ scales was significantly related to gender.

Convergent and Discriminant Correlations. For the full correlation matrices of measures included in Study 2, see Supplemental Table S8 (U.S. sample) and Supplemental Table

S9 (German sample). As the correlation coefficients were highly similar between U.S. and German participants, the following results apply to both samples. The correlational patterns suggested a convergence of CAMSQ scales with other mentalizing questionnaires (see Table 3). Specifically, Self-Certainty correlated most strongly with mentalizing measures that primarily pertain to the self (i.e., MentS Self, RFQ-6, MZQ), whereas Other-Certainty converged with
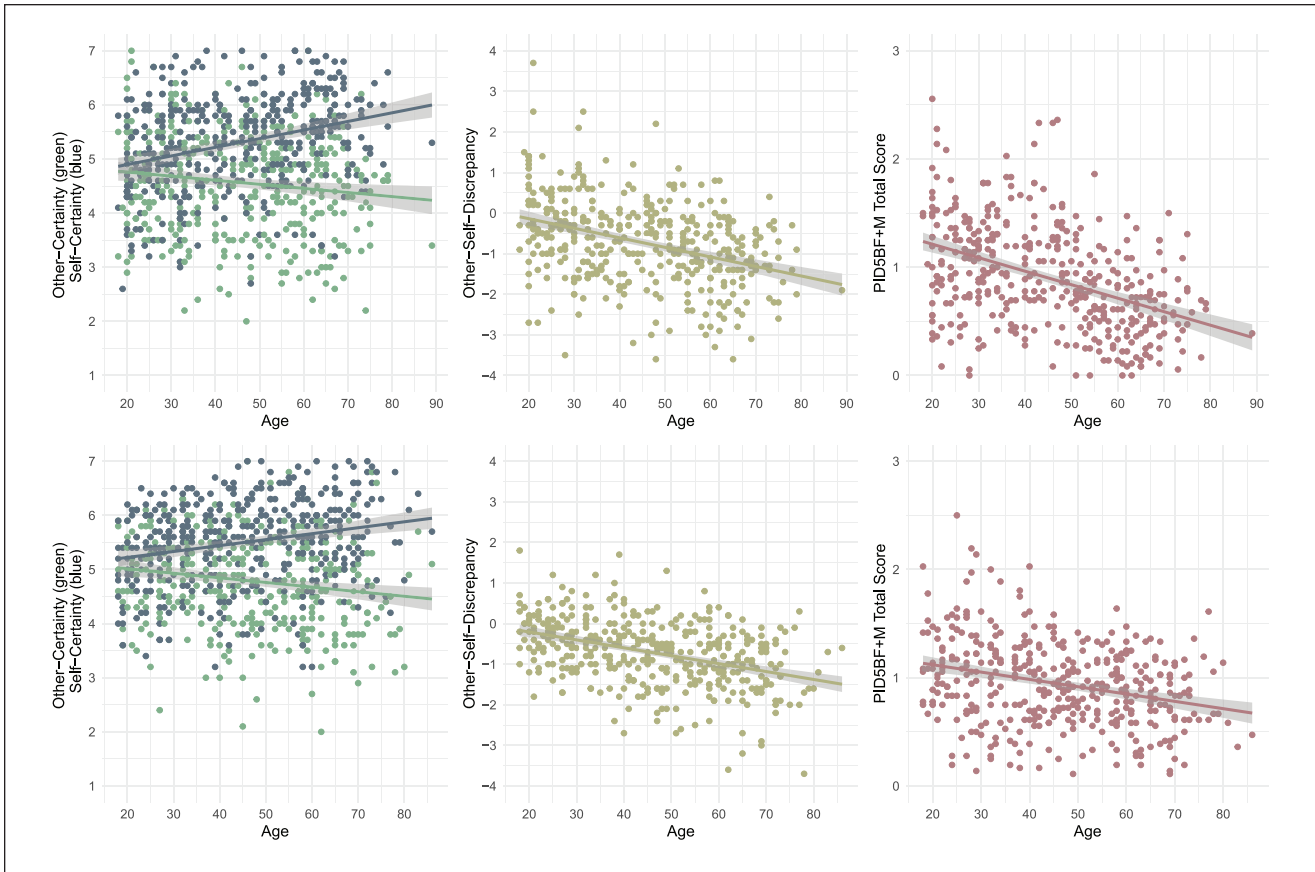
Figure 4. Linear Associations Between CAMSQ Scores and Age in Study 2 (With PID5BF+M Total Score for Comparison).
Note. Results from the U.S. sample are in the upper half (*N* = 403) and results from the German sample are in the lower half (*N* = 401).
CAMSQ = Certainty About Mental States Questionnaire; PID5BF+M = Personality Inventory for DSM-5 Brief Form + Modified.

Table 3. Convergent and Discriminant Correlations of the CAMSQ in the U.S. Sample and the German Sample in Study 2.

Measure	Self-Certainty		Other-Certainty	
	US	GER	US	GER
MentS				
Self	.58*	.44*	.05	.10
Other	.40*	.37*	.58*	.63*
Motivation to mentalize	.09	.15*	.25*	.31*
RFQ-6	−.55*	−.38*	−.07	−.01
MZQ total score	−.53*	−.42*	.00	−.13*
Emotional awareness	−.53*	−.44*	−.05	−.10*
Regulation of affect	−.49*	−.33*	.03	−.04
Psychic equivalence mode	−.43*	−.39*	.02	−.13*
Refusing self-reflection	−.32*	−.23*	.00	−.15*
RSES	.59*	.44*	.12*	.13*

Note. Results from the U.S. sample are displayed on the left (N = 403) and results from the German sample are on the right (N = 401).
CAMSQ = Certainty About Mental States Questionnaire; MentS = Mentalization Scale; RFQ-6 = Reflective Functioning Questionnaire (6-item version); MZQ = Mentalization Questionnaire; RSES = Rosenberg Self-Esteem Scale.
*p < .05.

