

German Translation of the Four-Item Mentalising Index (FIMI-G)

Alex Bertrams¹ , Max Blaise¹ , Ann Krispenz¹ 

[1] *Institute of Educational Science, University of Bern, Bern, Switzerland.*

Measurement Instruments for the Social Sciences, 2024, Vol. 6, Article e12249, <https://doi.org/10.5964/miss.12249>

Received: 2023-01-24 • **Accepted:** 2023-11-01 • **Published (VoR):** 2024-02-08

Handling Editor: Matthias Bluemke, GESIS – Leibniz Institute for the Social Sciences, Mannheim, Germany

Corresponding Author: Alex Bertrams, University of Bern, Division of Educational Psychology, Institute of Educational Science, Fabrikstrasse 8, 3012 Bern, Switzerland. E-mail: alexander.bertrams@unibe.ch

Supplementary Materials: Data, Materials [see [Index of Supplementary Materials](#)]



Abstract

Mentalising can be defined as the social-cognitive ability to understand and infer the mental non-emotional states of oneself and others. Recently, the Four-Item Mentalising Index (FIMI), a self-report scale, was developed to efficiently measure mentalising ability in English-speaking samples. This study presents a German translation of the FIMI—namely, the Four-Item Mentalising Index-German (FIMI-G). To assess the usefulness of the translation, initial evidence for the psychometric properties of the FIMI-G was gathered in a German-speaking sample from Austria, Germany, and Switzerland (N = 283). As expected, the corrected item-total correlations, the confirmatory factor analysis, and the inner consistency estimation indicated a homogenous, unidimensional measure which corresponds to the English original. In addition, the FIMI-G scores were related to the validation criteria as expected. Socially desirable responding did not undermine the validity. It is concluded that the German FIMI translation is a useful measure.

Keywords

measurement, mentalising, psychometrics, questionnaire, social-cognitive ability

Clutterbuck et al. (2021, p. 629) define *mentalising* as the “social-cognitive ability to understand and infer the mental states of oneself and others, including beliefs, intentions, and desires”. In the nomological landscape of social cognition, mentalising needs to be distinguished from neighbouring constructs such as cognitive empathy and reflective functioning. According to Happé et al. (2017), *empathy* is the ability to adopt the affective state of others. Such adoption may include (merely) *comprehending* the affective states



of others (so called cognitive empathy) and/or the vicarious *experience* of the affective state of those others (so called affective empathy). Accordingly, mentalising and *cognitive* empathy are similar constructs—both require the ability to *take the perspective* of another individual. However, both constructs focus on different mental states of others: While cognitive empathy encompasses the inference of others' *emotional* states, mentalising captures inferences about others' *non-emotional* states (Clutterbuck et al., 2021; Reniers et al., 2011). The necessity for a theoretical distinction between (cognitive) empathy and mentalising is also supported by studies which show that both constructs are based on different neurobiological mechanisms (Preckel et al., 2018).

Other researchers have equated mentalizing with *reflective functioning* (Fonagy et al., 2016).¹ Those authors define mentalizing as “the capacity to reflect on internal mental states such as feelings, wishes, goals, and attitudes, with regard to both the self and others” (p. 2). Even though this definition by Fonagy et al. (2016) is very close to the definition of mentalising by Clutterbuck et al. (2021), it additionally includes the reflection of “feelings” (i.e., affective mental states). Therefore, reflective functioning also seems to include aspects of cognitive empathy and is hence different from mentalising *sensu* Clutterbuck et al. (2021) which focuses on the ability to make inferences about others' *non-emotional* states.

Moreover, the definition of Fonagy et al. (2016) refers to the *capacity to reflect* internal mental states *accurately*. Accordingly, their measure for reflective functioning (i.e., the Reflective Functioning Questionnaire; RFQ) was constructed to measure *genuine* mentalizing vs. *hypomentalizing* (i.e., the inability to consider complex models of one's mind or the mind of others) vs. *hypermentalizing* (i.e., excessive mentalizing). As a result, the RFQ allows to investigate if different impairments in individuals' capacity to reflect on internal mental states (i.e., *hypomentalizing* and *hypermentalizing*) are differentially related to certain psychopathologies (such as borderline personality disorder and eating disorders). In contrast, the present research is based on a definition of mentalising that focuses on the *ability to understand* internal non-emotional states of oneself and others, therefore allowing to measure impairments or the lack of this ability (for example in individuals with autism).

