Measuring Theory of Mind In Mental Disorders: Validation and Exploration of Theory of Mind Deficits Using the Faux-Pas Recognition Test

---

Inaugural-Dissertation zur Erlangung des Doktorgrades der Philosophie an der Ludwig-Maximilians-Universität München

vorgelegt von

Silvia Gradl

aus Linz

Linz, 2025

Acknowledgments

Elaborate Summary

Theory of Mind (ToM) is a social-cognitive ability that forms part of the broader concept of social cognition, which encompasses a number of subdomains and processes. Social cognition is defined as the processing of social stimuli relevant to understanding the behavior of others, which is crucial for social interaction (e.g., Arioli et al., 2018). ToM is one of the most researched aspects of social cognition. It refers to the ability to represent the mental states of oneself and others, including desires, beliefs, emotions, or intentions (Apperly, 2012; Frith & Frith, 2012; Happé et al., 2017). A significant proportion of researchers (66%) use ToM interchangeably with mentalizing, while a further 61% use it as a synonym for mindreading (Quesque et al., 2024). ToM can also be seen as a component of an individual's empathic ability. This broader concept also encompasses the emotional aspect of recognizing and sharing the emotional experiences of another person (Dvash & Shamay-Tsoory, 2014). In the context of empathy, the term ToM is used interchangeably with the facet of cognitive empathy. The second facet pertains to emotional empathy, which encompasses emotional contagion or the automatic sharing of emotional states. These two different facets are dependent on different brain networks (Shamay-Tsoory, 2011). ToM can then further be distinguished into several different facets, as it is not a monolithic ability. It can be decomposed into affective/hot and cognitive/cold aspects (Molenberghs et al., 2016; Shamay-Tsoory & Aharon-Peretz, 2007), decoding and reasoning, or a social-perceptual and social-cognitive component (Sabbagh, 2004).

An individual's ability to understand another person's behavior (i.e., ToM understanding) is the foundation for appropriate social responses (Dvash & Shamay-Tsoory, 2014; Frith & Frith, 2012; Luke & Banerjee, 2013). This was corroborated by a meta-analysis, which revealed a correlation between higher ToM scores in children and popularity within their peer group (Slaughter et al., 2015). Conversely, impairments in ToM are associated with suboptimal social functioning (Adegboye et al., 2017), as well as cognitive and communication challenges (Sprung, 2010). In particular, a longitudinal study has demonstrated that difficulties in understanding social blunders (i.e., faux-pas) are predictive of increased peer rejection, while peer rejection may impair the understanding of faux-pas (Luke & Banerjee, 2013). ToM impairments have also been reliably associated with mental disorders. For example, individuals with borderline personality disorder (BPD) or major depressive disorder (MDD) showed moderate to severe ToM impairments, depending on the measurement instrument used (Berecz et

al., 2016; Bora, 2021; Bora & Berk, 2016; Németh et al., 2018; Nestor et al., 2022; Pagnoni et al., 2022). The review of meta-analyses by Cotter et al. (2018) identified ToM impairments in 30 clinical conditions and thus suggested that social-cognitive processes serve as a transdiagnostic clinical marker of various clinical conditions. The Research Domain Criteria (RDoC) framework also suggests that ToM is a transdiagnostic factor underlying mental disorders. RDoC offers an alternative approach to understanding mental disorders. Rather than relying on different categorical diagnoses, this approach aims to describe mental disorders in terms of their underlying impairment rather than by different categorical diagnoses (Michelini et al., 2021; National Institutes of Mental Health, 2020). The inclusion of ToM in the RDoC framework was driven by the prevalence of ToM impairments and their relevance to symptoms.

In mental disorders, ToM impairments can vary between no/less ToM and exceeding ToM. The distinction between these different types of ToM errors has been a focus of considerable research on mental disorders (Fretland et al., 2015; McLaren et al., 2022; Vaskinn et al., 2015). Mental states can either be oversimplified (no/less ToM) or over-ascribed (exceeding ToM), which can result in attributing intentions to random behavior (e.g., Vegni et al., 2021). A recent meta-analysis by McLaren et al. (2022) found that exceeding ToM was associated with a wide range of mental disorders, suggesting that exceeding ToM is a core feature of general psychopathology. However, ToM impairments have been investigated using the Movie for the Assessment of Social Cognition (MASC). Many studies have used the MASC to investigate specific ToM deficits. However, although the MASC is a valid ToM measure (Fossati et al., 2018), the validation of different ToM errors is still pending. Recent research has identified significant challenges in measuring ToM (Olderbak & Wilhelm, 2020; Quesque & Rossetti, 2020). The conceptual variability of ToM has led to a plethora of different measures that aim to assess ToM, but sometimes measure different constructs or not all facets of ToM. For example, a widely used ToM measurement instrument, the Reading the Mind in the Eyes Test (RMET), is intended to measure emotion recognition rather than ToM (Kittel et al., 2022; Quesque & Rossetti, 2020). In order to standardize the criteria for ToM measurement tools, Quesque and Rossetti (2020) suggest two criteria that a ToM test fulfill; test takers need to 1) represent others' mental states and 2) be able to distinguish these from their own. The MASC and the Faux-Pas Recognition Test (FPT) fulfill both criteria. The FPT

is, according to Osterhaus and Bosacki (2022), one of the five tests that account for more than 60% of ToM measurements used. While research has highlighted significant concerns regarding the validity and reliability of ToM measures (Hayward & Homer, 2017), the FPT demonstrated good reliability and validity (Şandor & İşcen, 2021). However, a more advanced evaluation of its psychometric properties, including a Rasch analysis, is still pending. Compared to the MASC, the FPT is easy to implement as it can be administered in a paper-and-pencil or digital format. Nevertheless, it utilizes an open-ended response format which complicates the application and scoring of the FPT (Stone et al., 1998). The development of a closed-answer format for the FPT would facilitate administration and increase the objectivity of evaluation. Furthermore, it would facilitate objectively validating the test. Moreover, as the validation of different ToM errors of the MASC is still pending, it would be necessary to develop and validate a polytomous answer format of the FPT that assesses specific ToM impairments. Furthermore, it would be beneficial to investigate ToM error patterns in clinical samples with the FPT, as thus far this has only been done with insufficiently valid ToM measures.

Study 1 examined the validity of the newly developed dichotomous answer format of the FPT in a community sample. The study assessed the accuracy of the closed answer format in measuring a single underlying construct (ToM) and estimated the difficulty of the items in this context. Results indicated that the FPT measures a single construct and that the dichotomous closed answer format has been successfully operationalized. The results demonstrate a strong alignm3ent between the model and the data, indicating the FPT's psychometric suitability and accuracy in measurement. This indicates that the FPT is able to measure correct ToM reliably and accurately. The FPT showed a strong correlation with the MASC and the RMET; however, the effect sizes were closer to the moderate range, suggesting related but distinct constructs. The level of correlation indicates that while the tests assess similar constructs, they also capture unique aspects of ToM, reinforcing the value of the tests as a distinct and valuable tool for ToM assessment. To summarize, the study confirms the dichotomous version of the FPT as a valid instrument for assessing ToM. Thus, the FPT represents a robust instrument for capturing ToM abillities that complements existing ToM measures. The results of the study also contribute to the existing knowledge about ToM and provide valuable insights for both the theoretical understanding and practical application of ToM measurements.

Study 2 examined the validity of the new polytomous answer format of the FPT in both a community and a clinical sample and examined associations between the FPT and the RMET. The clinical sample comprised patients with BPD. This new format categorizes responses into different levels of ToM abilities: no ToM, correct ToM, and exceeding ToM, providing a more nuanced assessment beyond traditional binary classifications. The development of this format was guided by findings in the literature indicating that individuals with mental disorders, such as BPD, exhibit specific ToM error patterns, particularly a tendency toward exceeding ToM errors (McLaren et al., 2022). Results confirmed that the polytomous format accurately captures ascending levels of difficulty, making the FPT an effective tool for discriminating between low and high ToM abilities. The model fit in the BPD sample was weaker, likely due to the frequent but inconsistent use of exceeding ToM responses. Although the FPT showed only moderate correlations with the RMET, this divergence does not undermine its validity, as the two tests assess different aspects of ToM. Notably, this study is the first to validate different ToM levels, previously observed using either insufficiently valid questionnaires or the MASC (e.g., Fretland et al., 2015; McLaren et al., 2022; Sharp & Vanwoerden, 2015), which has yet to be validated for ToM errors. By capturing a broader spectrum of ToM abilities, the polytomous format of the FPT enhances ToM assessments and contributes to a deeper understanding of ToM deficits.

Study 3 examined the differences in ToM levels (no ToM, correct ToM, exceeding ToM) using the newly validated polytomous answer format of the FPT. Specifically, the study investigated differences in ToM levels among individuals with BPD or MDD and a community sample, as well as differences in ToM levels between individuals with BPD and MDD, taking depression symptoms into account. Based on the literature, it was expected that exceeding ToM would be related to both BPD and MDD (McLaren et al., 2022). Moreover, the study investigated associations between the subscales of the MASC and the FPT, respectively. Results indicated that BPD was associated with exceeding ToM, regardless of depression symptoms. However, there was no association between exceeding ToM and MDD diagnosis. Individuals with MDD showed deficits in correct ToM, compared to the community sample. Importantly, the FPT's subscales for correct ToM and exceeding ToM demonstrated strong correlations with the corresponding subscales of the MASC, supporting the measure's convergent validity. However, the no ToM subscale of the FPT was not associated with the no or less ToM subscales of the MASC, reflecting differences in the constructs these tools

measure. This is the first study to examine different types of ToM errors using the FPT, as previous studies have predominantly used the MASC to analyze ToM errors, or used questionably valid questionnaires for this purpose. In the FPT, the different ToM levels have already been validated in Study 2. The results of Study 3 thus contribute to the growing body of literature by demonstrating specific ToM deficits of individuals with BPD and MDD using the FPT. As exceeding ToM is uniquely associated with BPD, this study challenges the notion that exceeding ToM is related to general psychopathology. The findings of Study 3 thus contribute to the expanding body of literature by demonstrating specific ToM deficits of individuals with BPD and MDD by using the FPT. As ToM serves as a treatment target for improving mental health (Kvarstein et al., 2020), these findings emphasize the need for further research on associations between different ToM impairments in mental disorders.

In conclusion, this dissertation validated the new closed answer formats of the FPT and examined ToM abilities across different samples. The studies supported the validity of the dichotomous and the polytomous answer formats and demonstrated that specific ToM impairments were present in BPD and MDD, which were captured by the polytomous response format. This enables a more differentiated assessment of ToM abilities in these groups. The results contribute to the current ToM literature on specific ToM deficits in mental disorders and emphasize the importance of developing and validating measurement instruments. Future research should further investigate ToM deficits in larger samples and in a broader ranges of mental disorders, such as autism or schizophrenia. Overall, this dissertation strengthens the FPT as a robust ToM measurement tool offering valuable insights for the identification of specific ToM deficits in clinical samples, especially in individuals with BPD and MD.

Deutsche Zusammenfassung

Die Theory of Mind (ToM) ist eine sozial-kognitive Fähigkeit, die Teil des umfassenderen Konzepts der sozialen Kognition ist, das eine Reihe von Teilbereichen und Prozessen umfasst. Soziale Kognition ist definiert als die Verarbeitung sozialer Stimuli, die für das Verständnis des Verhaltens anderer relevant sind (Arioli et al., 2018). Das bildet die Grundlage für soziale Interaktionen. ToM ist eine der am meisten erforschten Aspekte der sozialen Kognition. Sie bezieht sich auf die Fähigkeit, die mentalen Zustände (Wünsche, Überzeugungen, Emotionen oder Absichten) von sich selbst und anderen zu repräsentieren (Apperly, 2012; Frith & Frith, 2012; Happé et al., 2017). Ein großer Anteil der Forschenden (66 %) verwendet ToM synonym mit dem Begriff Mentalisieren, während weitere 61 % es als Synonym für Gedankenlesen verwenden (Quesque et al., 2024). ToM kann auch als eine Komponente der Empathiefähigkeit einer Person betrachtet werden. Dieses umfassendere Konzept beinhaltet auch den emotionalen Aspekt des Erkennens und Teilens der emotionalen Erfahrungen einer anderen Person (Dvash & Shamay-Tsoory, 2014). Im Zusammenhang mit Empathie wird der Begriff ToM synonym mit der Facette der kognitiven Empathie verwendet. Die zweite Facette betrifft die emotionale Empathie, die die emotionale Ansteckung oder das automatische Teilen von Gefühlszuständen umfasst. Diese beiden unterschiedlichen Facetten sind von verschiedenen Gehirnnetzwerken abhängig (Shamay-Tsoory, 2011). ToM kann in mehrere verschiedene Aspekte unterteilt werden, da es sich nicht um eine monolithische Fähigkeit handelt. Sie kann in affektive/heiße und kognitive/kalte Facetten (Molenberghs et al., 2016; Shamay-Tsoory & Aharon-Peretz, 2007), Dekodierung und Schlussfolgern oder eine sozial-perzeptive und sozial-kognitive Komponente (Sabbagh, 2004) unterteilt werden. Die sozial-perzeptive Komponente entspricht der affektiven/heißen ToM, wohingegen die sozial-kognitive Komponente der kognitiven ToM entspricht.