MentS Other, which pertains to inferring others' mental states. Both CAMSQ scales exhibited comparatively lower correlations with MentS Motivation to Mentalize. As in Study 1, Self-Certainty was rather strongly correlated with RSES self-esteem.

Associations With Psychopathology. The patterns of bivariate associations observed in Study 2 (Table 4) were consistent with those observed in Study 1. The following results apply to both samples. Low Self-Certainty was indicative of maladaptive personality traits (i.e., PID5BF+M, LPFS-BF, D16 dark personality, JSI victim sensitivity) and symptom distress (i.e., SCL-K9). By contrast, Other-Certainty exhibited overall weak and divergent bivariate associations with psychopathology measures (e.g., small positive correlations with antagonism and psychoticism, small negative correlations with detachment and interpersonal dysfunction), indicating that neither low nor high scores on Other-Certainty were maladaptive per se. Replicating the maladaptivity of a profile characterized by Other-Certainty approaching or exceeding Self-Certainty, however, we found that the algebraic difference score was robustly related to personality pathology and symptom distress. As per our preregistered hypothesis, CRA indicated significant discrepancy effects not only for the general factor of the PID5BF+M (H4) but for virtually all predicted criteria, including the PID-5 domains, LPFS-BF, D16, JSI victim sensitivity, and SCL-K9 (Table 4). The same discrepancy effects were also found using manifest variables in CRA (Supplemental Table S10).

As a proof of concept, we explored whether discrepancy effects in the sense of Other-Self-Discrepancies could also be observed for the MentS because the Self and Other scales bear similarities to the respective CAMSQ scales. Indeed, the manifest difference scores of CAMSQ and MentS were correlated at $r = .51/.53$ (US/GER). Using the Other and Self scales of the MentS in manifest CRA[4] produced significant discrepancy effects (Supplemental Table S11). However, the discrepancy effects were markedly weaker for the MentS (median $abs = .18$; Supplemental Table S11) as compared to those found for the CAMSQ using manifest CRA (median $abs = .35$; see Supplemental Table S10).

Discussion

The data supported the two-dimensional structure of the 20-item CAMSQ in both languages and strong measurement invariance between the United States and Germany, thereby suggesting that the psychometric performance of the solution optimized in Study 1 generalizes to independent samples. Other-Certainty and Self-Certainty exhibited high internal consistencies, and test-retest correlations further indicated that these scores reflect trait-like characteristics. The correlations of the CAMSQ scales with existing mentalizing questionnaires provided additional evidence for its convergent and discriminant validity.

The unique associations with various psychopathology measures were replicated. Low Self-Certainty was once more consistently associated with virtually all included measures of personality pathology and symptom distress which would align with an interpretation of hypomentalizing. As in Study 1, high Other-Certainty was not maladaptive per se but only when controlling for Self-Certainty. The average profile of relatively higher Self-Certainty and lower Other-Certainty was replicated, corroborating that most individuals are more certain about their own mental states as compared to others' mental states. We proposed that Other-Self-Discrepancy may be another important dimension of interindividual variation reflecting a form of maladaptive certainty that could align with hypermentalizing. The high test-retest correlation of Other-Self-Discrepancy attested to its reliability as a test score and the stability of the construct. Most importantly, the discrepancy effects found in Study 1 were replicated across numerous criteria. Furthermore, we found similar (yet markedly weaker) discrepancy effects for the Self and Other scales of the MentS, suggesting the generalizability of imbalances in mentalizing one's own and others' mental states as an aspect of impaired mentalizing beyond the CAMSQ. As we suspected that high Other-Self-Discrepancies could be a common feature among socially aversive personality traits, we investigated discrepancy effects in the context of the dark core of personality, victim sensitivity, and suspiciousness. This notion was generally supported by the overall pattern of discrepancy effects which would be consistent with hypermentalizing affecting interpersonal relations.

General Discussion

Mentalizing the self and others are conceptualized as key aspects of psychological functioning (e.g., AMPD of DSM-5). For example, perceptual distortions are thought to contribute to recurring maladaptive interpersonal experiences that are observed in individuals with personality pathology (Hopwood, 2018). The psychoanalytic concept of genuine mentalizing (e.g., Bateman & Fonagy, 2019; Fonagy et al., 2016) as adaptive mentalizing posits a medium level of certainty when inferring mental states, reflecting a modest and unassuming stance about one's own mentalizing capacity (i.e., perceived mentalizing capacity). By contrast, deviations from this optimal level of certainty are termed hypomentalizing (i.e., having too little certainty) and hypermentalizing (i.e., having too much certainty). Existing mentalizing questionnaires appear to be incapable of assessing a maladaptive form of having too much certainty about mental states (e.g., Müller et al., 2021). This research hence aimed to develop a measure of perceived mentalizing capacity that could detect these two forms of

Table 4. Associations of the CAMSQ With Psychopathology in Study 2.