Four-Item Mentalising Index

Existing measures of social cognitions typically focus on the assessment of empathy. For example, the widely used Interpersonal Reactivity Index (IRI; Davis, 1980; for a German version of the IRI, see Grevenstein, 2020) assesses empathy using four subscales:

1) Note that the term “mentalising” is spelled differently depending on the theoretical framework it is used in. In the present research, we rely on the definition of Clutterbuck et al. (2021) who equate mentalising with the “theory of mind” (p. 629) and spell the term using the letter “s”. However, other theoretical frameworks spell the term using the letter “z” (e.g., Fonagy et al., 2016) which in the present context also implicates a different meaning of the word.

(1) fantasy (i.e., the tendency to identify with fictional characters), (2) empathic concern (i.e., feelings of compassion and concern for the needs of others), (3) personal distress (i.e., the tendency to experience discomfort and stress when in the presence of distressed others), and (4) perspective taking (i.e., the tendency to adopt other individuals' point of view). One could argue that the IRI subscale "perspective taking" measures mentalising as defined by Clutterbuck et al. (2021). However, one needs to remember that those IRI items which measure perspective taking are answered by participants in very close proximity to the other IRI items which contain emotion-related content and wording. Therefore, in the context of the IRI, the perspective taking items may likely be perceived as a measure of (cognitive) empathy rather than mentalising as defined by Clutterbuck et al. (2021). Even more importantly, the IRI perspective taking items seem to measure individuals' *intention* to adopt another's viewpoint (e.g., "I try to look at everybody's side of a disagreement before I make a decision") rather than individuals' *ability* to do so. Therefore, the IRI items focusing on perspective taking cannot be used to measure individuals' mentalising ability or the lack thereof (e.g., in individuals with autism).

As no efficient self-report measure was available to measure individual differences in mentalising ability (i.e., the ability to understand the non-emotional states of oneself and others) in the general population, Clutterbuck et al. (2021) recently developed the Four-Item Mentalising Index (FIMI) by drawing items from different existing measures of social cognition. To ensure that these items reflected their conceptual definition of mentalising, the authors excluded any items that were rated by an expert panel to have emotion-related wording. This resulted in an initial selection of nine items. Based on the results of exploratory factor analysis (EFA) and confirmatory factor analyses (CFA) in a sample of $N = 660$ participants (Study 1a), four of the nine initial items were selected as suitable indicators for the mentalising scale. These results were replicated in another sample of $N = 669$ participants from the US and the UK using CFA (Study 1b). Before testing the selected items further, the wording of two of the four items was changed to ensure gender-neutral language and to further reduce emotional content. The resulting four items form the final English version of the FIMI.

Subsequently, the factor structure and reliability of the FIMI was examined (Study 2a: $N = 1,999$; Study 2b: $N = 116$). In those studies, Clutterbuck et al. (2021) report good psychometric properties in terms of the reliability (Study 2a: $\omega = .75$) and test-retest reliability (Study 2b: $r = .74$, $p < .001$). Results of another CFA showed an excellent fit for a 1-factor solution for the four FIMI items. Additionally, the authors conducted invariance analyses showing that the FIMI is invariant to sex.

In a last step, Clutterbuck et al. (2021) investigated the validity of the new measure by testing its construct validity against a cognitive mentalising task, autistic traits, and comparing scores in autistic and non-autistic people. In line with their expectations, the results of Study 3a ($N = 500$) revealed that the FIMI was positively related to the performance in the cognitive mentalising task ($r = .35$, $p < .001$) and negatively associated

with autistic traits ($r = -.43, p < .001$). In Study 3b ($N = 285$), the authors found statistically significantly lower FIMI scores in autistic adults ($N = 102$) than in non-autistic adults ($N = 183$). In concert, the results of the six studies conducted by Clutterbuck et al. (2021) show that the FIMI is a conceptually and methodologically robust scale for the assessment of mentalising.