Die Fähigkeit eines Individuums, das Verhalten einer anderen Person zu verstehen (d. h. das ToM-Verständnis), ist die Grundlage für angemessene soziale Interaktionen (Dvash & Shamay-Tsoory, 2014; Frith & Frith, 2012; Luke & Banerjee, 2013). Dies wurde durch eine Meta-Analyse bestätigt, die einen Zusammenhang zwischen höheren ToM-Werten bei Kindern und der Beliebtheit in ihrer Peer-Gruppe aufgezeigt hat (Slaughter et al., 2015). Im Gegensatz dazu, werden Beeinträchtigungen der ToM mit suboptimalem sozialen Verhalten (Adegboye et al.,

2017) sowie mit kognitiven und kommunikativen Herausforderungen in Verbindung gebracht (Sprung, 2010). Eine Längsschnittstudie hat insbesondere gezeigt, dass Schwierigkeiten beim Verstehen von sozialen Fehlern (d. h. Fauxpas) ein Prädiktor für verstärkte Ablehnung durch Gleichaltrige sind, während umgekehrt Ablehnung durch Gleichaltrige das Verständnis von Fauxpas beeinträchtigen kann (Luke & Banerjee, 2013).

Beeinträchtigungen der ToM wurden auch zuverlässig mit psychischen Störungen in Verbindung gebracht. So zeigten beispielsweise Personen mit einer Borderline Persönlichkeitsstörung (BPS) oder einer Major Depression (MD) mittlere bis starke ToM-Beeinträchtigungen, je nach verwendetem Messverfahren (Berecz et al., 2016; Bora, 2021; Bora & Berk, 2016; Németh et al., 2018; Nestor et al., 2022; Pagnoni et al., 2022). Ein Review über Meta-Analysen von Cotter et al. (2018) identifizierte ToM-Beeinträchtigungen bei 30 klinischen Störungen und schlug somit vor, dass sozial-kognitive Prozesse als transdiagnostischer klinischer Marker für verschiedene klinische Störungen dienen sollen. Auch die Research Domain Criteria (RDoC) legt nahe, dass ToM ein transdiagnostischer Faktor ist, der psychischen Störungen zugrunde liegt. RDoC bietet einen alternativen Ansatz zum Verständnis psychischer Störungen. Anstatt sich auf verschiedene kategorische Diagnosen zu stützen, zielt dieser Ansatz darauf ab, psychische Störungen im Hinblick auf die ihnen zugrunde liegende Beeinträchtigung zu beschreiben (Michelini et al., 2021; National Institutes of Mental Health, 2020). Die Aufnahme von ToM in RDoC wurde durch die Prävalenz von ToM-Beeinträchtigungen und deren Relevanz für Symptome begründet. Bei psychischen Störungen können die Beeinträchtigungen der ToM zwischen keine/weniger ToM und übermäßige ToM variieren. Die Unterscheidung zwischen diesen verschiedenen Arten von ToM-Fehlern ist ein Schwerpunkt der Forschung zu psychischen Störungen (Fretland et al., 2015; McLaren et al., 2022; Vaskinn et al., 2015). Mentale Zustände können entweder zu stark vereinfacht (keine/weniger ToM) oder überinterpretiert werden (übermäßige ToM), was dazu führen kann, dass einem zufälligen Verhalten Absichten zugeschrieben werden (Vegni et al., 2021). Eine kürzlich durchgeführte Meta-Analyse von McLaren et al. (2022) ergab, dass eine übermäßige ToM mit einer Vielzahl von psychischen Störungen assoziiert ist, was darauf hindeutet, dass eine übermäßige ToM ein Kernmerkmal der allgemeinen Psychopathologie ist. ToM-Beeinträchtigungen wurden jedoch mit dem Movie for the Assessment of Social Cognition (MASC) untersucht. Viele Studien haben den MASC

verwendet, um spezifische ToM-Defizite zu untersuchen. Doch obwohl der MASC ein valides ToM-Maß ist (Fossati et al., 2018), steht die Validierung der spezifischen ToM-Fehler noch aus.

Neuere Forschungen haben erhebliche Herausforderungen bei der Messung der ToM identifiziert (Olderbak & Wilhelm, 2020; Quesque & Rossetti, 2020). Die konzeptionelle Variabilität der ToM hat zu einer Fülle von verschiedenen Messungen geführt, die darauf abzielen, ToM zu erfassen, aber manchmal unterschiedliche Konstrukte oder nicht alle Facetten der ToM messen. So soll zum Beispiel ein weit verbreitetes ToM-Messinstrument, der Reading the Mind in the Eyes Test (RMET), eher der Messung der Emotionserkennung als der ToM dienen (Kittel et al., 2022; Quesque & Rossetti, 2020). Um die Kriterien für ToM-Messinstrumente zu standardisieren, schlagen Quesque und Rossetti (2020) zwei Kriterien vor, die eine ToM-Messung erfüllen sollte: Die Testteilnehmenden müssen 1) die mentalen Zustände anderer erfassen und 2) in der Lage sein, diese von ihren eigenen zu unterscheiden. Der MASC und der Faux-Pas Recognition Test (FPT) erfüllen beide Kriterien. Der FPT ist laut Osterhaus und Bosacki (2022) einer der fünf Tests, die mehr als 60 % der verwendeten ToM-Messungen ausmachen. Während Forschungen erhebliche Bedenken hinsichtlich der Validität und Reliabilität von ToM-Messungen geäußert haben (Hayward & Homer, 2017), zeigte der FPT eine gute Reliabilität und Validität (Şandor & İşcen, 2021). Eine weitergehende Bewertung seiner psychometrischen Eigenschaften, einschließlich einer Rasch-Analyse, steht jedoch noch aus. Im Vergleich zum MASC ist der FPT einfacher in der Anwendung, da er in einem Papier-und-Bleistift- oder digitalen Format durchgeführt werden kann. Es wird jedoch ein offenes Antwortformat verwendet, das die Anwendung und Bewertung des FPT erschwert (Stone et al., 1998). Die Entwicklung eines geschlossenen Antwortformats für den FPT würde die Durchführung erleichtern und die Objektivität der Auswertung erhöhen. Außerdem würde dies die objektive Validierung des Tests erleichtern. Da die Validierung verschiedener ToM-Fehler des MASC noch aussteht, wäre es außerdem notwendig, ein polytomes Antwortformat, das spezifische ToM-Beeinträchtigungen erfasst, zu entwickeln und zu validieren. Darüber hinaus wäre es von Vorteil, ToM-Fehlermuster in klinischen Stichproben mit dem FPT zu untersuchen, da dies bisher nur mit unzureichend validen Messinstrumenten gemacht worden ist.

Studie 1 untersuchte die Gültigkeit des neu entwickelten dichotomen Antwortformats des FPT in einer Gemeinschaftsstichprobe. Die Studie bewertete die

Genauigkeit des geschlossenen Antwortformats bei der Messung eines einzigen zugrunde liegenden Konstrukts (ToM) und schätzte die Schwierigkeit der Items in diesem Zusammenhang ein. Die Ergebnisse zeigten, dass der FPT ein einziges Konstrukt misst und dass das dichotome geschlossene Antwortformat erfolgreich operationalisiert wurde. Die Ergebnisse zeigen eine starke Übereinstimmung zwischen dem Modell und den Daten, was auf die psychometrische Eignung und Messgenauigkeit des FPT hinweist. Dies deutet darauf hin, dass der FPT in der Lage ist, korrekte ToM zuverlässig und genau zu messen. Der FPT zeigte eine mäßige Korrelation mit dem MASC und einer anderen Messung der sozialen Kognition, dem RMET. Das Ausmaß der Korrelation deutet darauf hin, dass die Tests zwar ähnliche Konstrukte erfassen, aber auch einzigartige Aspekte der ToM, was den Wert des Tests als eigenständiges und wertvolles Instrument zur Bewertung der ToM unterstreicht. Zusammenfassend lässt sich sagen, dass die Studie die dichotome Version des FPT als valides Instrument zur Erfassung von ToM bestätigt. Somit stellt die dichotome Version des FPT ein robustes Instrument zur Erfassung von ToM-Fähigkeiten dar, das bestehende Messinstrumente ergänzt. Die Ergebnisse der Studie tragen darüber hinaus zum bestehenden Wissen über ToM bei und liefern wertvolle Erkenntnisse sowohl für das theoretische Verständnis als auch für die praktische Anwendung von ToM-Messungen.

Studie 2 untersuchte die Gültigkeit des neuen polytomen Antwortformats des FPT sowohl in einer Gemeinschafts- als auch in einer klinischen Stichprobe. Die klinische Stichprobe umfasste Personen mit BPD. Dieses neue Format kategorisiert die Antworten in verschiedene Stufen von ToM-Fähigkeiten: keine ToM, korrekte ToM und überdurchschnittliche ToM, was eine bessere Bewertung als die traditionellen binären Klassifizierungen ermöglicht. Die Entwicklung dieses Formats wurde aus Erkenntnissen in der Literatur abgeleitet, die darauf hinweisen, dass Personen mit psychischen Störungen, wie z. B. BPS, spezifische ToM-Fehlermuster aufweisen, vor allem einer Tendenz zum überinterpretieren von mentalen Zuständen (McLaren et al., 2022). Die Ergebnisse bestätigten, dass das polytome Antwortformat aufsteigende Schwierigkeitsgrade erfasst, was den FPT zu einem effektiven Instrument zur Unterscheidung zwischen niedrigen und hohen ToM-Fähigkeiten macht. Die Ergebnisse für die BPS-Stichprobe fielen etwas schlechter aus. Dies entspricht jedoch den Erwartungen, da Personen mit BPS zum ToM-Fehler des Überinterpretierens mentaler Zustände neigen, was die Antwortvielfalt verringert. Obwohl der FPT nur

schwache Korrelationen mit dem RMET aufwies, untergräbt diese Divergenz nicht seine Validität, da die beiden Tests unterschiedliche Aspekte der ToM erfassen. Insbesondere ist diese Studie die erste, die verschiedene ToM-Stufen validiert. Zuvor wurden zur Untersuchung verschiedener ToM-Stufen entweder mit unzureichend validen Fragebögen oder dem MASC (Fretland et al., 2015; McLaren et al., 2022; Sharp & Vanwoerden, 2015) durchgeführt, wobei der MASC bisher noch keine ToM-Fehler validiert hat. Durch die Erfassung eines breiteren Spektrums von ToM-Fähigkeiten verbessert das polytome Format des FPT die ToM-Bewertungen und trägt zu einem tieferen Verständnis von ToM-Defiziten bei.

Studie 3 untersuchte die Unterschiede im ToM-Niveau (kein ToM, korrekte ToM, übersteigerte ToM) zwischen Personen mit BPS oder MD und einer Gemeinschaftsstichprobe, sowie Unterschiede im ToM-Niveau zwischen Personen mit BPS und MD, unter Berücksichtigung der Depressionssymptomatik. Des Weiteren untersuchte die Studie Zusammenhänge zwischen den Subskalen des MASC mit denen des FPT. Basierend auf der Literatur wurde erwartet, dass eine übersteigerte ToM sowohl mit der BPS als auch mit der MD zusammenhängt (McLaren et al., 2022). Die Ergebnisse zeigten, dass die BPS-Diagnose, unabhängig vom Schweregrad der Depressionssymptomatik, mit einer übersteigerten ToM zusammen hing. Es zeigte sich jedoch kein Zusammenhang zwischen der MD-Diagnose und einer übersteigerten ToM. Personen mit MD zeigten im Vergleich zur Gemeinschaftsstichprobe jedoch Defizite in der korrekten ToM. Das ist die erste Studie, die verschiedene Arten von ToM-Fehlern mit dem FPT untersucht hat. Frühere Studien haben überwiegend den MASC zur Untersuchung von ToM-Fehlern verwendet, oder fragwürdig valide Fragebögen dafür herangezogen. Beim FPT wurden die verschiedenen ToM-Stufen bereits in Studie 2 validiert. Wichtig ist, dass die Subskalen des FPT für korrekte ToM und übersteigerte ToM mäßige bis starke Korrelationen mit den entsprechenden Subskalen des MASC aufwiesen, was die konvergente Validität der Messung unterstützt. Die Subskala „keine ToM" des FPT war jedoch nicht mit den Subskalen „keine" oder „weniger ToM" des MASC assoziiert, was die Unterschiede in den Konstrukten widerspiegelt, die diese Instrumente messen. Die Ergebnisse von Studie 3 leisten somit einen Beitrag zur wachsenden Literatur, indem sie spezifische ToM-Defizite von Personen mit BPD und MDD unter Verwendung des FPT aufzeigen. Somit wird die Vorstellung in Frage gestellt, dass eine übersteigerte ToM mit einer allgemeinen Psychopathologie in Zusammenhang steht. Da die ToM als

Behandlungsziel zur Verbesserung der psychischen Gesundheit dient (Kvarstein et al., 2020), unterstreichen diese Ergebnisse den Bedarf an weiterer Forschung über Zusammenhänge zwischen verschiedenen ToM-Beeinträchtigungen bei psychischen Störungen.