| | U.S. sample (N = 403) | | | | | German sample (N = 401) | | | | |
| | Bivariate (semipartial) correlations | | | Latent CRA | | Bivariate (semipartial) correlations | | | Latent CRA | |
Outcome	Self-Certainty	Other-Certainty	Other – Self	abs	p value	Self-Certainty	Other-Certainty	Other – Self	abs	p value
PID5BF+M total score	−.43* (−.49*)	.08 (.24*)	.44*	0.75*	<.001	−.34* (−.39*)	.00 (.20*)	.33*	0.78*	<.001
Negative affectivity	−.45* (−.49*)	.05 (.21*)	.42*	0.68*	<.001	−.30* (−.33*)	−.02 (.15*)	.27*	0.54*	.003
Antagonism	−.13* (−.20*)	.18* (.23*)	.26*	0.60*	<.001	−.14* (−.23*)	.11* (.21*)	.25*	0.65*	<.001
Psychoticism	−.28* (−.35*)	.15* (.25*)	.36*	0.72*	<.001	−.17* (−.28*)	.14* (.26*)	.31*	1.00*	<.001
Detachment	−.41* (−.40*)	−.08 (.06)	.27*	0.15	.148	−.40* (−.36*)	−.18* (.02)	.20*	0.23	.120
Disinhibition	−.45* (−.47*)	−.01 (.14*)	.36*	0.50*	.001	−.35* (−.37*)	−.06 (.13*)	.27*	0.58*	.003
Anankastia	−.14* (−.17*)	.09 (.14*)	.19	0.38*	.003	−.07 (−.11*)	.05 (.09)	.12*	0.34*	.028
PID-5-FBF suspiciousness	−.31* (−.35*)	.05 (.16*)	.30*	0.51*	<.001	−.24* (−.25*)	−.05 (.08)	.19*	0.27	.064
LPFS-BF total score	−.59* (−.60*)	−.06 (.14*)	.44*	0.57*	<.001	−.51* (−.54*)	−.09 (.20*)	.41*	0.94*	<.001
Self-functioning	−.62* (−.64*)	−.03 (.18*)	.49*	0.69*	<.001	−.52* (−.58*)	−.05 (.25*)	.46*	1.11*	<.001
Interpersonal functioning	−.46* (−.46*)	−.10 (.06)	.30*	0.28*	.049	−.40* (−.39*)	−.12* (.09)	.26*	0.39*	.023
D16	−.25* (−.30*)	.09 (.18*)	.29*	0.56*	<.001	−.10* (−.13*)	.02 (.09)	.12*	0.25*	.048
JSI victim sensitivity	−.17* (−.21*)	.09 (.15*)	.22*	0.40*	<.001	−.18* (−.22*)	.02 (.13*)	.20*	0.41*	.004
SCL-K9	−.46* (−.52*)	.10* (.27*)	.48*	0.83*	<.001	−.36* (−.41*)	−.01 (.19*)	.34*	0.70*	<.001

Note. Estimates are standardized. Semipartial correlations controlling for Self-Certainty or Other-Certainty, respectively. The sign of all discrepancy effects is positive (*a3* parameter).
CAMSQ = Certainty About Mental States Questionnaire; CRA = condition-based regression analysis; Other − Self = Other-Self-Discrepancy operationalized by the algebraic difference between Other-Certainty and Self-Certainty; *abs* = |Other − Self| − |Other + Self|; PID5BF+M = Personality Inventory for DSM-5 Brief Form + Modified; PID-5-FBF = Personality Inventory for DSM-5 Faceted Brief Form; LPFS-BF = Level of Personality Functioning Scale—Brief Form (high values reflect dysfunction); D16 = dark core of personality; JSI = Justice Sensitivity Inventory; SCL-K9 = Symptom Checklist K9.
*p < .05.

maladaptive deviations from genuine mentalizing. Starting with a pool of 40 items, we constructed a psychometrically sound 20-item self-report questionnaire in two languages. Tests of convergent and discriminant validity, internal consistency, test-retest reliability, and measurement invariance across the United States and Germany indicated that the newly derived CAMSQ performed equivalently well in both U.S. and German samples of which three out of four approximated the respective general populations. The findings presented herein indicate that the CAMSQ assesses two maladaptive variants of subjective certainty about mental states that can be linked to hypomentalizing and hypermentalizing. Whereas the average ("healthy") individual was more certain about their own mental states than about others' mental states, a maladaptive profile was characterized by low Self-Certainty in conjunction with Other-Certainty approaching or exceeding the level of Self-Certainty. The CAMSQ appears to be a promising new instrument to study personality functioning in general and mentalizing in particular because it assesses clinically relevant as well as adaptive configurations of perceiving and interpreting the self and others in terms of mental states.

Configurations of the CAMSQ Scales and Personality Functioning

This study is the first to present evidence suggesting that a balanced stance toward interpreting oneself and others (as evident in the configuration of Self- and Other-Certainty) might be a distinct competence that indicates mental health. We propose that being certain about oneself but not too certain about others could be an adapted take on the genuine mentalizing concept. In this sense, hypermentalizing as a presumed feature of impaired personality functioning could thus be reflected in elevated levels of Other-Certainty that do not follow the average healthy pattern of being more certain about oneself than about others (i.e., Other-Self-Discrepancy). Moreover, hypomentalizing could be reflected in low Self-Certainty of the CAMSQ as it exhibited consistent associations with a broad range of psychopathology markers. Research using the RFQ to measure uncertainty about mental states is in line with this notion (e.g., Fonagy et al., 2016). Hypomentalizing as a deviation from psychological functioning could relate to having an unstable self-image that manifests itself in uncertainty about one's own patterns of feeling, thinking, and behaving (i.e., identity). The strong intercorrelations between Self-Certainty, RSES self-esteem, and LPFS self-functioning as found in Study 2 illustrate the close link between a poorly understood self and a negative self-evaluation. However, in line with findings for the RFQ (e.g., Müller et al., 2021), the present studies did not provide any indication that individuals can be too certain about their own mental states in a maladaptive manner (i.e., hypermentalizing in regard to

the self). In the same vein, we did not find evidence that too little certainty about others' mental states is maladaptive (i.e., hypomentalizing in regard to others). It thus remains to be seen whether hypermentalizing the self and hypomentalizing others exist, whether they are less prevalent in the population, or whether the CAMSQ is incapable of capturing these phenomena.

Findings were also consistent with a developmental perspective on mentalizing. The adaptive profile of the CAMSQ that aligns with genuine mentalizing was increasingly prevalent in older participants because Self-Certainty tended to be higher and Other-Certainty tended to be lower compared to younger participants. In general, and in line with the idea of personality maturation, the CAMSQ Other-Self-Discrepancy as well as other indicators of personality pathology such as PID5BF+M indicated a pattern of more adaptive psychological functioning with increasing age.