Present Research

The aim of the present study was to provide initial evidence for the reliability and validity of a German translation of the FIMI, the Four-Item Mentalising Index-German (FIMI-G). It was expected that the unidimensional factor structure of the original FIMI would fit the German version as well, as the FIMI-G measures the same construct as the original English version. Furthermore, responses to the FIMI-G were assumed to be strongly positively related to habitually considering others' viewpoints (i.e., *perspective taking*; Grevenstein, 2020), as such a habit may most likely be built up in individuals who are usually successful in inferring other individuals' inner states. The FIMI-G responses were also predicted to be positively related to individual differences in *empathic concern*—that is, how frequently one is typically feeling compassion and concern for the needs of others (Grevenstein, 2020). While mentalising may be helpful in recognising the cognitive preconditions of others' emotions (e.g., thwarted intentions) and may therefore often facilitate compassion, understanding others' inner lives does not necessarily elicit intense empathic responses in the perceiver (Majdandžić et al., 2016). Thus, only a moderate relationship between mentalising and empathic concern was assumed. As a weak mentalising ability may complicate social communication and interaction (Frith & Frith, 2001; Parelman et al., 2021), it was further predicted that lower FIMI-G scores would be related to lower *communication and reciprocity* skills, which are an aspect of autistic traits. In line with this assumption, Clutterbuck et al. (2021) found the FIMI scores moderately to highly related with self-reported autistic traits.

Regarding the discriminant validity, it was presumed that mentalising is relatively independent of *reasoning ability* and *social motivation*. Reflecting on people's mental states seems to be a different process than reasoning about non-social entities on the neurological level (van Overwalle, 2011). Therefore, FIMI-G scores should only be weakly related to scores in a non-social reasoning test, if at all. As reflected in the weak relationship found between theory-of-mind skills and social motivation (Devine & Apperly, 2022), the desire to affiliate with others may not substantially be decreased by low mentalising ability or increased by high mentalising ability. Thus, at most, a weak relationship between FIMI-G scores and a social motivation measure should emerge. Finally, it was explored whether responding to the FIMI-G is associated with two facets of socially desirable responding (i.e., *self-deceptive enhancement* and *impression management*) and whether such tendencies would threaten the validity of the FIMI-G by affecting the relationships to the other measured validity criteria.

Method

Translation of the FIMI

For the German translation of the FIMI, we followed a hybrid method for cross-cultural adaptation. This approach combines both the efficiency of neural machine translation (NMT) with the nuanced understanding of a language expert. As Goyal et al. (2021) found, a hybrid method yields highly accurate translations while effectively preserving content validity. We implemented this approach as follows: First, the four items of the English FIMI were translated into German using the online software DeepL, which translates texts using artificial neural networks (DeepL, 2021). The resulting German items were then back-translated to English by an English-German bilingual sociologist with a PhD and native fluency in both languages. There were only minor differences between the English original and the English back-translation, indicating that the German translation was appropriate. Next, the bilingual sociologist and the first author discussed the differences between the original English FIMI and the back-translation. Based on the results of this discussion, minor adjustments to three of the German items were made. The final German translation of the FIMI, the FIMI-G, is presented in Table 1.

Table 1

Item Wordings and Parameters

#	Item wording in German	Item content in English	<i>M</i>	<i>SD</i>	<i>r</i> _{it}
1	Ich finde es einfach, mich in die Lage eines anderen zu versetzen.	Putting oneself in somebody else's shoes	3.08	0.70	.70
2 ^a	Ich finde es manchmal schwierig, die Dinge aus der Sicht anderer zu sehen.	Seeing things from other people's point of view	2.77	0.80	.66
3	Ich versuche manchmal, meine Freunde besser zu verstehen, indem ich mir vorstelle, wie die Dinge aus ihrer Perspektive aussehen.	Trying to understand friends	3.20	0.65	.57
4	Ich kann in der Regel die Sichtweise einer anderen Person verstehen, auch wenn sie sich von meiner eigenen unterscheidet.	Understanding another person's viewpoint	3.08	0.69	.64