Zusammenfassend lässt sich sagen, dass in dieser Dissertation die neuen geschlossenen Antwortformate des FPT validiert und die ToM-Fähigkeiten in verschiedenen Stichproben untersucht wurden. Die Studien untermauerten die Validität des dichotomen und des polytomen Antwortformats und zeigten, dass spezifische ToM-Beeinträchtigungen bei BPS und MD vorlagen, die durch das polytome Antwortformat erfasst wurden. Das ermöglicht eine differenziertere Bewertung der ToM-Fähigkeiten in diesen Gruppen. Diese Ergebnisse tragen nicht nur zur aktuellen ToM-Literatur zu ToM-Defiziten in klinischen Störungen bei, sondern unterstreichen auch die Bedeutung der Verwendung von validen Messinstrumenten. Zukünftige Forschungsarbeiten sollten ToM-Defizite in größeren Stichproben und bei einem breiteren Spektrum psychischer Störungen, wie beispielsweise Autismus oder Schizophrenie, weiter untersuchen. Insgesamt stärkt diese Dissertation den FPT als ein robustes ToM-Messinstrument und bietet wertvolle Erkenntnisse für die Identifizierung von spezifischen ToM-Defiziten in klinischen Stichproben, insbesondere bei Personen mit BPS und MD.

Table of Contents

1   General Introduction

1.1   Theory of Mind: Definition and Historical Outline

The first attempts to measure and define Theory of Mind (ToM) were made on chimpanzees by Premack and Woodruff (1978), who published a groundbreaking study on the ability of chimpanzees to infer a person's intentions. In the study, they showed chimpanzees videos of humans confronted with various problems and examined the chimpanzees' ability to recognize the actors' intentions. They investigated this by presenting the chimpanzee with several photos of possible solutions to the various problems. The chimpanzee's consistent selection of the photos presenting the correct solution can be explained by the fact that the animal recognized the problem in the video, understood the actor's purpose, and chose alternatives that were compatible with this purpose. Premack and Woodruff (1978) defined ToM at the time as the imputation of states of mind to the human actor (presented in the videos).

Subsequently, ToM research was extended to children. Wimmer and Perner (1983) conducted the first systematic study on children's understanding of belief with their famous experiment on Maxi and the chocolate. Over the years, research has increased in scope and depth ToM has been studied across different ages (e.g., Wellman et al., 2001). Researchers have also explored the foundations for understanding ToM in neural processes (Dvash & Shamay-Tsoory, 2014; e.g., Mazza et al., 2015; Pitman et al., 2012) and development of ToM in typical and atypical populations (e.g., individuals with autism or schizophrenia; Chung et al., 2014). They have also expanded research on people with mental disorders, such as posttraumatic stress disorder, borderline personality disorder (BPD), or MDD (Cotter et al., 2018; Janssen et al., 2022; McLaren et al., 2022; Nestor et al., 2022; Richman & Unoka, 2015). The different traditions and approaches in ToM research start from a different understanding of what ToM is: a set of conceptual knowledge, cognitive processes, or a social competence that can vary from person to person. The different aspects then impact ToM research because they affect the selected ToM measure.

Concurrent with ToM research becoming more widespread, the vocabulary became highly heterogeneous. Various overlapping concepts and constructs of ToM were further developed, and definitions are often vague. Thus, differentiating them is not entirely possible. Nevertheless, the nomological network of ToM and relevant constructs will be explained below.

ToM refers to the ability to represent the mental states of oneself and others (Happé et al., 2017) and ascribe mental states such as desires, beliefs, emotions, or intentions to oneself and others (Apperly, 2012; Frith & Frith, 2012). That implies the ability to distinguish between one's own mental state and that of others (Quesque & Rossetti, 2020). ToM understanding enables an individual to explain and predict the behavior of another person and is therefore the basis for appropriate social responses (Dvash & Shamay-Tsoory, 2014; Frith & Frith, 2012; Luke & Banerjee, 2013). This was confirmed by a meta-analysis that found that higher ToM scores in children are related to popularity in their peer group (Slaughter et al., 2015). Particularly, one longitudinal study has shown that difficulties in understanding social blunders (i.e., faux-pas) predict increased peer rejection, while peer rejection may impair the understanding of faux-pas (Luke & Banerjee, 2013). Conversely, ToM impairments are related to poor social functioning (Adegboye et al., 2017) as well as cognitive and communication problems (Sprung, 2010). Many researchers use ToM interchangeably with mentalizing (66%) or mindreading (61%; Quesque et al., 2024).

1.2 Theory of Mind and Relevant Concepts

1.2.1 Social Cognition

ToM is a social-cognitive ability, embedded in the multifaceted concept of social cognition that encompasses a number of subdomains and processes. Social cognition has been defined as the processing of social stimuli relevant to understanding the behavior of others, which is crucial for social interaction (Adolphs, 1999; Arioli et al., 2018; Frith & Frith, 2012; Happé et al., 2017). The National Institutes of Mental Health integrated social cognition into the RDoC framework (RDoC), a transdiagnostic approach to understanding psychopathology. RDoC currently divides the domain of social processes into four constructs: 1) affiliation and attachment, 2) social communication, 3) perception and understanding of self, and 4) perception and understanding of others. The constructs are then divided into various subconstructs. For instance, the domain perception and understanding of others is divided into animacy perception, action perception, and understanding mental states (i.e., Theory of Mind).

Despite the clusters provided by RDoC, there is no clear consensus in the literature as to whether social cognition can be divided into three or four subdomains. The authors hardly agree on which social-cognitive processes should be differentiated and which are interrelated. For example, Arioli and colleagues (2018) clustered social

cognition into three domains, which include social perception, social understanding (i.e., empathy and ToM), and social decision-making. Penn et al. (2008) suggested the three constructs, emotion perception, ToM, and attributional style. Pinkham et al. (2013), however, identified four core domains: social perception, emotion processing, ToM, and attributional style. Green et al. (2015) named the four subdomains similarly: perception of social cues, experiencing and regulating emotions, mentalizing, and experience sharing.

Appropriate social interactions require several different processes: First, a social being must perceive the social cues of others, such as facial expressions, gestures, posture, body language, and voice. Once this information is integrated, the individual is able to resonate directly with the affective states of others (i.e., empathy), which includes experiencing and regulating emotions. Furthermore, the individual is able to interpret the observable behaviors in terms of mental states (i.e., ToM; Frith & Frith, 2012). The individual can then explain the causes of behaviors or make sense of social events or interactions (Arioli et al., 2018).

1.2.2 The Nomological Network of Theory of Mind

After mentioning empathy and ToM as parts of social cognition, it becomes apparent that there is no consensus in research on social cognition as to whether empathy and ToM are in the same (Arioli et al., 2018) or in different subdomains (Pinkham et al., 2013) of social cognition. There is a plethora of overlapping concepts and related constructs (Olderbak & Wilhelm, 2020). Progress in understanding these constructs has been hindered by challenges in assessment. The measurement problems will be discussed in another section. The following will discuss the distinction between empathy and ToM.

ToM can be considered as a component of an individual's empathic ability. This broader concept also includes the emotional aspect of recognizing and sharing the emotional experiences of another person (Dvash & Shamay-Tsoory, 2014). As part of empathy, ToM is equated with the facet of cognitive empathy. The second facet refers to emotional empathy, which refers to emotional contagion (i.e., automatically sharing emotional states). The different facets depend on separate brain networks (Shamay-Tsoory, 2011). Figure 1 illustrates the nomological network of ToM and the relationship between empathy and ToM. As shown in Figure 1, it is assumed that the cognitive and affective ToM are part of cognitive empathy (Dvash & Shamay-Tsoory, 2014). Affective or hot ToM refers to ascribing emotions and feelings to others, whereas cognitive ToM

3

refers to the ability to make inferences about others' thoughts, intentions, or beliefs (Molenberghs et al., 2016; Shamay-Tsoory & Aharon-Peretz, 2007).

ToM is not a monolithic construct. It should be noted that a number of subprocesses and precursors are involved in understanding the mental states of other people. In addition to cognitive (cold) and affective (hot) aspects, it can also be subdivided into the two processes of decoding and reasoning (Sabbagh, 2004). The process of decoding the mental states of others is based on the direct observation of the environment (i.e., of information that is readily available). In order to reason about these mental states, additional information such as knowledge and facts is required to explain and predict the actions of others (Harkness et al., 2005).

Another proposed decomposition of ToM suggests that it is based on a two-stage process. (Sabbagh, 2004). Thus, it comprises a social–perceptual component (i.e., using perceptual characteristics such as facial expressions, gestures, or body movements to recognize the mental states of others) and a social–cognitive component (i.e., drawing conclusions about the mental state of others without relying on perceptual cues; Nettle & Liddle, 2008; Sabbagh, 2004). Affective ToM is equivalent to the social-perceptual component, whereas cognitive ToM corresponds to the social-cognitive component. It has been shown that the social-perceptual component develops before the social-cognitive component and that these components affect different regions of the brain (Sabbagh, 2004).

The different facets are measured with different tasks. Schurz and colleagues (2021) found in their meta-analysis that tasks that are assumed to measure only one aspect, simultaneously measure both cognitive (e.g., explaining the speaker's intention in strange stories) and affective (e.g., recognizing facial emotions) aspects. Tasks that are linked to both aspects (cognitive and affective ToM), such as the Faux-pas Recognition Test (FPT), show high clinical discrimination. However, their interindividual differences could be due to various causes: differences only in cognitive processes, only in affective processes only, or in both combined.

Figure 1

Theory of Mind and Empathy as Multidimensional Constructs
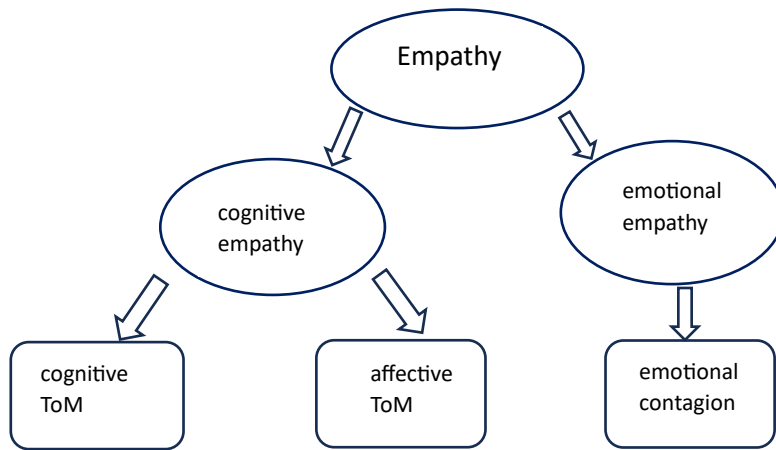
Note. Reproduced from Davsh and Shamay-Tsoory, 2014.

1.2.3 Theory of Mind Impairments

Even though Apperly (2012) argues that ToM is not a one-dimensional entity that individuals simply possess to a greater or lesser extent, the distinction between different ToM capabilities/types of error has inspired much research (Fretland et al., 2015; Vaskinn et al., 2015). ToM impairments can vary between no/less ToM and exceeding ToM. The complete lack of inferring mental states of others and only focusing on non-mental, such as physical, states is defined as no ToM (Dziobek et al., 2006; Happé, 1994). Less ToM is characterized by rather focusing on concrete events than on internal states when interpreting behavior (Cortés-García et al., 2021; Quek et al., 2018). This means that fewer mental states are attributed to others, statements are understood literally, and conclusions are oversimplified (Fretland et al., 2015; Montag et al., 2011; Sharp et al., 2011; Vaskinn et al., 2015). While some studies differentiate between no ToM and less/reduced ToM (Dziobek et al., 2006; Vaskinn et al., 2015), one study subsumes no and reduced ToM under the term "inframentalisation" (Lahera et al., 2014). Exceeding ToM, also called overmentalization (Vegni et al., 2021) or hypermentalization (Sharp & Vanwoerden, 2015) refers to the tendency to over-ascribe mental states to others (i.e., attributing intentions to random behavior).