The CAMSQ is the first self-report questionnaire of mentalizing that appears to capture two qualitatively different forms of mentalizing impairment as characterized by having too little or too much certainty about mental states. It is also the first measure providing a comprehensive and balanced coverage (see Supplemental Table S1) of different content domains of mental states (e.g., feelings, thoughts, attitudes, motivations, goals). As we have elaborated on before, existing questionnaires might conflate mentalizing content with assumed psychopathological consequences of impaired mentalizing (e.g., emotion dysregulation) by collapsing such indicators within the same scales, which could impede theory testing in some contexts. Except for the MentS, other questionnaires of mentalizing neglect the distinction between inferring the mental states of self and others. This research highlights that separating self- and other-mentalizing adds to a more comprehensive picture with respect to mentalizing impairment as the two constructs put each other into perspective, giving rise to adaptive and maladaptive configurations of Self-Certainty and Other-Certainty.

Limitations and Future Directions

This research is subject to some limitations. First and foremost, we deliberately chose to study a clinically relevant construct in the general population to delineate the construct across the full range of psychological functioning. Future studies should thus examine the utility of the CAMSQ in clinical populations and settings. It would be interesting to examine how self- and other-perception in terms of the CASMQ change during psychotherapy and whether such changes are related to symptom reduction. Moreover, this study is limited to showing that configurations of mentalizing certainty as measured by the CAMSQ are associated with personality dysfunction, but it still needs to be clarified whether such dispositions are a contributing factor to, a consequence of, or simply indicative of

maladjustment. For example, it could be insightful to investigate how mentalizing the self and others is linked to dynamic processes of person perception that may be involved in recurring maladaptive interpersonal sequences (Hopwood, 2018). In addition to the current focus on trait certainty, one might consider adapting the CAMSQ for this purpose so that it can measure the momentary state of certainty in intensive longitudinal studies, which would allow for a process-oriented understanding of mentalizing impairments. Another important limitation is that we solely relied on self-report measures to assess personality functioning. Thus, future research should examine how CAMSQ configurations relate to informant or clinician-rated personality dysfunction and mentalizing capacity. Additionally, in arguing that configurations of the CAMSQ may be indicative of the concepts of hypo- and hypermentalizing, we adhered to a narrower interpretation of these concepts that prioritizes the certainty aspect of individual differences in mentalizing. Theoretical accounts posit a variety of other features that may also be present in hypomentalizing or hypermentalizing individuals (e.g., behavioral dispositions to provide simplistic or lengthy narratives; Bateman & Fonagy, 2019). More generally, we have argued that certainty as assessed by the CAMSQ should not be equated with the actual accuracy in inferring mental states, which is also considered an important part of mentalizing (e.g., Luyten et al., 2020). Empirically, these aspects may be rather independent because mentalizing accuracy as for example assessed by performance tasks tends to exhibit minimal overlap with self-report measures of mentalizing. This notwithstanding, both mentalizing certainty and accuracy may contribute uniquely to healthy psychological functioning.

Conclusion

We have presented the CAMSQ as a new self-report measure of mentalizing that seeks to overcome the potential shortcomings of existing measures. Providing two equivalent and psychometrically sound versions of the CAMSQ in English and German, the measure is openly accessible and ready for use. Based on observations made with the CAMSQ, we propose a new and more nuanced conceptualization of genuine mentalizing as a configuration of being certain about oneself but not too certain about others.

Appendix

English Version of the Certainty About Mental States Questionnaire (CAMSQ)

People regularly interpret the feelings, thoughts, and behaviors of themselves or others. In the following, various statements will be presented that can be used to describe oneself.

For each statement, please evaluate which answer option best applies to you.

1. I know my innermost wishes and desires.
2. I can tell whether another person is at peace with themselves.
3. I know how other people will react to something.
4. I understand why certain things make me happy.
5. I know what I am trying to achieve with my behavior.
6. I can tell when a person in a group is feeling awkward.
7. I know why I am interested in certain things.
8. I can tell when other people don't give their honest opinions.
9. I understand my feelings.
10. I know when other people are hiding their thoughts.
11. I know what my virtues are.
12. I can tell when other people are just pretending to find something funny.
13. I know why I have a strong opinion on a subject.
14. When I'm in a bad mood, I know the reason why.
15. I know if a person is trustworthy.
16. I know the reasons for my behavior.
17. I can tell when other people are taking advantage of someone.
18. I can tell if another person is bored by what I am saying.
19. I know how a person feels when I look at their face.
20. I know what the best decision is for me in a difficult life situation.

Note. Items are scored by taking the mean of item responses (i.e., *never* = 1, *almost never* = 2, *sometimes* = 3, *half of the time* = 4, *often* = 5, *almost always* = 6, *always* = 7). Other-Certainty = 2, 3, 6, 8, 10, 12, 15, 17, 18, 19. Self-Certainty = 1, 4, 5, 7, 9, 11, 13, 14, 16, 20. Other-Self-Discrepancy = Other-Certainty – Self-Certainty. We recommend analyzing Other-Self-Discrepancy using condition-based regression analysis (Humberg et al., 2018) due to the algebraic difference being confounded with the scales of Self-Certainty and Other-Certainty. The English and the German versions of the CAMSQ for use in paper-pencil format are openly accessible at https://osf.io/ea5nx/. The CAMSQ is subject to a CC BY-NC-ND 4.0 license. Please note that publication or distribution of translations or other derivatives (e.g., short forms) require the authors' permission.

Authors' Note

Acknowledgments

Declaration of Conflicting Interests

Funding

ORCID iDs

Sascha Müller [iD] https://orcid.org/0000-0002-8663-4543

Leon P. Wendt [iD] https://orcid.org/0000-0003-2229-2860

Johannes Zimmermann [iD] https://orcid.org/0000-0001-6975-2356

Supplemental Material

Supplemental material for this article is available online.