Note. *N* = 283. *r*_{it} = item-total correlation (corrected). The FIMI-G is presented with the following German instruction: "Lesen Sie bitte jede Aussage und geben Sie an, inwieweit Sie zustimmen oder nicht zustimmen" ("Please read each statement and indicate the extent to which you agree or disagree"). The German items are rated on the following scale: *starke Ablehnung* (strongly disagree; 1), *leichte Ablehnung* (slightly disagree; 2), *leichte Zustimmung* (slightly agree; 3), and *starke Zustimmung* (strongly agree; 4).

^aThe responses to Item 2 were inverted.

Participants

The final sample for the analyses consisted of 283 German-speaking individuals from Austria, Germany, and Switzerland (59% male, 41% female; $M_{\text{age}} = 41.52$, $SD_{\text{age}} = 12.50$, range of age: 18–74 years), recruited from the Clickworker platform (<https://www.clickworker.de/>). Another 69 individuals were excluded from the analyses due to failing the attention check ($n = 66$), German not being their first language ($n = 1$), or not completing the FIMI-G ($n = 2$). The respondents received €2 for participation.

Procedure

After giving informed consent, the participants provided sociodemographic data, followed by an attention-check measure (see Bertrams, 2021b; Bertrams & Schlegel, 2020). Next, the participants completed the measures described below, including the FIMI-G. The measures were presented in randomised order, except for the reasoning ability test, which was administered at the end. The complete materials are available in the [Supplementary Materials](#). Finally, the participants were thanked, debriefed, and paid.

Measures

Mentalising

The FIMI-G was presented with the items, instructions, and response scale (see [Table 1](#) and the respective table notes). Higher overall FIMI-G scores express higher mentalising ability.

Perspective Taking

The German version of the subscale *perspective taking* of the Interpersonal Reactivity Index (Grevenstein, 2020; Paulus, 2009) was applied. This measure consists of four items (e.g., “I try to look at everybody’s side of a disagreement before I make a decision”), which are completed on a scale from *never* (1) to *always* (5). Higher overall scores indicate higher perspective taking.

Empathic Concern

The subscale *empathic concern* of the German Interpersonal Reactivity Index (Grevenstein, 2020; Paulus, 2009) was used. The four items (e.g., “I have warm feelings for people who are less well off than I am”) are responded to on a scale from *never* (1) to *always* (5). The higher the overall score, the higher the self-reported empathic concern.

Communication and Reciprocity

The participants completed the ten items (e.g., “I frequently find that I don’t know how to keep a conversation going”) of the subscale *communication and reciprocity* of the

German Autism-Spectrum Quotient (Freitag et al., 2007). A six-point scale ranging from *definitely disagree* (1) to *definitely agree* (6), instead of the usual four-point scale, was applied to increase reliability (Bertrams, 2021a). Higher overall scores mean higher skills in communication and social reciprocity.

Reasoning Ability

The mini-q (Baudson & Preckel, 2016), which is directly based on Baddeley's Grammatical Reasoning Test (Baddeley, 1968), was applied. In this test, the participants are presented with 64 statements, each describing the spatial position of a circle, a triangle, and a square to each other in social wording (e.g., "the triangle prefers the circle", "the triangle is not rejected by the circle"). The wording of the statements varies in voice—active (e.g., "prefers") or passive (e.g., "is preferred")—and is either positive (e.g., "prefers") or negative (e.g., "does not prefer"). Each provided statement is either true or false depending on the spatial distances between the circle, the triangle, and the square. For each statement, the participant decides as quickly as possible whether it is true or false. A higher number of correct decisions within a time limit of 3 minutes indicates a higher reasoning ability.

Social Motivation

The subscale *affiliation* of the Unified Motive Scales (Schönbrodt & Gerstenberg, 2012) was used to measure social motivation. The ten items (e.g., "I try to be in the company of friends as much as possible") are answered on a six-point scale ranging from *does not apply to me at all* (0) to *applies to me perfectly* (5). Higher overall scores reflect higher social motivation.