1.3 Developmental Trajectories of Theory of Mind

Before measuring ToM in children, it is important to note that infants exhibit precursors of ToM that they express via nonverbal communication. Infants exhibit a preference for social stimuli from a very early age. It is crucial to identify others as social partners before being able to attribute beliefs and intentions to them. One study found that one-month-old infants showed the ability to differentiate between human and monkey faces and three-month-old infants even showed a preference for human faces (Sanefuji et al., 2014). However, looking at faces is not sufficient for understanding the mental states of others. An important preliminary stage is to be aware of what a person sees or perceives. One study showed that higher gaze-following abilities at 10.5 months preceded later ToM abilities at 4.5 years in children (Brooks & Meltzoff, 2015). Declarative joint attention, characterized as the shared focus on an object with another individual, predicts later ToM development, specifically false-belief understanding (Nelson et al., 2008; Sodian et al., 2020; Sodian & Kristen-Antonow, 2015).

Children's comprehension of false beliefs (FB; Wimmer & Perner, 1983) undermines an important milestone that researchers have focused on in the measurement of ToM. Understanding FBs means recognizing that individuals can hold beliefs that do not correspond to reality. ToM develops gradually, which means that more complex aspects of ToM emerge only after simpler aspects have been mastered (Moran, 2013). Research therefore supports a conceptual continuity in ToM development, with implicit understanding in infancy (such as gaze-following abilities), predicting explicit ToM in preschool years (Kloo et al., 2020; Sodian et al., 2020). Clements and Perner (1994) investigated implicit knowledge of FB by recording eye movements. They found that from the age of 2 years and 11 months, approximately 90% of children looked at the belief-based location. The third year of life also represents a critical period in children's socio-cognitive development. This period is characterized by a shift in language skills and development from implicit understanding to more explicit reasoning about mental states (Kaltefleiter et al., 2021). Explicit FB understanding requires verbal articulation skills from the protagonists. Many researchers have investigated at which age children come to understand explicit FB. To address the empirical inconsistencies and theoretical controversies, Wellman and colleagues (2001) conducted a meta-analysis that included 178 studies. They found that at 44 months of age, 50% of children were successful in FB tasks. After this age

there is a dramatic improvement in performance with increasing age. By the age of 48 months (4 years), children performed significantly above chance in explicit FB tasks, which represents a significant advance in their social-cognitive abilities.

Despite the predominant focus of research on ToM development during the preschool years, the development of ToM extends beyond these early years (Banerjee et al., 2011; Lecce et al., 2017; Osterhaus & Bosacki, 2022). Significant ToM progress is made during middle childhood (5-10 years; Osterhaus & Koerber, 2021). During elementary school, children acquire a more complex understanding of mental states (Hughes & Devine, 2015; Miller, 2009). With regard to the most important milestone in ToM research (i.e., understanding first-order FB), during this period, children acquire a more complex understanding of FB. The development of second-order FB understanding appeared to reach an asymptote around age 7, after which no substantial developmental progressions occurred (Osterhaus & Koerber, 2021). At this age, children acquire the understanding that mental states can be recursive. Understanding first- and second-order FB i.e., Advanced ToM or AToM, predicted later AToM at age 10 (Devine et al., 2016). In addition to understanding second-order FB, children begin to understand sarcasm, faux-pas, or white lies (Devine & Hughes, 2013; Osterhaus & Koerber, 2021). There are significant individual differences (Devine, 2021; Hughes & Devine, 2015) that are associated with various social and cognitive outcomes. These include children's relationships with peers (Banerjee et al., 2011; Slaughter et al., 2015), their experiences of loneliness (Koerber & Osterhaus, 2020), and their academic performance (Lecce & Devine, 2021; Osterhaus & Koerber, 2023).

The meta-analysis by Wellman et al. (2001) identified several factors that improve children's performance in all age groups, thus leaving the underlying developmental trajectory unchanged: motive for the underlying transformation, participation in the transformation, salience of the protagonist's mental state, and real presence of the target object (in this case: children are more likely to answer correctly when the target is not real and present).

In the literature, the vast majority of ToM research has focused on investigating ToM in childhood. Fewer studies have been conducted to assess ToM in adolescents and adults (Hughes & Devine, 2015), although adolescence is an equally important period for socio-emotional development (Fuhrmann et al., 2015; Steinberg, 2005). During adolescence, self- and social understanding are shaped by social interactions (Carpendale & Lewis, 2004; Zerwas et al., 2004). Dealing with the socio-emotional

challenges of adolescence thus reflects the high relevance of ToM. Studies investigating ToM performance in adolescence showed that adolescents show better ToM than children (Bosco et al., 2014; Devine & Hughes, 2013; Im-Bolter et al., 2016), but worse than young adults (Dumontheil et al., 2010; Humphrey & Dumontheil, 2016; Symeonidou et al., 2016; Tousignant et al., 2017; Valle et al., 2015).

These age-related improvements appear to vary according to the ToM measure used (Tousignant et al., 2017). One study found age-related improvements in the social-perceptual component (measured with the RMET) and the social-cognitive component (measured with the FPT) of ToM during adolescence (Meinhardt-Injac et al., 2020). Age-related improvement in social-cognitive ToM was accompanied by improvements in logical thinking, inhibition, and language processing. Thus, a specific age-related improvement is only recognizable for the social-perceptual ToM component, measured with the RMET. The finding that the RMET is not invariant to aging effects has been extended by another recent review (Osterhaus & Bosacki, 2022). They showed that performance on the RMET increased during middle childhood and adolescence but decreased in older adults. This fits in with the finding that adolescents have weaker emotion recognition than young adults (Tousignant et al., 2017; Vetter et al., 2013).

Research suggests that children make significant progress in ToM into adolescence, while ToM declines in older adults (Bernstein et al., 2011). A meta-analytic review demonstrated that older adults have poorer ToM than younger adults in all tasks, regardless of verbal or visual tasks (Strange Stories, RMET, Videos, FB, and FPT; Henry et al., 2013).

1.4 Developmental Progression of Theory of Mind Measures

1.4.1 Basic Theory of Mind Measures: First-Order False-Belief Tasks

The first measures used to assess ToM were FB tasks (Wellman et al., 2001; Wimmer & Perner, 1983). These are the fundamental measures that aim to assess a child's ability to understand that others can have beliefs that are different from reality and their own. First-order FB tasks assess participants' understanding that another person may hold FBs about the world around them. Participants must therefore be able to distinguish between the perspective of another person and reality (Lecce et al., 2014; Miller, 2013). It has been demonstrated that children between the ages of four and five are capable of comprehending first-order FBs (Happé, 1995). Among first-

order FB tasks, the three main forms of FB tasks are change-of-location tasks, unexpected-contents tasks, and unexpected-identity tasks (Wellman et al., 2001).

In change-of-location tasks, a hidden object is moved from one location to another in an agent's absence. The most prominent change-of-location task is the chocolate task (Wimmer & Perner, 1983). In this experiment, the character Maxi puts chocolate in cupboard x. While he is away, his mum takes the chocolate from cupboard x to cupboard y. Maxi comes back and wants to eat the chocolate. The test subjects are asked in which cupboard Maxi will look for the chocolate. Only if they can connect Maxi's FB (the chocolate is in x) with what they themselves know (the chocolate is in y), can they correctly point to cupboard x. The subjects are therefore asked about the FB and the real fact. Thus, this experiment tests whether subjects have an explicit and clear representation of the other's FB. Another prominent example of such tasks is the Sally-Anne test (Baron-Cohen et al., 1985).

In the unexpected-content task (Perner et al., 1987), also known as the Smarties task (Davis & Pratt, 1995), the test taker is presented with a container of a familiar candy brand (Smarties) and is asked to guess the contents. However, when they open the box, there is no candy inside. Instead, the box is filled with pencils. Subsequently, the child is posed with two control questions about the actual contents of the box and their previous opinion (FB) of the content of the box. Then, the pencils were returned to the box and the box was closed again. Subsequently, the test taker was informed that it was their friend's turn. The friend was then presented with the closed box as they had originally seen it and asked to state what was in it. Then the test taker is asked what their friend would guess is inside the box. According to Perner et al. (1987), half of the four-year-olds and the majority of the five-year-olds manage to solve this FB task. A variation of the unexpected content task that is very similar is the unexpected identity task. In this task, a FB is elicited by an object that presents a deceptive identity, for example, a sponge that resembles a rock (Buttelmann et al., 2015; Wellman et al., 2001).

In the tasks described, the participants were required to answer questions. Explicit FB comprehension therefore requires verbal articulation skills. At the age of 4 years, children's performance on explicit FB tasks in which they had to articulate themselves verbally was significantly above chance (Wellman et al., 2001). It is therefore possible that children implicitly understand FB earlier than this age but are not able to articulate it beforehand. Clements and Perner (1994) provided the first

evidence that explicit FB understanding may be preceded by implicit knowledge of FB. They performed an anticipatory FB task by extending the famous change-of-location task of Maxi and the chocolate (see Wimmer & Perner, 1983) by recording eye movements. Before asking the typical FB test question, they asked the children to anticipate the protagonist's action by saying, "I wonder where he is going to look." Meanwhile, they recorded the children's anticipatory gaze to see whether it was directed at the believed (false) or true location of the target object. They found that from the age of 2 years and 11 months, around 90% of children looked at the belief-based location. Conversely, only about 45% of children at this age gave the correct verbal response. Thus, implicit measurements of understanding are contrasted with explicit responses to the experimenter's questions because they are independent of language ability (Frith & Frith, 2008; Tager-Flusberg, 2007). This finding was replicated by other studies that also found that typically developing children are able to pass implicit before explicit FB tasks (Ruffman et al., 2001). Implicit abilities to reason from the mental state are not only present in infants and young children but also in older children and adults. One study used anticipatory-looking study tasks to investigate implicit ToM ability in older children and adults and found that adults looked longer at the place without an object when they held FB than when they held true beliefs, even though they also had a distraction task (Schneider et al., 2012). Thus, older children and adults also have implicit abilities to deal with FB that appear to function unconsciously.

1.4.2 Advanced Theory of Mind Measures

1.4.2.1 Second-Order False-Belief Tasks. At 4 years of age, children typically pass these basic ToM tasks (Wellman et al., 2001) and around the age of six, develop more advanced ToM (AToM) abilities, involving the interpretation of more subtle mental states, emotions, and social cues (Perner & Wimmer, 1985). AToM refers to the age-related progressions in the ability to infer mental states that occur after preschool (Hughes & Devine, 2015). To test these, researchers have attempted to develop more AToM tests that extend the understanding of ToM beyond first-order FB tasks. At the age of seven or eight years, children begin to understand more-complex forms of FB, such as second- or higher-order FB (Miller, 2009). While first-order ToM tasks require participants to take the perspective of another person, second-order ToM tasks require the participants to take the perspective of a character who takes the perspective of another character (Longobardi et al., 2019). In other words, second-order ToM ability

means attributing a mental state to a person who attributes a different mental state (e.g., A thinks that B thinks that x…). This ability requires the realization that mental states can be recursive (Osterhaus & Koerber, 2021).

Second-order FB tasks were first introduced by Perner and Wimmer (1985). An example of a second-order FB task is the following scenario: a mother falsely tells her son, Peter, that she has bought him a toy for his birthday. However, the genuine gift was a puppy (Lecce et al., 2014). If the previous scenario is continued, Peter finds the birthday puppy without his mother knowing. Peter's grandmother then asks Peter's mother what Peter will get for his birthday. An example of a second-order FB question is "What does the mother say to the grandmother?". To test understanding of the story, two additional control questions were posed: "Did the mother see Peter find the present?" and "What did Peter say to the grandmother?". According to Perner and Wimmer (1985), second-order FB tasks would be more challenging than first-order FB tasks. Empirical evidence supported their hypothesis. It was not until the age of 7 or 8 that the majority of children were able to assign a second-order FB. Subsequent research has also demonstrated that second-order FB tasks necessitate the development of skills that extend beyond those required for success in first-order FB tasks (Miller, 2009). What first-order and second-order FB tasks have in common is that test takers are required to form beliefs about a certain situation that differ from the protagonists in the stories (Hollebrandse et al., 2014). In second-order FB tasks, test takers additionally have to know the belief one protagonist has about the belief of the other protagonist and how these two beliefs differ from each other.