Notes

1. We also explored interaction effects between Self- and Other-Certainty in predicting PID5BF+M total score and PID-5 domains in both samples but found no significant results.
2. All participants who took part in the main study were invited to the retest assessment but only 100 slots were available.
3. Although relying on dynamic cutoffs would be more consistent with our methodological approach, the functionality to derive dynamic cutoffs for measurement invariance testing has not yet been implemented.
4. We used the manifest variables for analyzing discrepancy effects of the MentS because there is no established confirmatory measurement model.

References

Allen, J. G., & Fonagy, P. (Eds.). (2006). *The handbook of mentalization-based treatment*. John Wiley.

Allen, J. G., Fonagy, P., & Bateman, A. (2008). *Mentalizing in clinical practice*. American Psychiatric Publishing.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Publishing.

Antonsen, B. T., Johansen, M. S., Rø, F. G., Kvarstein, E. H., & Wilberg, T. (2016). Is reflective functioning associated with clinical symptoms and long-term course in patients with personality disorders? *Comprehensive Psychiatry*, *64*, 46–58. https://doi.org/10.1016/j.comppsych.2015.05.016

Bach, B., Kerber, A., Aluja, A., Bastiaens, T., Keeley, J. W., Claes, L., Fossati, A., Gutierrez, F., Oliveira, S. E. S., Pires, R., Riegel, K. D., Rolland, J.-P., Roskam, I., Sellbom, M., Somma, A., Spanemberg, L., Strus, W., Thimm, J. C., Wright, A. G. C., & Zimmermann, J. (2020). International assessment of DSM-5 and ICD-11 personality disorder traits: Toward a common nosology in DSM-5.1. *Psychopathology*, *53*, 179–188. https://doi.org/10.1159/000507589

Back, M. D., Küfner, A. C., Dufner, M., Gerlach, T. M., Rauthmann, J. F., & Denissen, J. J. (2013). Narcissistic admiration and rivalry: Disentangling the bright and dark sides of narcissism. *Journal of Personality and Social Psychology*, *105*(6), 1013–1037. https://doi.org/10.1037/a0034431

Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: An investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders*, *34*(2), 163–175. https://doi.org/10.1023/B:JADD.0000022607.19833.00

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "reading the mind in the eyes" test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry*, *42*(2), 241–251. https://doi.org/10.1111/1469-7610.00715

Bateman, A. W., & Fonagy, P. (2016). *Mentalization-based treatment for personality disorders: A practical guide*. Oxford University Press.

Bateman, A. W., & Fonagy, P. (Eds.). (2019). *Handbook of mentalizing in mental health practice*. American Psychiatric Publishing.

Bender, D. S., Morey, L. C., & Skodol, A. E. (2011). Toward a model for assessing level of personality functioning in DSM-5, part I: A review of theory and methods. *Journal of Personality Assessment*, *93*, 332–346. https://doi.org/10.1080/00223891.2011.583808

Bleidorn, W. (2015). What accounts for personality maturation in early adulthood? *Current Directions in Psychological Science*, *24*(3), 245–252. https://doi.org/10.1177/0963721414568662

Bo, S., Sharp, C., Fonagy, P., & Kongerslev, M. (2017). Hypermentalizing, attachment, and epistemic trust in adolescent BPD: Clinical illustrations. *Personality Disorders: Theory, Research, and Treatment*, *8*(2), 172–182. https://doi.org/10.1037/per0000161

Bocklisch, F., Bocklisch, S. F., & Krems, J. F. (2012). Sometimes, often, and always: Exploring the vague meanings of frequency expressions. *Behavior Research Methods*, *44*(1), 144–157.

Bouchard, M. A., Target, M., Lecours, S., Fonagy, P., Tremblay, L. M., Schachter, A., & Stein, H. (2008). Mentalization in adult attachment narratives: Reflective functioning, mental states, and affect elaboration compared. *Psychoanalytic Psychology*, *25*(1), 47–66. https://doi.org/10.1037/0736-9735.25.1.47

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233–255.

Davies, M. (2008). *The Corpus of Contemporary American English (COCA)* [Word frequency data]. https://www.english-corpora.org/coca/

Davies, M. (2011). *Most frequent 100,000 word forms in English (based on data from the COCA corpus)*. Corpus of Contemporary American English. https://www.wordfrequency.info/

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*(1), 113–126. https://doi.org/10.1037/0022-3514.44.1.113

de Meulemeester, C., Lowyck, B., Vermote, R., Verhaest, Y., & Luyten, P. (2017). Mentalizing and interpersonal problems in

borderline personality disorder: The mediating role of identity diffusion. *Psychiatry Research*, *258*, 141–144. https://doi.org/10.1016/j.psychres.2017.09.061

Derogatis, L. R. (1977). *Symptoms Checklist-90: Administration, scoring, and procedures manual for the revised version*.

Diener, E. D., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, *49*(1), 71–75. https://doi.org/10.1207/s15327752jpa4901_13

Dimitrijević, A., Hanak, N., Altaras Dimitrijević, A., & Jolić Marjanović, Z. (2018). The Mentalization Scale (MentS): A self-report measure for the assessment of mentalizing capacity. *Journal of Personality Assessment*, *100*(3), 268–280. https://doi.org/10.1080/00223891.2017.1310730

Dziobek, I., Fleck, S., Kalbe, E., Rogers, K., Hassenstab, J., Brand, M., Kessler, J., Woike, J. K., Wolf, O. T., & Convit, A. (2006). Introducing MASC: A movie for the assessment of social cognition. *Journal of Autism and Developmental Disorders*, *36*(5), 623–636. 10.1007/s10803-006-0107-0

Fonagy, P. (1991). Thinking about thinking: Some clinical and theoretical considerations in the treatment of a borderline patient. *International Journal of Psychoanalysis*, *72*(4), 639–656.