Socially Desirable Responding

Two facets of social desirability were captured with the German version of the Balanced Inventory of Desirable Responding (Musch et al., 2002). A total of ten items (e.g., "I always know why I like things") measure *self-deceptive enhancement*, and another ten items (e.g., "There have been occasions when I have taken advantage of someone", reverse scored) measure *impression management*. The participants gave their responses on seven-point scales ranging from *completely disagree* (1) to *completely agree* (7). The higher the overall scores, the higher the tendency to respond in a socially desirable manner in terms of self-deceptive enhancement or impression management.

Results

The raw data are available in the [Supplementary Materials](#) (note that the Clickworker IDs have been deleted from the raw data file to protect the participants' personal rights).

Item Parameters

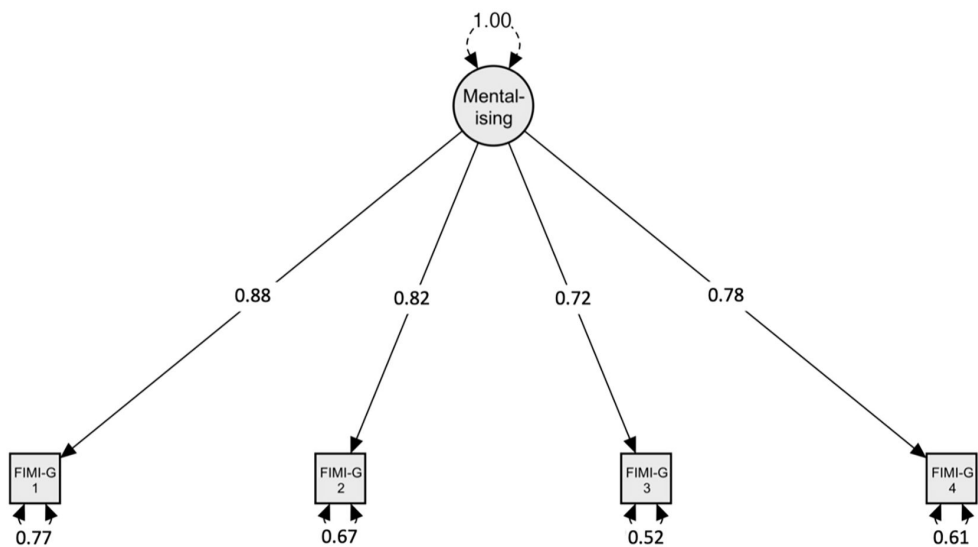
Table 1 shows the means, standard deviations, and corrected item-total correlations for each of the four FIMI-G items. There was no floor or ceiling effect apparent. The corrected item-total correlations were sufficiently high ($\geq .57$), indicating a homogenous overall scale.

Factor Structure

To examine the factor structure, a CFA was applied using Mplus (Muthén & Muthén, 1998–2017) and the weighted least squares mean and variance adjusted (WLSMV) estimator to account for the ordinal character of the data. Clutterbuck et al.'s (2021) findings suggest a structural model with all four items incorporated into one common factor. Given that no alternative models were evident, no model comparisons were made and only the model fit of the one-factor model was examined. Figure 1 depicts the model and the factor loadings, all of which were significant at $p < .001$.

Figure 1

Model Plot Reporting Standardised Parameters



Note. $N = 283$. The model was estimated using the weighted least squares mean and variance adjusted (WLSMV) estimator. Item 2 was inverted prior to the analysis.

For the evaluation of the model fit, we relied on the guidelines given by Schermelleh-Engel et al. (2003) for *acceptable* model fit ($RMSEA \leq .08$, $CFI \geq .95$, $SRMR \leq .10$) and *good* model fit ($RMSEA \leq .05$, $CFI \geq .97$, $SRMR \leq .05$). According to those guidelines, the

model fit was good, $\chi^2(2) = 6.47$, $p = .039$, SRMR = .018, CFI = .997, with the exception of the RMSEA = .089, 90% CI [.017, .169]. However, we do not consider RMSEA as a reliable indicator of model fit for the present research because our one-factor model has only 2 degrees of freedom and was tested in a rather small sample size (Kenny et al., 2015). The present findings agree with those of Clutterbuck et al.'s (2021) unidimensional factor solution for the original English FIMI. Additionally, we analysed the data using the maximum likelihood (ML) estimation which yielded very similar results regarding the model fit, $\chi^2(2) = 5.72$, $p = .057$, SRMR = .020, CFI = .990, RMSEA = .081, 90% CI [.000, .163], and the significance of the factor loadings (all $ps < .001$).