The focus of FB tasks is on making predictions, understanding beliefs, or predicting actions (e.g., Hughes et al., 2000; Wellman et al., 2001). However, the results of a single task may be inconclusive and insufficient for understanding the level of ToM ability. Furthermore, it should be noted that FB understanding is just one of several aspects of ToM that emerge during the developmental period. It is also important to consider other aspects, such as understanding that people can have different desires about the same objects (diverse desires), understanding that people can have different thoughts about the same object (diverse beliefs), understanding that people can have limited access to knowledge (knowledge access), understanding that internal emotions of people can be different than those shown outside (hidden emotions), and understanding ambiguous situations, such as irony and idioms (Peterson et al., 2005; Wellman & Liu, 2004). Furthermore, many FB tasks have been

found to have inconsistent reliability (Hughes et al., 2000; Mayes et al., 1996). There are also some concerns regarding the ecological validity of these measures (Peskin & Ardino, 2003; Ronald et al., 2005). These issues highlight the need to find measures that address the limitations described above.

A plethora of other tasks have been developed to measure AToM. AToM tests not only consist of social stories that require a form of second-order reasoning (Miller, 2009). Some tasks assess the ability to interpret complex social situations, such as the recognition of a faux-pas (Osterhaus & Koerber, 2021) or the ability to draw conclusions from non-verbal cues (Osterhaus et al., 2016; Warnell & Redcay, 2019). To present an overview of these measures and their frequency of use, Osterhaus and Bosacki (2022) recently conducted a systematic review and meta-analysis. The authors identified five main definitional operationalizations, which are also presented in Table 1. The most common operationalization of AToM is the recognition of nonliteral speech, such as irony, jokes, or sarcasm, followed by drawing inferences from nonverbal cues (e.g., the eyes), higher-order FB understanding, the interpretation of social situations, and the recognition of faux-pas. For each main operationalization, the most frequently used test was identified. Five tasks accounted for more than 60% of ToM measures used and include second-order FB tasks (Perner & Wimmer, 1985), Strange Stories (Happé, 1994), the RMET (Baron-Cohen et al., 2001), the FPT (Stone et al., 1998), and the Frith-Happé Triangles test (Abell et al., 2000). The Triangles test is, however, subsumed under the operationalization of the ascription of mental states to non-human, inanimate objects. This operationalization occurs less frequently than the interpretation of social situations and given my focus on the frequency of operationalizations, I have included the MASC (MASC; Dziobek et al., 2006) in Table 1, which is also commonly used to test ToM in clinical samples.

The different conceptualizations raise the question of whether the different AToM tests measure the same underlying ability, implying that they are interchangeable measures of the same construct. Previous work (Hayward & Homer, 2017; Warnell & Redcay, 2019) suggests that correlations between distinct measures are low, which may be interpreted as evidence suggesting minimal coherence. This reflects one major challenge in the ToM assessment process, the existence of terms with similar meanings and disparate measurements for identical terms (Olderbak & Wilhelm, 2020). Olderbak and Wilhelm (2020) refer to this phenomenon as the 'jingle and jangle fallacies'. According to them, jingle fallacies occur when measurements are presumed

to assess the same construct, yet in practice, they assess different constructs. Jangle fallacies occur when measurements are assumed to measure different constructs, yet in fact measure the same construct. Many constructs are not uniformly described in the literature (e.g., ToM, perspective taking, mentalizing, mindreading). These unclear definitions lead to considerable differences in the measurement of the concepts, which is accompanied by a lower validity of the measurement methods (Olderbak & Wilhelm, 2020; Quesque & Rossetti, 2020). Despite these issues, there is however consensus about the core components of ToM: being able to represent another person's mental state and distinguish it from one's own (Quesque & Rossetti, 2020). Due to these problems, Quesque and Rossetti (2020) established two criteria to qualify as a ToM test that are among the core components of ToM. First, a task must require subjects to represent the mental states of others, and success on the task must not be due to lower-level processes such as attention orientation or associative learning (mentalizing criterion). Secondly, subjects must be able to distinguish between their own mental states and those of others (non-merging criterion). The assessment of whether tasks fulfill these criteria is presented in Table 1.

Notably, ToM measures load on different factors, as ToM is a multifaceted skill. To identify the underlying factors of different ToM tasks (e.g., RMET, FPT), Osterhaus et al. (2016) analyzed items from seven ToM tasks. The researchers employed Rasch scaling and factor analyses to ascertain that there are three distinct ToM factors: 1) social reasoning, 2) reasoning about ambiguity, and 3) recognizing transgressions of social norms. Table 1 provides an overview of the factors responsible for the respective ToM test loads. For the MASC, such a factor analysis is not yet available. It is therefore unclear which aspects the MASC measures and whether they differ from the FPT.

Table 1

Overview of the 5 Most Relevant AToM Task Operationalizations

	Operationalization	Task	Factor	Utility	MC	NC
1)	Recognition of nonliteral speech (irony, jokes, sarcasm)	Strange Stories (Happé, 1994)	social reasoning	PP	Yes	Yes
2)	Drawing of inferences from nonverbal cues (eyes)	Reading the Mind in the Eyes Test (RMET)	social reasoning	PP	No	No
3)	Higher-order FB reasoning	Second-Order FB Tasks	social reasoning	PP	Yes	Yes
4)	Interpretation of social situations	Movie for the Assessment of Social Cognition (MASC)	(not yet investigated)	Videos (monitors and speakers)	Yes	Yes
5)	Recognition of faux-pas	Faux-Pas Recognition Test	recognizing transgressions of social norms	PP	Yes	Yes

Note. MC = mentalizing criterion; NC = non-merging criterion.

1.4.2.2 The Strange Stories Task. The most recent systematic review and meta-analysis (Osterhaus & Bosacki, 2022) indicates that the Strange Stories task (Happé, 1994) is the most commonly used measure of higher-order mental state comprehension. It particularly assesses the ability to understand nonliteral speech such as irony, sarcasm, and double bluffs. The task was created to capture subtle social-cognitive difficulties in adults with autism spectrum disorders (ASD; Murray et al., 2017).

The story-based task comprises 24 short vignettes, each accompanied by a picture and two test questions, one comprehension question ("Was it true, what X said?") and one justification question ("Why did X say that?"). Each vignette is about a protagonist who does something unexpected. The Strange Stories task consists of 12 different story types, with two examples of each. The 12 story types were as follows: lie, white lie, joke, pretend, misunderstanding, persuade, appearance/reality, figure of speech, sarcasm, forget, double bluff, and contrary emotions. Additionally, a set of six control stories (i.e., physical stories) was provided to the subjects. These control stories

did not address mental states; instead, they described an unforeseen outcome with a mechanical physical cause (e.g., a power cut causing a meal to be undercooked). In the Strange Stories test, participants are presented with brief scenarios about a protagonist who takes an unanticipated action. To illustrate, a child might offer a compliment about her aunt's appearance in an unfortunate choice of headwear (white lie) or suggest to a friend that she will forgo her own dinner to accommodate their request for lunch (persuasion). In this task, participants are required to provide an explanation for such statements. In each trial, participants are tasked with evaluating whether they can interpret the statements based on the contrast between what is expressed and what is intended. The Strange Stories task is administered using a paper-and-pencil format with open-ended answers (Happé et al., 1998). The responses were subsequently evaluated according to a pre-established scoring system, with a second rater providing consistent results despite being unaware of the participant group and hypothesis. Answers were scored on a scale of 0 to 2, with 2 being awarded for a comprehensive and explicitly correct response and 1 for a partial or implicit answer. In the case of the jumbled passages, the test questions required a simple yes or no answer, which were scored 0 (incorrect or don't know) or 1 (correct). The Strange Stories task is considered as both valid and reliable (Devine & Hughes, 2016). Moreover, research has consistently demonstrated its utility in distinguishingindividuls ith ASD from neurotypical controls (Murray et al., 2017; Velloso et al., 2013). Studies have shown its effectiveness in differentiating individuals with autism spectrum disorders (ASD) from neurotypical controls (Velloso et al., 2013; Murray et al., 2017). Specifically, children with ASD exhibit significantly lower performance on this task compared to their typically developing peers (Freed et al., 2015; Velloso et al., 2013). Notably, gender differences have been observed, with girls outperforming boys (Devine & Hughes, 2013).

1.4.2.3 The Faux-pas Recognition Test (FPT). A very common measure to assess the ability to interpret complex social situations is the FPT (Stone et al., 1998). According to a recent systematic review and meta-analysis on an overview of ToM measures, the FPT is one of the five tests that account for more than 60% of ToM measurements used (Osterhaus & Bosacki, 2022). The extensive number of translations provides evidence of the FPT's popularity (see *Autism Research Center*, 2023). It was initially developed for clinical trials in autism research and was then slightly modified for use with brain-injured adults (Stone et al., 1998b) and patients with

dementia (Gregory et al., 2002). The test instrument is capable of reliably capturing both developmental changes and interindividual differences. Recently, the FPT demonstrated good psychometric properties for faux-pas stories (α = .94) and for control stories (α = .92; Şandor & İşcen, 2021). The FPT is a measure that presents social blunders (= faux-pas) embedded in stories, labeled as vignettes (Stone et al., 1998). A faux-pas occurs when an individual says something without considering whether the listener might not want to hear or know it. The individual is usually unaware of the mistake, which then has negative and unintended consequences (Baron-Cohen et al., 1999). The FPT is available in a number of different versions. Originally, the FPT was developed with 10 vignettes that presented a faux-pas (targets; Baron-Cohen et al., 1999; Stone et al., 1998). Later, 10 vignettes without faux-pas (foils) were added to control for response biases (Baron-Cohen et al., 1999). Test takers are instructed to detect the vignettes that contain a faux-pas. If so, they are then required to answer up to six ToM-related questions and two questions assessing basic comprehension of the scenario (control questions). The original six items included a 1) detection question (whether a faux-pas occurred), 2) person identification question (who made the inappropriate statement), two FB questions about 3) why somebody should not have said/done something and 4) why they did it anyway, 5) a question about whether the person knew they committed a faux-pas, and 6) affective question about how the victim felt. The original authors (Stone et al., 1998) revised the analysis procedure and made it available online. However, no validation was made available for publication.

Recognizing a faux-pas is one of the more demanding cognitive skills associated with ToM. To identify a faux-pas, it is essential for test takers to represent the mental states of both protagonists, comprehend the emotional impact of the statement, and be aware of the social norms that were sometimes violated in these situations. As mentioned above, the FPT meets the two criteria that a ToM task should fulfill: test takers are required to represent the mental states of others (mentalizing criterion), and test takers are required to differentiate between their own and the other's mental states (nonmerging criterion; Quesque & Rossetti, 2020). The identification of faux-pas encompasses both the cognitive and the affective components of ToM. Understanding the difference between the speaker's and listener's level of knowledge (i.e., understanding that the person making the inappropriate comment does not know that they should not say it) comprises the cognitive ToM. Understanding the listener's emotional state (i.e., realizing that the person hearing the comment would likely feel

offended) comprises the affective component (Baron-Cohen et al., 1999; Stone et al., 1998b).

As previously reported, age affects faux-pas understanding in children. Before the age of 10, children's ability to detect faux-pas is unstable, with girls apparently being able to recognize faux-pas slightly earlier than boys (Baron-Cohen et al., 1999). A longitudinal study found that difficulties in understanding faux-pas predicted increased peer rejection, while peer rejection may impair understanding of faux-pas (Luke & Banerjee, 2013). It has been shown that children's ability to recognize a faux-pas depends on the extent to which they acquire the basic AToM concepts (i.e., understanding that someone can hold a FB about a belief of someone else). The age at which children understand AToM tasks, such as recognizing a faux-pas, is influenced by information-processing skills such as language and intelligence (Osterhaus & Koerber, 2021). From the age of 20 onwards, the FPT appears to be relatively independent of the effects of aging, and may therefore be less susceptible to changes in cognitive function in late adulthood (Cavallini et al., 2013).

1.4.2.4 The Reading the Mind in the Eyes Test (RMET). A very frequently used instrument for measuring the ability to draw conclusions from non-verbal cues is the RMET (Baron-Cohen et al., 2001). It was originally developed for use in clinical populations. According to a recent review and meta-analysis, it accounts for more than 60% of the most frequently-used AToM tests (Osterhaus & Bosacki, 2022). The RMET is often used in clinical contexts (Bora & Berk, 2016; Nestor et al., 2022; Olderbak et al., 2015; Richman & Unoka, 2015) due to its straightforward implementation. The RMET requires the participants to demonstrate their ability to understand mental state terms and match them to the eye region of faces (Pfaltz et al., 2013). The capacity to discern emotions in the facial expressions of others is a basic skill that is acquired during childhood and continues throughout life. This ability is a fundamental aspect of social communication, as it allows us to assess the emotions and mental state of another person. To measure this ability, Baron-Cohen et al. (1997) first developed the RMET. In the original version of the RMET, 25 pictures of the eye region were presented and subjects were required to select two words that best described the thoughts and feelings of the depicted individual (Baron-Cohen et al., 1997). Subsequently, an elaborated version was published exhibiting enhanced psychometric properties (Baron-Cohen et al., 2001). This version contains 36 items (pictures of the eye region), in which the subjects have to select one of four terms for the mental state.