Fonagy, P., Leigh, T., Steele, M., Steele, H., Kennedy, R., Mattoon, G., Target, M., & Gerber, A. (1996). The relation of attachment status, psychiatric classification, and response to psychotherapy. *Journal of Consulting and Clinical Psychology*, *64*(1), 22–31. https://doi.org/10.1037/0022-006X.64.1.22

Fonagy, P., & Luyten, P. (2016). A multilevel perspective on the development of borderline personality disorder. In D. Cicchetti (Ed.), *Developmental psychopathology: Maladaptation and psychopathology* (3rd ed., pp. 726–792). John Wiley. https://doi.org/10.1002/9781119125556.devpsy317

Fonagy, P., Luyten, P., Allison, E., & Campbell, C. (2017). What we have changed our minds about: Part 1. Borderline personality disorder as a limitation of resilience. *Borderline Personality Disorder and Emotion Dysregulation*, *4*, 1–11. https://doi.org/10.1186/s40479-017-0061-9

Fonagy, P., Luyten, P., Moulton-Perkins, A., Lee, Y.-W., Warren, F., Howard, S., Ghinai, R., Fearon, P., & Lowyck, B. (2016). Development and validation of a self-report measure of mentalizing: The Reflective Functioning Questionnaire. *PLOS ONE*, *11*(7), Article e0158678. https://doi.org/10.1371/journal.pone.0158678

Fonagy, P., & Target, M. (2006). The mentalization-focused approach to self pathology. *Journal of Personality Disorders*, *20*(6), 544–576. https://doi.org/10.1521/pedi.2006.20.6.544

Fonagy, P., Target, M., Steele, H., & Steele, M. (1998). *Reflective Functioning Scale manual* [Unpublished manuscript].

Fossati, A., Borroni, S., Dziobek, I., Fonagy, P., & Somma, A. (2018). Thinking about assessment: Further evidence of the validity of the Movie for the Assessment of Social Cognition as a measure of mentalistic abilities. *Psychoanalytic Psychology*, *35*(1), 127–141. https://doi.org/10.1037/pap0000130

Fossati, A., Somma, A., Krueger, R. F., Markon, K. E., & Borroni, S. (2017). On the relationships between DSM-5 dysfunctional personality traits and social cognition deficits: A study in a sample of consecutively admitted Italian psychotherapy patients. *Clinical Psychology & Psychotherapy*, *24*(6), 1421–1434. https://doi.org/10.1002/cpp.2091

Franke, H. (2000). *The Brief Symptom Inventory: Deutsche Version*. Manual. Beltz.

Fuchs, T. (2007). Fragmented selves: Temporality and identity in borderline personality disorder. *Psychopathology*, *40*(6), 379–387. https://doi.org/10.1159/000106468

George, C., Kaplan, N., & Main, M. (1985). *The adult attachment interview*. Department of Psychology, University of California, Berkeley.

Geukes, K., van Zalk, M., & Back, M. D. (2018). Understanding personality development: An integrative state process model. *International Journal of Behavioral Development*, *42*(1), 43–51. https://doi.org/10.1177/0165025416677847

Gomez, R., Watson, S., & Stavropoulos, V. (2020). Personality inventory for DSM-5, Brief Form: Factor structure, reliability, and coefficient of congruence. *Personality Disorders: Theory, Research, and Treatment*, *11*, 69–77. https://doi.org/10.1037/per0000364

Hausberg, M. C., Schulz, H., Piegler, T., Happach, C. G., Klöpper, M., Brütt, A. L., Sammet, I., & Andreas, S. (2012). Is a self-rated instrument appropriate to assess mentalization in patients with mental disorders? Development and first validation of the Mentalization Questionnaire (MZQ). *Psychotherapy Research*, *22*, 699–709. https://doi.org/10.1080/10503307.2012.709325

Hilbig, B. E., Thielmann, I., Klein, S. A., Moshagen, M., & Zettler, I. (2021). The dark core of personality and socially aversive psychopathology. *Journal of Personality*, *89*(2), 216–227. https://doi.org/10.1111/jopy.12577

Hopwood, C. J. (2018). Interpersonal dynamics in personality and personality disorders. *European Journal of Personality*, *32*(5), 499–524. https://doi.org/10.1002/per.2155

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

Humberg, S., Dufner, M., Schönbrodt, F. D., Geukes, K., Hutteman, R., Van Zalk, M. H., Denissen, J. J. A., Nestler, S., & Back, M. D. (2018). Enhanced versus simply positive: A new condition-based regression analysis to disentangle effects of self-enhancement from effects of positivity of self-view. *Journal of Personality and Social Psychology*, *114*(2), 303–322. https://doi.org/10.1037/pspp0000134

Ickes, W. (1993). Empathic accuracy. *Journal of Personality*, *61*, 587–610. https://doi.org/10.1111/j.1467-6494.1993.tb00783.x

Janke, S., & Glöckner-Rist, A. (2014). Deutsche Version der Satisfaction with Life Scale (SWLS) [German version of the Satisfaction with Life Scale (SWLS)]. *Zusammenstellung Sozialwissenschaftlicher Items und Skalen*. https://doi.org10.6102/zis147

Janssen, A. B., Schultze, M., & Grötsch, A. (2017). Following the ants. *European Journal of Psychological Assessment*, *33*, 409–421. https://doi.org/10.1027/1015-5759/a000299

Jolliffe, D., & Farrington, D. P. (2006). Development and validation of the Basic Empathy Scale. *Journal of Adolescence*, *29*(4), 589–611. https://doi.org/10.1016/j.adolescence.2005.08.010

Kretzschmar, A., & Gignac, G. E. (2019). At what sample size do latent variable correlations stabilize? *Journal of*

Research in Personality, *80*, 17–22. https://doi.org/10.1016/j.jrp.2019.03.007

Leite, W. L., Huang, I.-C., & Marcoulides, G. A. (2008). Item selection for the development of short forms of scales using an ant colony optimization algorithm. *Multivariate Behavioral Research*, *43*(3), 411–431. https://doi.org/10.1080/00273170802285743