Reliability

For the FIMI-G, McDonald's ω was .82 independently of whether it was based on ML or WLSMV estimation, indicating sufficiently high reliability in terms of inner consistency.

Validity

Table 2 shows the descriptive statistics and intercorrelations for the applied measures. The FIMI-G scores were correlated with perspective taking, empathic concern, and communication and reciprocity in the expected manner. Moreover, the responses to the FIMI-G were distinct from reasoning ability and social motivation, as the correlations with these variables were non-existent or very small, respectively. These empirical findings suggest that the FIMI-G is a valid measure.

However, there was some overlap between responding to the FIMI-G and responding in a socially desirable way, as indicated by the statistically significant small- to medium-sized correlations with self-deceptive enhancement and impression management. To estimate the influence of self-deceptive enhancement and impression management, partial correlations were performed. When the two facets of socially desirable responding were statistically controlled for, the FIMI-G scores were still substantially related to perspective taking ($r_{\text{partial}} = .55$, $p < .001$), empathic concern ($r_{\text{partial}} = .49$, $p < .001$), as well as communication and reciprocity ($r_{\text{partial}} = .45$, $p < .001$). In addition, controlling for self-deceptive enhancement and impression management did not change the correlation with reasoning ability ($r_{\text{partial}} = .02$, $p = .77$) and changed the correlation with social motivation only slightly ($r_{\text{partial}} = .08$, $p = .17$). In sum, the correlations found provided initial evidence for the validity of the new FIMI-G, as they were in line with the assumptions and could not be alternatively explained by socially desirable responding.

Table 2
 McDonald's Omegas (ω), Means (M), Standard Deviations (SD), and Correlations

Variable	ω	M	SD	Correlation														
				1	2	3	4	5	6	7	8	9						
1. Mentalising (FIMI-G)	.82	3.03	0.57	—														
2. Perspective taking (IRI-PT)	.82	3.70	0.64	.60***	—													
3. Empathic concern (IRI-EC)	.79	3.52	0.66	.47***	.36***	—												
4. Communication and reciprocity (AQ)	.80	2.80	0.69	.54***	.34***	.16**	—											
5. Reasoning ability (BGRT)	—	15.79	4.79	.02	.02	-.03	.11	—										
6. Social motivation (UMS)	.92	2.32	1.05	.12*	.08	.15*	.21***	.06	—									
7. Self-deceptive enhancement (BIDR-SDE)	.77	4.25	0.82	.34***	.35***	.03	.51***	.02	.12*	—								
8. Impression management (BIDR-IM)	.75	3.93	1.02	.21***	.22***	.24***	.22***	<.001	.06	.39***	—							
9. Age	—	41.52	12.50	.11	.10	.11	.19**	-.29***	-.23***	.18**	.08	—						
10. Sex	—	—	—	.21***	-.03	.23***	.07	.05	-.02	-.07	.14*	-.09	—					

Note. $N = 283$ (for empathic concern; $n = 282$; for cognitive ability; $n = 279$). The overall scores of the self-report measures were obtained by averaging the responses to the items of the scales. The cognitive ability score represents the number of correct responses. FIMI-G = Four-Item Mentalising Index-German; IRI = Interpersonal Reactivity Index; PT = Perspective Taking; EC = Empathic Concern; AQ = Autism-Spectrum Quotient; BGRT = Baddeley's Grammatical Reasoning Test; UMS = Unified Motive Scales; BIDR = Balanced Inventory of Desirable Responding; SDE = Self-Deceptive Enhancement; IM = Impression Management. Coding for sex: 1 = male, 2 = female.