The pictures represent either emotionally charged (positive vs. negative) or neutral mental states. The pictures are gender balanced. Internal consistency was acceptable (α = .70; Charernboon & Lerthattasilp, 2017). A German translation is also available (Pfaltz et al., 2013).

In the RMET, the test subject must have a lexicon of mental states and know the semantics of these terms, such as suspicious or annoyed. The instrument then involves assigning these terms to fragments of facial expressions of mental states - only the part of the face around the eyes. On a supposedly unconscious, rapid, and automatic level, subjects must compare the eyes in each picture with examples of eye expressions stored in memory and seen in the context of certain mental states to arrive at a judgment about which word the eyes most closely match. It is assumed that people sort and assign the eyes to the faces stored in their memory on a quick and unconscious level and make corresponding judgments about the mental state of the faces (Baron-Cohen et al., 2001).

The RMET is not invariant to aging effects. In fact, results of a recent review and meta-analysis show that performance on the RMET increased during middle childhood and adolescence, but decreased in older adults (Osterhaus & Bosacki, 2022). Osterhaus et al. (2016) conducted a factor analysis of items from seven ToM tasks in order to identify the underlying factors. They concluded that there are three distinct ToM factors: 1) social reasoning, 2) reasoning about ambiguity, and 3) recognizing transgressions of social norms. The RMET primarily measures social reasoning. Regarding the validity of the RMET, the test is not homogenous as it does not have a single-factor solution (Olderbak et al., 2015). An alternative three-factor solution, that distinguishes the items based on valence (8 positive items, 12 negative items, 16 neutral items) was unsatisfactory (Vellante et al., 2013). In the RMET, subjects are only required to decode, but not to reason about the mental states of the protagonists (Maleki et al., 2020).

The RMET does not capture either of the two criteria (mentalizing and non-merging criterion) to qualify as a ToM measure, as proposed by Quesque and Rossetti (2020). According to their argument, the RMET rather measures lower-level processes, such as emotion recognition or attention orientation, rather than genuine ToM-abilities. Although ToM is typically associated with the ability to perceive emotions on the face, a recent meta-analysis found that performance on the RMET test was strongly related to emotion perception and only moderately related to the performance of participants

on other ToM tasks (Kittel et al., 2022). The results of this meta-analysis suggest that the construct validity of the RMET is questionable, as it shares only 15% of its variance with other ToM measurements.

1.4.2.5 The Movie for the Assessment of Social Cognition (MASC). To measure the interpretation of social situations, especially in clinical samples, the Movie for the Assessment of Social Cognition (MASC; Dziobek et al., 2006) is a very prominent ToM measure. As the FPT, the MASC meets the two mentalizing and non-merging criteria a ToM task should fulfill (Quesque & Rossetti, 2020). It is a reliable and valid ToM measurement (Fossati et al., 2018) that demonstrated acceptable internal consistency (Cronbach's α = .70 - .78).

The aim of the MASC is to operationalize ToM through videos in which social interactions as they take place in everyday life are shown (Dziobek et al., 2006). It presents a 15-minute film about four people meeting for a dinner party. The four characters have different motives for attending this evening of cooking, dining, and playing a board game. Throughout the film, friendship and dating issues are the predominant themes, as each character develops their own dynamics with each of the other characters. To vary the amount of intimacy in their interactions, the relationships between the characters differed between those of either strangers or friends. Throughout the film, they experience different situations that elicit different mental states and emotions. The movie is divided into 46 segments, which represent the test items. Each video sequence is followed by a question on the three different mental states' modalities. In sum, test takers are required to answer 45 mentalizing criterion (MC) questions about the mental states' modalities of the four film characters. The different mental state modalities are emotions, thoughts, and intentions that vary in positive, negative, and neutral valence. In addition to the mental state modalities, the conversational content of the items captured both verbal (19 items) and non-verbal (16 items) communication, such as the recognition of facial expressions or the interpretation of body language and gestures.

Regarding the evaluation of the different mental state modalities of the MASC, it uses 15 items to measure the interpretation of emotions, 14 items to measure intensions, and 4 items to measure thoughts (Preißler et al., 2010). Various concepts of social cognition are adopted in the MASC: first- and second-order FB, deception, faux-pas, persuasion, metaphor, sarcasm, or irony. In addition, a multidimensional approach was implemented for the first time: the subdivisions of the different mental

state modalities, valence, and degree of linguistic involvement in the social cognition process could be adequately captured (Dziobek et al., 2006).

Following previous MASC investigations, the mental state modalities intentions and thoughts are combined into 18 items that measure the cognitive ToM and the 15 items that measure the interpretation of emotions focus on measuring the affective ToM (Buhlmann et al., 2015; Knopp et al., 2024). Thus, the MASC provides a total score for all 45 mental state decoding items and further allows to differentiate between the cognitive and the affective ToM subscales. The remaining 12 items do not differentiate between cognitive and affective ToM. Additionally, participants are assigned to answer 6 control questions to examine non-social inferencing and control for memory and general comprehension effects (Dziobek et al., 2006).

All items of the MASC are in an MC format. Each item has four possible answer categories that differentiate between right/wrong responses (one correct response out of four possible responses). Specifically, the MC responses differentiate between three different types of ToM errors: 1) non-mental state inferences such as physical states, referred to as no ToM, 2) insufficient mental state inferences—referred to as less ToM, and 3) too excessive mental-state inferences—referred to as exceeding ToM (Dziobek et al., 2006). Nevertheless, the validation of these error types is still pending.

1.5   Theory of Mind and Mental Disorders

Mental disorders are one of the main causes of illness-related stress worldwide, according to the latest Global Burden of Disease Study (GBD) from 2019 (Murray, 2022). They are one of the major public health problems and have serious consequences, such as high school dropout rates, low economic productivity, imprisonment, suicide, and homelessness (Erskine et al., 2015). Using data from the 2019 GBD study, one recent study demonstrated that the average prevalence for mental disorders was 11.63%, with peaks between the ages of 5 and 9, and 20 and 24 (Kieling et al., 2024). This is relevant because AToM ability develops nonlinearly between the ages of 5 and 10 (Osterhaus & Koerber, 2021) and people show age-related ToM improvements from adolescence to young adulthood (18-24 years; Spenser et al., 2020).

The first studies dealing with ToM deficits in mental disorders were conducted on ASD (Baron-Cohen et al., 1985; Happé & Frith, 1996). They found that ToM impairments are a key feature of ASD. Subsequently, ToM deficits have been identified as core cognitive deficits in schizophrenia (Chung et al., 2014; Savla et al., 2013).

Soon, studies of social cognitive deficits have become an emerging area of interest for a number of other clinical groups. Disorders that have been repeatedly associated with ToM deficits include MDD (Berecz et al., 2016; Bora & Berk, 2016; Nestor et al., 2022; Pagnoni et al., 2022; Richman & Unoka, 2015) post-traumatic stress disorder (PTSD; Couette et al., 2020; Janssen et al., 2022), and BPD (Bora, 2021; Németh et al., 2018; Richman & Unoka, 2015). This is in line with a review of meta-analyses by Cotter et al. (2018) which suggested that social-cognitive deficits are a transdiagnostic clinical marker of various clinical conditions.

The prevalence of these impairments and their relevance to symptoms was the reason that ToM was included in the RDoC framework. The RDoC framework is an alternative approach to understanding mental disorders (Michelini et al., 2021; National Institutes of Mental Health, 2020). Within the RDoC framework, ToM was proposed as a transdiagnostic factor underlying mental disorders. Another dimensional approach is provided by the Hierarchical Taxonomy of Psychopathology (HiTOP) framework (Conway et al., 2019; Michelini et al., 2021). The two frameworks were introduced with the aim of overcoming the limitations of categorical diagnostic systems as previously used in the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) and ICD. The validity of categorical diagnoses was questioned due to the high prevalence of comorbidities and diagnostic heterogeneity within the same diagnosis. In addition, the diagnoses were found to have low symptom specificity and low reliability (Conway et al., 2019). Since, the disorders ASD, schizophrenia, BPD, PTSD, and MDD have repeatedly shown ToM deficits in the literature, they are discussed below.

1.5.1 Theory of Mind and Autism Spectrum Disorder

ASD, encompassing autism (including both low-functioning autism and high-functioning autism, HF) and Asperger syndrome, are neurodevelopmental disorders (DSM-5; American Psychiatric Association, 2013). The impairments typically manifest before the age of three years. The core diagnostic criteria for ASD involve enduring challenges in social communication and interaction across multiple contexts. The deficits in social interactions may result from interpretation or identification difficulties of verbal and non-verbal social stimuli of other individuals. Understanding non-verbal social information, such as emotions and mental states from the other's perspective is crucial for ToM development (Baron-Cohen et al., 1995; Mundy et al., 1986). Therefore, impairments in social communication and social interaction haven often been attributed

to a core ToM deficit (Baron-Cohen et al., 1985, 2001). Baron-Cohen (2009) suggested that the social communication deficits in ASD stem from a delay in the development of the empathy system (i.e., the ability to recognize mental states in others and elicit appropriate emotional responses). The early ability to understand and infer emotions and mental states predicts later AToM abilities (Harwood & Farrar, 2006; O'Brien et al., 2011). Thus, early disruptions in ToM development may contribute to a cascade of later AToM deficits in people with ASD (Baron-Cohen et al., 1995).

While ToM impairments have been consistently reported in individuals with ASD (Chung et al., 2014; Kimhi, 2014; Mathersul et al., 2013; Scheeren et al., 2013), studies examining ToM abilities in individuals with HFA (i.e., normally intelligent) have yielded mixed results. Scheeren et al. (2013) found no differences in performance on ToM tasks in school-aged children and adolescents with HFA compared to typically developing peers. Conversely, Mathersul et al. (2013) found significant ToM deficits even in HFA compared to nonclinical controls.

ToM impairments in people with ASD differ in ToM tasks. A plethora of instruments have been developed for the purpose of measuring ToM, or certain aspects thereof, in individuals diagnosed with ASD. The initial research in this field was influenced by studies examining ToM in young children with typical development. These studies employed variants of the FB tasks that have been widely used in mainstream developmental research (Baillargeon et al., 2010; Baron-Cohen et al., 1985). Research indicates that young children with ASD are less successful in first-order FB tasks than their typically developing peers (Baron-Cohen et al., 1985; Wellman et al., 2001). However, research on explicit and implicit FB tasks in children with ASD shows a more complex picture. While individuals with ASD often perform similarly to neurotypical controls on explicit FB tasks, they show deficits in implicit FB understanding, such as in anticipatory-looking tasks (Schuwerk et al., 2015, 2016). This suggests that social cognition alterations in ASD may be primarily implicit in nature (Callenmark et al., 2014). Interestingly, experience can influence implicit FB performance in individuals with ASD (Schuwerk et al., 2015). In contrast to typically developing children, who pass implicit before explicit FB tasks, children with ASD tend to pass explicit before implicit FB tasks (Ruffman et al., 2001). Explicit FB understanding is associated with linguistic abilities and executive functions, while implicit FB tasks are not (Wiesmann et al., 2017). This could explain the better abilities of children with ASD in explicit FB tasks: some children with HFA have developed a

way of dealing with FB tasks (Peterson et al., 2005) without viewing them as FBs. These children can treat standard FB tasks as logical reasoning problems, relying primarily on language and other non-social cognitive processes. However, they cannot use this strategy when solving implicit FB tasks, which are assumed to be important for social functioning, i.e., for prosocial behavior (Frith & Frith, 2008; Tager-Flusberg, 2007). This could also explain why children and adolescents with HFA are theoretically able to master AToM tasks, but still fail to apply these principles in everyday life (Scheeren et al., 2013).

The impairment of AToM in people with ASD was also analyzed using various AToM tasks. One meta-analysis compared verbal (i.e., intention/belief inferences) and visual (i.e., emotion recognition) AToM tasks in patients with ASD (Chung et al., 2014). The verbal tasks included the Strange Stories (Happé, 1994), the FPT (Baron-Cohen et al., 1999), and the visual task consisted of the RMET (Baron-Cohen et al., 2001). They found that individuals with ASD showed ToM deficits in both verbal and visual tasks.

Although many studies using various ToM instruments have already investigated ToM impairments in individuals with ASD, Brewer et al. (2017) nevertheless criticize the lack of empirical data on the extent and variability of ToM performance in individuals diagnosed with ASD. They point out the limitations of existing ToM measurement instruments, such as the non-standardized application of the instruments and the large number of test variants.