Lis, S., Schaedler, A., Liebke, L., Hauschild, S., Thome, J., Schmahl, C., Stahlberg, D., Kleindienst, N., & Bohus, M. (2018). Borderline personality disorder features and sensitivity to injustice. *Journal of Personality Disorders*, *32*(2), 192–206. https://doi.org/10.1521/pedi_2017_31_292

Luyten, P., Campbell, C., Allison, E., & Fonagy, P. (2020). The mentalizing approach to psychopathology: State of the art and future directions. *Annual Review of Clinical Psychology*, *16*, 297–325. https://doi.org/10.1146/annurev-clinpsy-071919-015355

Luyten, P., Mayes, L. C., Nijssens, L., & Fonagy, P. (2017). The Parental Reflective Functioning Questionnaire: Development and preliminary validation. *PLOS ONE*, *12*(5), Article e0176218. https://doi.org/10.1371/journal.pone.0176218

Maples, J. L., Carter, N. T., Few, L. R., Crego, C., Gore, W. L., Samuel, D. B., Williamson, R. L., Lynam, D. R., Widiger, T. A., Markon, K. E., Krueger, R. F., & Miller, J. D. (2015). Testing whether the DSM-5 personality disorder trait model can be measured with a reduced set of items: An item response theory investigation of the Personality Inventory for DSM-5. *Psychological Assessment*, *27*(4), 1195–1210.

McLaren, R., Hopwood, C., Gallagher, M., & Sharp, C. (2021). *The relative strength of the relationship between hypermentalizing and borderline personality disorder: A meta-analytic review* [Manuscript submitted for publication].

McNeish, D., & Wolf, M. G. (2021). Dynamic fit index cutoffs for confirmatory factor analysis models. *Psychological Methods*. Advance online publication. https://psycnet.apa.org/doiLanding?doi=10.1037%2Fmet0000425

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

Moshagen, M., Zettler, I., & Hilbig, B. E. (2020). Measuring the dark core of personality. *Psychological Assessment*, *32*(2), 182–196. https://doi.org/10.1037/pas0000778

Müller, S., Wendt, L. P., Spitzer, C., Masuhr, O., Back, S. N., & Zimmermann, J. (2021). A critical evaluation of the Reflective Functioning Questionnaire (RFQ). *Journal of Personality Assessment*. Advance online publication. https://doi.org/10.1080/00223891.2021.1981346

Murphy, B. A., & Lilienfeld, S. O. (2019). Are self-report cognitive empathy ratings valid proxies for cognitive empathy ability? Negligible meta-analytic relations with behavioral task performance. *Psychological Assessment*, *31*, 1062–1072. https://doi.org/10.1037/pas0000732

Muthén, L. K., & Muthén, B. O. (1998–2019). *Mplus user's guide* (8th ed.).

Olaru, G., Schroeders, U., Hartung, J., Wilhelm, O., & Wrzus, C. (2019). Ant colony optimization and local weighted structural equation modeling. A tutorial on novel item and person sampling procedures for personality research. *European Journal of Personality*, *33*(3), 400–419. https://doi.org/10.1002/per.2195

Olderbak, S., & Wilhelm, O. (2020). Overarching principles for the organization of socioemotional constructs. *Current Directions in Psychological Science*, *29*(1), 63–70. https://doi.org/10.1177/0963721419884317

Olderbak, S., Wilhelm, O., Olaru, G., Geiger, M., Brenneman, M. W., & Roberts, R. D. (2015). A psychometric analysis of the reading the mind in the eyes test: Toward a brief form for research and applied settings. *Frontiers in Psychology*, *6*, Article 1503. https://doi.org/10.3389/fpsyg.2015.01503

Open-Source Psychometrics Project. (2020, October 1). *Vocabulary IQ test*. https://openpsychometrics.org/tests/VIQT/

Orth, U., Trzesniewski, K. H., & Robins, R. W. (2010). Self-esteem development from young adulthood to old age: A cohort-sequential longitudinal study. *Journal of Personality and Social Psychology*, *98*(4), 645–658. https://doi.org/10.1037/a0018769

Petrowski, K., Schmalbach, B., Kliem, S., Hinz, A., & Brähler, E. (2019). Symptom-Checklist-K-9: Norm values and factorial structure in a representative German sample. *PLOS ONE*, *14*(4), Article e0213490. https://doi.org/10.1371/journal.pone.0213490

Pincus, A. L., Hopwood, C. J., & Wright, A. G. C. (2020). The interpersonal situation: An integrative framework for the study of personality, psychopathology, and psychotherapy. In D. Funder, J. F. Rauthmann, & R. Sherman (Eds.), *Oxford handbook of psychological situations* (pp. 124–142). Oxford University Press.

Poznyak, E., Morosan, L., Perroud, N., Speranza, M., Badoud, D., & Debbané, M. (2019). Roles of age, gender and psychological difficulties in adolescent mentalizing. *Journal of Adolescence*, *74*, 120–129. https://doi.org/10.1016/j.adolescence.2019.06.007

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Realo, A., Allik, J., Nolvak, A., Valk, R., Ruus, T., Schmidt, M., & Eilola, T. (2003). Mind-reading ability: Beliefs and performance. *Journal of Research in Personality*, *37*(5), 420–445. https://doi.org/10.1016/S0092-6566(03)00021-7

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton University Press.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*, 1–36. http://www.jstatsoft.org/v48/i02/

Schipolowski, S., Wilhelm, O., Schroeders, U., Kovaleva, A., Kemper, C. J., & Rammstedt, B. (2013). BEFKI GC-K: A short scale for the measurement of crystallized intelligence. *Methods, Data, Analyses*, *7*(2), 153–181. https://doi.org/10.12758/mda.2013.010

Schlegel, K., & Scherer, K. R. (2018). The nomological network of emotion knowledge and emotion understanding in adults: Evidence from two new performance-based tests. *Cognition and Emotion*, *32*(8), 1514–1530. https://doi.org/10.1080/02699931.2017.1414687