* $p < .05$. ** $p < .01$. *** $p < .001$.

It is worth mentioning that the FIMI-G total score did not correlate with age. There was a small effect for sex, in that female compared to male participants reported higher FIMI-G scores. These findings on age and sex agree with Clutterbuck et al.'s (2021) findings for the original English FIMI. The sex differences found in the present study are also in line with very recent findings of Clutterbuck et al. (2023) who found that women reported higher FIMI scores than men. Moreover, controlling for sex did not change how the FIMI-G was related to the other applied measures.

Conclusion

To conclude, the FIMI-G appears to be an appropriate translation of the FIMI given the initial findings on its item parameters, factor structure, inner consistency, and relationships with relevant validity criteria. Thus, the FIMI-G may be useful for measuring mentalising ability in German-speaking samples. However, some questions regarding the new measure still need to be answered, including its applicability to specific research contexts (e.g., autism research; see Clutterbuck et al., 2021) and the stability or sensitivity to change of the responses over time.

Funding: The authors have no funding to report.

Acknowledgments: The authors thank Sara Möser for her help in translating the scale.

Competing Interests: The authors have declared that no competing interests exist.

Ethics Statement: The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

Data Availability: The complete materials and the dataset generated for this study are available (see Bertrams et al., 2023).

Supplementary Materials

For this article, the complete materials and the dataset are available (see Bertrams et al., 2023).

Index of Supplementary Materials

Bertrams, A., Blaise, M., & Krispenz, A. (2023). *ResearchBox #2323: 'German translation of the Four-Item Mentalising Index (FIMI-G)' [Data, surveys]*. ResearchBox. <https://researchbox.org/2323>

References

- Baddeley, A. D. (1968). A 3 min reasoning test based on grammatical transformation. *Psychonomic Science*, 10(10), 341–342. <https://doi.org/10.3758/BF03331551>
- Baudson, T. G., & Preckel, F. (2016). mini-q: Intelligenzscreening in drei Minuten [mini-q: A three-minute intelligence screening]. *Diagnostica*, 62(3), 182–197. <https://doi.org/10.1026/0012-1924/a000150>
- Bertrams, A. (2021a). Internal reliability, homogeneity, and factor structure of the ten-item Autism-Spectrum Quotient (AQ-10) with two additional response categories. *Experimental Results*, 2, Article e3. <https://doi.org/10.1017/exp.2020.70>
- Bertrams, A. (2021b). Less illusion of a just world in people with formally diagnosed autism and higher autistic traits. *Journal of Autism and Developmental Disorders*, 51(10), 3733–3743. <https://doi.org/10.1007/s10803-020-04831-7>
- Bertrams, A., & Schlegel, K. (2020). Speeded reasoning moderates the inverse relationship between autistic traits and emotion recognition. *Autism*, 24(8), 2304–2309. <https://doi.org/10.1177/1362361320937090>
- Clutterbuck, R. A., Callan, M. J., & Shah, P. (2023). Socio-demographic and political predictors of Theory of mind in adulthood. *PloS ONE*, 18(5), Article e0284960. <https://doi.org/10.1371/journal.pone.0284960>
- Clutterbuck, R. A., Callan, M. J., Taylor, E. C., Livingston, L. A., & Shah, P. (2021). Development and validation of the Four-Item Mentalising Index. *Psychological Assessment*, 33(7), 629–636. <https://doi.org/10.1037/pas0001004>
- Davis, M. H. (1980). *Interpersonal Reactivity Index (IRI)* [Database record]. APA PsycTests. <https://doi.org/10.1037/t01093-000>
- DeepL. (2021, November 1). *How does DeepL work?* <https://www.deepl.com/en/blog/how-does-deepl-work>
- Devine, R. T., & Apperly, I. (2022). Willing and able? Theory of mind, social motivation and social competence in middle childhood and early adolescence. *Developmental Science*, 25(1), Article e13137. <https://doi.org/10.1111/desc.13137>
- 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>
- Freitag, C. M., Retz-Junginger, P., Retz, W., Seitz, C., Palmason, H., Meyer, J., Meyer, J., Rösler, M., & von Gontard, A. (2007). Evaluation der deutschen Version des Autismus-Spektrum-Quotienten (AQ): Die Kurzversion AQ-k [German adaptation of the Autism-Spektrum Quotient (AQ): Evaluation and short version AQ-k]. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 36(4), 280–289. <https://doi.org/10.1026/1616-3443.36.4.280>
- Frith, U., & Frith, C. (2001). The biological basis of social interaction. *Current Directions in Psychological Science*, 10(5), 151–155. <https://doi.org/10.1111/1467-8721.00137>