1.5.2 Theory of Mind and Schizophrenia

Schizophrenia is a persistent and severe mental disorder that is characterized by disturbances in thought processes, perception, emotional reactivity, and social interactions (DSM-5; American Psychiatric Association, 2013). Individuals with schizophrenia typically show the following symptoms: positive symptoms, such as delusions, hallucinations, disorganized speech, and highly disorganized or catatonic behavior, and negative symptoms, such as reduced emotional expression or avoidance. In addition to these symptoms, the DSM-5 also postulates that some individuals may have deficits in social cognition, such as ToM. For example, individuals with schizophrenia may perceive irrelevant events or stimuli and interpret them as meaningful, possibly leading to the development of explanatory delusions. According to the DSM-5, these impairments frequently persist during symptomatic remission.

This is also confirmed by studies that showed that the extent of ToM impairment is related to psychotic symptoms, as remitted patients showed lower (but still significant) ToM deficits (Balogh et al., 2014; Bora et al., 2009; Sprong et al., 2007). Conversely, ToM deficits may be considered a risk factor for the transition to psychosis and are associated with developing psychotic disorders (Van Donkersgoed et al., 2015).

ToM deficits have also been reported by several meta-analyses (Bora et al., 2009; Bora & Pantelis, 2013; Chung et al., 2014; Fett, 2011; Savla et al., 2013; Sprong et al., 2007). The respective meta-analyses used different ToM tasks to assess ToM and its different aspects. It is therefore very important to have reliable and validated tools for the assessment of ToM. The main challenges in investigating ToM in schizophrenia are the large number of different ToM tests and the lack of standardized measures. The literature shows that people with schizophrenia are able to represent basic cognitive mental states. However, when they are confronted with more complex tasks, such as tasks that require making inferences of affective mental states or the integration of both cognitive and affective mental states, their ToM abilities appear to fail (Abu-Akel & Shamay-Tsoory, 2013). It is therefore necessary to investigate ToM with an instrument that captures more complex ToM and includes both affective and cognitive ToM. One commonly used test is the FPT (Stone et al., 1998), which captures both cognitive and affective ToM.

Several meta-analyses included the FPT (Bora et al., 2009; Bora & Pantelis, 2013; Chung et al., 2014; Fett, 2011; Savla et al., 2013), except for Sprong et al. (2007). Savla et al. (2013) found that individuals with schizophrenia performed poorly on all ToM tasks, including the FPT, the MASC, and the RMET. The systematic review and meta-analysis by Bora and Pantelis (2013) found significant ToM deficits in individuals with a first episode of psychosis, individuals at extremely high risk of psychosis, and first-degree relatives with schizophrenia compared to healthy controls (HC). Relatives of patients showed a less severe deficit, with performance levels between those of HC and individuals with the first episode of psychosis. ToM measures included in the analysis were verbal and visual ToM tasks, such as the FPT, FB, Strange Stories, the RMET, or the MASC. Sprong et al. (2007), who used several ToM tasks such as the RMET and FB tasks, found significant ToM deficits with large effect sizes across all symptom subgroups, including remitted patients. The impairments were not moderated by IQ, gender, or age. Conversely, the meta-analysis conducted by Bora et al. (2009),

which found large ToM deficits in schizophrenic patients across all ToM tasks (including the FPT, the RMET, the MASC, and FB tasks), demonstrated a significant impact of IQ impairment on ToM deficits when the analysis was limited to remitted patients. Therefore, they suggest that a general and selective cognitive deficit in individuals with schizophrenia may contribute to ToM impairments in symptom-free patients. Chung et al. (2014) compared ToM in verbal (i.e., the Strange Stories and the FPT) and visual (i.e., the RMET) ToM tasks. They found that patients with schizophrenia showed greater impairments in verbal than in visual ToM tasks.

Further research on ToM deficits in individuals with schizophrenia investigated with the FPT shows that patients with schizophrenia and control subjects do not differ in the understanding of control stories or general story comprehension (of both the FP stories and the control stories). This indicates that individuals with schizophrenia do not show any problems in understanding the content of the stories. However, patients with schizophrenia showed impairments in the understanding of a faux-pas. Most studies found differences in understanding inappropriateness and understanding intentions (e.g., Hasson-Ohayon et al., 2015; Pijnenborg et al., 2013; Shur et al., 2008). Thus, the results show that individuals with schizophrenia have trouble understanding other people's emotions and mental states.

The positive and negative symptoms of schizophrenia were found to correspond to different ToM impairments. While those exhibiting positive symptoms displayed a tendency to overinterpret mental states, those exhibiting negative symptoms displayed an impaired ability to interpret the mental states of others (Montag et al., 2012; Peyroux et al., 2019).

The question of whether a ToM deficit can be considered a stable trait in schizophrenia is currently a matter of debate. Some researchers propose that it is a clinical characteristic that depends on the presence and severity of the symptoms (Balogh et al., 2014; Pickup & Frith, 2001), whereas most argue that it should be regarded as a stable trait (e.g., Bora et al., 2009; Green et al., 2012; Martino et al., 2007; Sprong et al., 2007, 2007). In the study by Fernández-Modamio et al. (2018), patients with minimal levels of symptom severity also exhibited ToM deficits, which supports the proposition that ToM impairments may be regarded as a stable trait. This result was also confirmed by the meta-analyses of Bora et al. (2009) and Sprong et al. (2007). Both analyses revealed ToM deficits in remitted patients and thus support the assumption that there are trait-related ToM deficits in schizophrenia.

1.5.3 Theory of Mind and Major Depressive Disorder

MDD is the most widespread mood disorder (Baxter et al., 2013). Besides its high prevalence, it can manifest in an episodic or recurrent manner (DSM-5; American Psychiatric Association, 2013). MDD is characterized by the presence of persistent low mood, loss of pleasure and interest, impaired cognitive function, and vegetative symptoms, including disturbed sleep or appetite (Otte et al., 2016). The proportion of heredity is estimated at around 35%.

In addition to impaired cognitive function, several meta-analyses (Bora & Berk, 2016; Nestor et al., 2022; Richman & Unoka, 2015) have shown that ToM is impaired in MDD. Two meta-analyses that included studies that employed various ToM tasks, such as the FPT, the RMET, and the MASC to measure ToM, reported moderate ToM impairments (Bora & Berk, 2016; Nestor et al., 2022). This contradicts the findings of large effects of the meta-analysis conducted by Richman and Unoka (2015), which only included one measure (i.e., RMET). Hence, the choice of the ToM measure has a significant influence on the results. Bora and Berk (2016) and Nestor et al. (2022) included several ToM measures, including verbal and visual tasks, in their analyses. They found that the severity of depressive symptoms is significantly related to ToM impairments. In addition to that, both meta-analyses found ToM deficits in cognitive and affective facets and in both decoding and reasoning process domains. These findings were evident in both verbal and visual tasks. However, Nestor et al. (2022) found no significant differences between effect sizes between affective and cognitive facets or between decoding and reasoning domains.

Although research suggests a clear correlation between ToM and depressive symptoms in clinical samples, this is not the case in community samples (Nestor et al., 2022). This is consistent with other findings showing that the severity of depressive symptoms is significantly related to ToM impairments. The systematic-review by Berecz et al. (2016) included acutely depressed as well as mild and unspecified depressed to HC. They found that those with mild and unspecified depression did not significantly underperform HC or differ significantly on ToM tasks. They also found that depressed individuals performed comparably to healthy controls on cognitive ToM tasks relying on reasoning abilities (such as tasks that measure faux-pas detection or understanding the characters' intentions). Hence, mixed findings emerge depending on the ToM measurement. It is possible that the heterogeneity of the ToM

measurements and the heterogeneity of the forms of depression may contribute to the inconsistencies in results (Quesque & Rossetti, 2020).

Symptoms of depression correspond to different ToM deficits. Some studies show that symptoms of depression are associated with reduced ToM (De Coninck et al., 2021; Safiye et al., 2023; Scandurra et al., 2020; Wolkenstein et al., 2011), while others suggest that MDD is related to both reduced and exceeding ToM (McLaren et al., 2022).

A number of factors may be responsible for the observed impairment in ToM in individuals with MDD. First, individuals with MDD frequently disengage from social interactions with others, whether due to low mood, fatigue, or other depressive symptoms (Porcelli et al., 2019). This may, in turn, result in a reduction in the frequency of interpersonal interactions. Consequently, individuals with MDD may become less proficient in identifying the thoughts and feelings of others due to a lack of practice. Nestor et al. (2022) suggest that engagement in social interactions may improve the ability to understand the emotional and mental states of others.

Secondly, MDD is often characterized by a tendency towards negative self-focus or self-directed attention (Mor & Winquist, 2002). This tendency towards introspection and negative rumination about one's own thoughts and feelings can reduce or impair empathy and involvement in understanding the mental and emotional states of others (Nestor et al., 2022). The generally negative perspective of individuals with major depression can lead to misinterpretations and misidentifications of the cognitive and affective states of others (Nestor et al., 2022). The predominantly negative outlook of individuals with MDD can contribute to misinterpretations and inaccuracies in recognizing others' cognitive and emotional states. For instance, individuals with MDD may misjudge the mental states of others due to their tendency to view others through a lens characterized by negativity.

1.5.4 Theory of Mind and Borderline Personality Disorder

The characteristics of a general personality disorder are persistent and maladaptive patterns of cognition, affectivity, interpersonal functioning, and impulse control (DSM-5; American Psychiatric Association, 2013). This pattern of inner experience and behavior significantly deviates from the expectations of one's own culture, is omnipresent and inflexible, and remains consistent over time, leading to suffering or impairment. It typically occurs in adolescence or early adulthood but cannot be diagnosed before the age of 18. Personality disorders differ from personality traits

in that they are inflexible and maladaptive and cause considerable functional impairment or subjective stress. In accordance with the DSM-5, there are various different personality disorders. One subtype that has been extensively studied regarding ToM impairments is BPD. This disorder is marked by instability in interpersonal relationships, self-perception, and emotions, along with significant impulsivity. Furthermore, patients with BPD frequently experience difficulties in emotion regulation, engage in self-injury, and exhibit chronic suicidal tendencies. Genetic predispositions and childhood trauma, including physical and sexual abuse, may contribute to the onset and progression of BPD (Lieb et al., 2004). Pronounced difficulties in social behavior (Winsper et al., 2015) and the accompanying unstable relationships (Gunderson, 2007) are key features of BPD across the entire lifespan.

A growing body of research indicates that patients with BPD have social-cognitive deficits in general (Herpertz & Bertsch, 2014; Roepke et al., 2013). Deficits in the subcomponent ToM in particular have been repeatedly observed in patients with BPD (Bora, 2021; Németh et al., 2018; Richman & Unoka, 2015). One reason for the inconsistent results is that different ToM measurement instruments were used, which had an influence on the results (Quesque & Rossetti, 2020). The discrepancies are discussed below.

Research suggests that patients with BPD have a basic understanding of ToM, measured by the RMET (Baez et al., 2015; Németh et al., 2018; Pourmohammad et al., 2021; Zabihzadeh et al., 2017). In these studies, the ToM abilities of BPD patients did not differ from HC. However, several studies revealed deficits in AToM abilities, as measured using the FPT (Stone et al., 1998), in BPD patients (e.g., Baez et al., 2015; Harari et al., 2010; Petersen et al., 2016; Zabihzadeh et al., 2017). One study found that BPD patients did not differ from HC on basic FB tasks but made significantly more ToM errors in affective and cognitive faux-pas understanding than HC (Petersen et al., 2016). These discrepancies might be due to the fact that the FPT examines the perception of others' emotions in more complex situations of the social context that resemble real-life events and contextual cues, which are more sensitive for detecting ToM deficits in BPD patients. On overall ToM ability, some studies found that patients with BPD underperformed HC on the FPT (Baez et al., 2015; Harari et al., 2010; Petersen et al., 2016; Pluta et al., 2018; Pourmohammad et al., 2021; Zabihzadeh et al., 2017), while another study revealed that patients with BPD or MDD and HC do not

differ (Yeh et al., 2017). However, a closer look at the various ToM facets reveals more contrasting results.

In the meta-analysis by Németh et al. (2018), the studies that used the FPT and the MASC (Dziobek et al., 2006), revealed that patients with BPD differ from HC in their reasoning abilities and cognitive ToM but did not differ in mental state decoding. This result was replicated by another meta-analysis, which also found that BPD patients underperformed HC in ToM reasoning but not decoding (Bora, 2021). In addition to that, the study revealed that BPD patients show deficits in faux-pas recognition compared to HC. Two studies found significant differences between BPD patients and HC, particularly in faux-pas recognition and cognitive understanding of faux-pas (Baez et al., 2015; Harari et al., 2010). One study found impairments in the cognitive and affective components of ToM (Petersen et al., 2016).