Schmitt, M., Baumert, A., Gollwitzer, M., & Maes, J. (2010). The Justice Sensitivity Inventory: Factorial validity, location in the personality facet space, demographic pattern, and normative data. *Social Justice Research*, *23*(2), 211–238. https://doi.org/10.1007/s11211-010-0115-2

Schmitt, M., Gollwitzer, M., Maes, J., & Arbach, D. (2005). Justice sensitivity. *European Journal of Psychological Assessment*, *21*(3), 202–211. https://doi.org/10.1027/10155759.21.3.202

Schultze, M. (2020). *stuart: Subtests using algorithmic rummaging techniques* (R Package version 0.9.1). https://CRAN.R-project.org/package=stuart

Sharp, C., Pane, H., Ha, C., Venta, A., Patel, A. B., Sturek, J., & Fonagy, P. (2011). Theory of mind and emotion regulation difficulties in adolescents with borderline traits. *Journal of the American Academy of Child & Adolescent Psychiatry*, *50*(6), 563–573. https://doi.org/10.1016/j.jaac.2011.01.017

Sharp, C., Steinberg, L., McLaren, V., Weir, S., Ha, C., & Fonagy, P. (2021). Refinement of the Reflective Function Questionnaire for Youth (RFQY) Scale B using item response theory. *Assessment*. Advance online publication. https://doi.org/10.1177/10731911211003971

Sharp, C., & Vanwoerden, S. (2015). Hypermentalizing in borderline personality disorder: A model and data. *Journal of Infant, Child, and Adolescent Psychotherapy*, *14*(1), 33–45. https://doi.org/10.1080/15289168.2015.1004890

Simonsohn, U. (2018). Two lines: A valid alternative to the invalid testing of U-shaped relationships with quadratic regressions. *Advances in Methods and Practices in Psychological Science*, *1*, 538–555. https://doi.org/10.1177/2515245918805755

Skodol, A. E., Clark, L. A., Bender, D. S., Krueger, R. F., Morey, L. C., Verheul, R., Alarcon, R. D., Bell, C. C., Siever, L. J., & Oldham, J. M. (2011). Proposed changes in personality and personality disorder assessment and diagnosis for DSM-5 Part I: Description and rationale. *Personality Disorders: Theory, Research, and Treatment*, *2*(1), 4–22. https://doi.org/10.1037/a0021891

Slade, A., Aber, J. L., Bresgi, I., Berger, B., & Kaplan, C. A. (2004). *The parent development interview—Revised*. The City University of New York.

Somma, A., Krueger, R. F., Markon, K. E., & Fossati, A. (2019). The replicability of the personality inventory for DSM–5 domain scale factor structure in U.S. and non-U.S. samples: A quantitative review of the published literature. *Psychological Assessment*, *31*(7), 861–877. https://doi.org/10.1037/pas0000711

Spitzer, C., Müller, S., Kerber, A., Hutsebaut, J., Brähler, E., & Zimmermann, J. (2021). Die deutsche Version der Level of Personality Functioning Scale-Brief Form 2.0 (LPFS-BF): Faktorenstruktur, konvergente Validität und Normwerte in der Allgemeinbevölkerung [The German version of the Level of Personality Functioning Scale-Brief Form 2.0 (LPFS-BF): Latent structure, convergent validity and norm values in

the general population]. *Psychotherapie–Psychosomatik–Medizinische Psychologie*, *71*(7), 284–293. https://doi.org/10.1055/a-1343-2396

Spitzer, C., Zimmermann, J., Brähler, E., Euler, S., Wendt, L. P., & Müller, S. (2021). Die deutsche Version des Reflective Functioning Questionnaire (RFQ): Eine teststatistische Überprüfung in der Allgemeinbevölkerung [The German version of the Reflective Functioning Questionnaire (RFQ): A psychometric evaluation in the German general population]. *Psychotherapie–Psychosomatik–Medizinische Psychologie*, *71*(3/4), 124–131. https://doi.org/10.1055/a-1234-6317

Taubner, S., Hörz, S., Fischer-Kern, M., Doering, S., Buchheim, A., & Zimmermann, J. (2013). Internal structure of the Reflective Functioning Scale. *Psychological Assessment*, *25*(1), 127–135. https://doi.org/10.1037/a0029138

Thornton, M. A., Weaverdyck, M. E., & Tamir, D. I. (2019). The brain represents people as the mental states they habitually experience. *Nature Communications*, *10*(1), Article 2291. https://doi.org/10.1038/s41467-019-10309-7

Vachon, D. D., & Lynam, D. R. (2016). Fixing the problem with empathy: Development and validation of the affective and cognitive measure of empathy. *Assessment*, *23*(2), 135–149. https://doi.org/10.1177/1073191114567941

Vazire, S. (2010). Who knows what about a person? The self-other knowledge asymmetry (SOKA) model. *Journal of Personality and Social Psychology*, *98*(2), 281–300. https://doi.org/10.1037/a0017908

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H. L. J., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science*, *7*(6), 632–638. https://doi.org/10.1177/1745691612463078

Weekers, L. C., Hutsebaut, J., & Kamphuis, J. H. (2019). The Level of Personality Functioning Scale-Brief Form 2.0: Update of a brief instrument for assessing level of personality functioning. *Personality and Mental Health*, *13*(1), 3–14. https://doi.org/10.1002/pmh.1434

Wolf, M. G., & McNeish, D. (2020). *Dynamic model fit* (R Shiny Application Version 1.1.0). https://dynamicfit.app/connect/

World Health Organization. (2018). *International classification of diseases for mortality and morbidity statistics* (11th rev.). https://icd.who.int/browse11/l-m/en

Zimmermann, J., Müller, S., Bach, B., Hutsebaut, J., Hummelen, B., & Fischer, F. (2020). A common metric for self-reported severity of personality disorder. *Psychopathology*, *53*, 161–171. https://doi.org/10.1159/000507377