- Goyal, A. K., Bakshi, J., Panda, N. K., Kapoor, R., Vir, D., Kumar, K., Aneja, P., Singh, B., Gupta, M., & Walia, S. S. (2021). A hybrid method for the cross-cultural adaptation of self-report measures. *International Journal of Applied Positive Psychology*, 6, 45–54.
<https://doi.org/10.1007/s41042-020-00039-3>
- Grevenstein, D. (2020). Factorial validity and measurement invariance across gender groups of the German version of the Interpersonal Reactivity Index. *Measurement Instruments for the Social Sciences*, 2, Article 8. <https://doi.org/10.1186/s42409-020-00015-2>
- Happé, F., Cook, J. L., & Bird, G. (2017). The structure of social cognition: In(ter)dependence of sociocognitive processes. *Annual Review of Psychology*, 68, 243–267.
<https://doi.org/10.1146/annurev-psych-010416-044046>
- Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Research*, 44(3), 486–507.
<https://doi.org/10.1177/0049124114543236>
- Majdandžić, J., Amashauffer, S., Hummer, A., Windischberger, C., & Lamm, C. (2016). The selfless mind: How prefrontal involvement in mentalizing with similar and dissimilar others shapes empathy and prosocial behavior. *Cognition*, 157, 24–38.
<https://doi.org/10.1016/j.cognition.2016.08.003>
- Musch, J., Brockhaus, R., & Bröder, A. (2002). Ein Inventar zur Erfassung von zwei Faktoren sozialer Erwünschtheit [An inventory for the assessment of two factors of social desirability]. *Diagnostica*, 48(3), 121–129. <https://doi.org/10.1026//0012-1924.48.3.121>
- Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus User's Guide* (8th ed.). Muthén & Muthén.
- Parelman, J., Doré, B., & Falk, E. B. (2021). The role of mentalizing in communication behaviors. In M. Gilead & K. N. Ochsner (Eds.), *The neural basis of mentalizing* (pp. 579–595). Springer.
https://doi.org/10.1007/978-3-030-51890-5_30
- Paulus, C. (2009). *Der Saarbrücker Persönlichkeitsfragebogen SPF (IRI) zur Messung von Empathie: Psychometrische Evaluation der deutschen Version des Interpersonal Reactivity Index* [The SPF (IRI) for the measure of empathy: Psychometric evaluation of the German interpersonal reactivity index]. <http://hdl.handle.net/20.500.11780/3343>
- Preckel, K., Kanske, P., & Singer, T. (2018). On the interaction of social affect and cognition: Empathy, compassion and theory of mind. *Current Opinion in Behavioral Sciences*, 19, 1–6.
<https://doi.org/10.1016/j.cobeha.2017.07.010>
- Reniers, R. L., Corcoran, R., Drake, R., Shryane, N. M., & Völlm, B. A. (2011). The QCAE: A questionnaire of cognitive and affective empathy. *Journal of Personality Assessment*, 93(1), 84–95. <https://doi.org/10.1080/00223891.2010.528484>
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equations models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, 8(2), 23–74.
- Schönbrodt, F. D., & Gerstenberg, F. X. R. (2012). An IRT analysis of motive questionnaires: The Unified Motive Scales. *Journal of Research in Personality*, 46(6), 725–742.
<https://doi.org/10.1016/j.jrp.2012.08.010>

van Overwalle, F. (2011). A dissociation between social mentalizing and general reasoning. *NeuroImage*, 54(2), 1589–1599. <https://doi.org/10.1016/j.neuroimage.2010.09.043>