As with other disorders (e.g., schizophrenia or MDD), the symptoms of BPD also correspond with various ToM deficits. It has long been considered that exceeding ToM is a central feature of BPD (Bora, 2021; Sharp & Vanwoerden, 2015). However, a recent meta-analysis has demonstrated that exceeding ToM is not exclusive to BPD but is in fact widespread in many mental disorders (McLaren et al., 2022). The findings are consistent with a dimensional approach to mental health, as exemplified by the RDoC (RDoc; Michelini et al., 2021; National Institutes of Mental Health, 2020) or the HiTOP framework (Conway et al., 2019). A study comparing the ToM abilities of individuals with BPD, other personality disorders (OPD), and HC found that ToM was related to the severity of personality psychopathology and severity of symptom distress (Normann-Eide et al., 2020). A further recent study comparing ToM performance in patients with BPD and OPD also found that exceeding ToM was associated with the severity of personality psychopathology and symptom distress in both groups (Burghardt et al., 2023). These findings indicate that personality disorder severity may be a more reliable predictor of exceeding ToM than BPD in particular and that general symptom distress may be a more pertinent factor in exceeding ToM than specific personality disorder symptoms.

ToM deficits were assessed not only in patients with BPD alone but also in patients with BPD with comorbid MDD (Richman & Unoka, 2015; Zabihzadeh et al., 2017). However, there are mixed results on ToM deficits in patients with BPD and comorbid MDD. A meta-analysis showed that patients with BPD and comorbid MDD performed better on the RMET than patients with BPD alone (Richman & Unoka,

2015). Contradictory results were found in a study by Zabihzadeh et al. (2017), who found that BPD patients with comorbid MDD showed poorer ToM performance, assessed with the FPT and the RMET, than those with BPD alone. The findings of a recent study by Burghardt et al. (2023) complement these results. In their study, patients who showed symptoms of BPD and depression performed poorer on overall ToM abilities, assessed with the MASC. As already mentioned, however, the measuring instrument used influences the results (Quesque & Rossetti, 2020).

1.5.5 *Theory of Mind and Posttraumatic Stress Disorder*

PTSD is a common psychiatric disorder that is characterized by debilitating symptoms such as avoidance (of thoughts or activities reminiscent of the event), re-experiencing (e.g., flashbacks, intrusive thoughts, or nightmares), and hyperarousal (Eilers & Rosner, 2021). According to the DSM-5, the symptoms of PTSD are categorized into four symptom clusters, namely avoidance, re-experiencing, hyperarousal, and negative changes in mood and perception (American Psychiatric Association, 2013). These symptoms lead to a considerable impairment of personal, social, educational, or occupational functioning. An estimate seven out of ten people worldwide experience at least one traumatic event during their lifetime (Kessler et al., 2017). However, the lifetime prevalence of developing PTSD is estimated to be only around 6% (Koenen et al., 2017).

The most important risk factors for the development of PTSD include the type of trauma, previous trauma, and gender (Karam et al., 2014), as well as the severity of the traumatic event and the lack of social support following the traumatic event, as shown in two meta-analyses (Brewin et al., 2000; Ozer et al., 2003). This relationship is also evident in the opposite direction, as PTSD in turn also influences the utilization of social support. It seems reasonable to posit that the symptoms of PTSD themselves, such as negative changes in mood and cognition, have the effect of reducing the ability to perceive, interpret, and respond to other people (i.e., social cognition; Green et al., 2008). These abilities are crucial to effectively make use of social support (Sharp et al., 2012).

Social-cognitive deficits in patients with PTSD, especially in the subdomain ToM (Green et al., 2008) were identified in two meta-analyses (Janssen et al., 2022; Plana et al., 2014) and one systematic-review (Couette et al., 2020). Janssen et al. (2022) focused only on PTSD patients in their meta-analysis. They demonstrated a medium effect size for general social-cognitive deficits in individuals with PTSD compared to

controls, regardless of trauma type, gender, and age. Two of the five included studies used AToM tests (e.g., FPT, Strange Stories) to measure ToM. Their analysis demonstrated a large effect-size for ToM impairments in PTSD patients compared to controls. The results confirm the findings of an earlier meta-analysis by Plana et al. (2014), which, however, only included one study on ToM deficits in PTSD patients (measured with the RMET and Strange Stories test). The systematic-review also found impaired social cognition in individuals with PTSD and comprehensive disturbances in their cognitive and affective ToM (Couette et al., 2020). Results revealed affective ToM impairments in all included studies but cognitive ToM impairments in only 50% of the studies. Notably, the studies assessed ToM with the RMET and AToM tasks (i.e., FPT, Strange Stories), and in the study that measured ToM with the FPT (Nietlisbach et al., 2010), patients with PTSD did not differ from controls. So far, this is the only study that has investigated ToM differences in PTSD patients using the FPT.

As with MDD, PTSD has been both associated with reduced and exceeding ToM (Berthelot et al., 2019). Research on the effects of childhood trauma on ToM shows that there is an association between childhood trauma and PTSD through reduced ToM (Doba et al., 2022; Huang et al., 2020; Sharp et al., 2012) and an association between childhood trauma and both reduced and exceeding ToM (Doba et al., 2022).

Several factors may account for the observed ToM impairment in people with PTSD. First, ToM impairments may be explained by PTSD symptoms. There is evidence to suggest that PTSD symptoms, for example, emotional numbing, emotional unresponsiveness, or feeling detached from others (American Psychiatric Association, 2013), may potentially make it more challenging to process and integrate social information (Mazza et al., 2012; Plana et al., 2014). Secondly, neuroimaging research shows that PTSD symptoms and ToM deficits are associated with overlapping disrupted brain regions, including hyperactive amygdala reactivity and lower activation in the medial prefrontal cortex (Pitman et al., 2012; Zoladz & Diamond, 2013). Conversely, ToM impairments might also serve as a risk factor for the development of PTSD symptoms. ToM deficits may impair the use of social support and therefore may increase the risk of developing PTSD after a trauma, as a lack of social support is an important risk factor (Brewin et al., 2000; Ozer et al., 2003). Moreover, it is known that childhood trauma can change the development of the brain, particularly through differences in the areas responsible for social cognition (Hanson et al., 2015; Lupien et al., 2009). Thus, in accordance with the latent vulnerability model (McCrory & Viding,

2015), ToM impairments may be regarded as a latent vulnerability factor in the emergence of psychiatric symptoms subsequent to trauma in later life.

1.6 The Present Study

1.6.1 Main Research Issues

In summary, it can be stated that ToM impairments represent a potential risk factor for the onset of psychiatric symptoms and that mental disorders have been reliably associated with ToM impairments. However, as outlined in Section 1.4, ToM research is currently confronted with certain methodological problems. It is therefore essential to implement reliable measurement methods in order to investigate findings on ToM abilities in clinical groups. In the disorders described above (ASD, Schizophrenia, MDD, PTSD, and BPD), the main ToM instruments used were Strange Stories, the RMET, the MASC, and the FPT. The RMET test has been criticized for not measuring affective ToM, but rather emotion recognition (Kittel et al., 2022) and for not fulfilling the criteria of a ToM test (nonmerging and mentalizing criterion; Quesque & Rossetti, 2020). The Strange Stories task, the MASC, and the FPT are in line with the criteria established by Quesque and Rosetti (2020). The Strange Stories task is easy to administer as it is used in a paper-and-pencil format. However, it uses an open-ended answers format (Happé et al., 1998). Despite the use of a pre-established scoring system and the input of a second rater who provides consistent results despite being unaware of the participant group and hypothesis, this format has reduced objectivity and is still more time-consuming than a multiple-choice format. The MASC is administered using a multiple-choice answer format. However, the implementation is challenging as speakers and monitors are required for the presentation of a series of short videos. It may not be possible to implement this testing environment in most clinical settings. Therefore, it is necessary to further develop an alternative ToM test that is both easy to administer and easy to score and falls into a category of the most common AToM operationalizations, as outlined by Osterhaus and Bosacki (2022).

The FPT falls into the fifth most common AToM operationalization (i.e., recognition of faux-pas; Osterhaus & Bosacki, 2022) and loads on a different factor (i.e., recognizing transgressions of social norms) than the RMET and the Strange Stories task (Osterhaus et al., 2016). In comparison to the MASC, the FPT is easy to implement as it can be administered in a paper-and-pencil or digital format. As described earlier in Section 1.4.2.3, the FPT represents social blunders (i.e., faux-pas), resulting from unintended statements or behaviors by one protagonist that negatively

impact another protagonist (Baron-Cohen et al., 1999; Osterhaus et al., 2016). Although the FPT can be administered in a paper-and-pencil or digital format, the fact that it utilizes an open-ended response format still complicates the application and scoring of the FPT (Stone et al., 1998). This makes the implementation and interpretation time-consuming and labor-intensive and limits the objectivity of results interpretation. The development of a closed-answer format for the FPT would facilitate administration and evaluation.

Mental disorders are linked to different types of ToM impairments. ToM impairments van vary between no/less ToM and exceeding ToM. In other words, individuals can show either a limited tendency to ascribe mental states to others or a tendency to over-interpret others' behaviors (Vegni et al., 2021). According to a recent meta-analysis by McLaren et al. (2022), exceeding ToM serves as a clinical marker across a broad range of mental disorders (e.g., schizophrenia, ASD, BPD, and MDD). It is important to note, however, that they have only included studies that used the MASC as a measure of exceeding ToM, as they criticized that validity of the Reflective Functioning Questionnaire (RFQ; Müller et al., 2022) and the Hypermentalizing Questionnaire. The classification of ToM errors in the MASC is divided into three categories: no ToM, less ToM, and exceeding ToM. Nevertheless, the validation of these error types is still pending. It therefore would be highly beneficial to develop a polytomous closed-answer format for the FPT to be able to analyze the degree of ToM, which ranges from reduced ToM to correct ToM to exceeding ToM.

Furthermore, it is important to acknowledge that the aforementioned studies that identified specific ToM impairments in mental disorders were conducted using questionnaires that may not have been entirely valid. It would thus be beneficial to investigate ToM impairments in various mental disorders with a valid ToM measure.

In order to address the three issues described above, this thesis reports three studies. To facilitate the interpretation of the responses and thus increase objectivity, this thesis reports on a further development of the FPT with a newly developed closed answer format for the German adult version of the FPT (Ströbele, n.d.). The first study validates the newly developed dichotomous closed-answer format of the FPT. This enables us to compare the results of the FPT with other ToM measurements. The second study validates the newly developed polytomous closed-answer format of the FPT, which assesses the degree of ToM impairments. The third study compares the

level of ToM ability in two disorders (MDD and BPD) and a community sample to examine the extent to which ToM differs between clinical and community samples.

1.6.1   Overarching Research Questions

The main objectives of this thesis are the following: 1) to evaluate a new closed-answer format of the FPT (correct vs. incorrect answers) and to validate the FPT by comparing it with other ToM measures, 2) to evaluate a new closed-answer format that examines different ToM levels (ranging from reduced ToM to correct ToM to exceeding ToM), and 3) to compare the level of ToM ability in two disorders (MDD and BPD) and a community sample to examine the extent to which ToM differs between clinical and community samples.

More specifically, the present thesis aims to address the following research questions:

1) Does the FPT reliably measure ToM and is the newly developed dichotomous answer format valid?

> Study 1 distinguished between correct/incorrect responses, irrespective of the various ToM impairments. The objective was to ascertain whether the items of the FPT measure a single underlying concept and, if so, whether this concept was ToM.

2) Does the newly developed polytomous response format of the FPT measure different levels of ToM?

> The objective of Study 2 was to provide a more detailed assessment of ToM. The study therefore distinguished between various ToM impairments (i.e., no ToM, correct ToM, and exceeding ToM) and sought to investigate whether the polytomous answer format, as opposed to the dichotomous correct/incorrect format, diminished the validity of the FPT.

3) Are there discrepancies between a mental disorder sample and a community sample in terms of their specific ToM abilities (no ToM, correct ToM, exceeding ToM)?

> The aim of Study 3 was to ascertain whether and to what extent individuals with the mental disorders BPD and MDD exhibit ToM deficits compared to individuals in a community sample. The study thus sought to investigate in which ToM level individuals with MDD or BPD show difficulties, compared to a community sample.

1.7 Further Development of the FPT

As all studies related to the thesis focus on the newly developed closed-answer format of the FPT, it is essential to provide a detailed account of the development process of the closed-answer format. As outlined in Section 1.4.2.3, the FPT presents social blunders (= faux-pas) embedded in stories, labeled as vignettes (Stone et al., 1998). A faux-pas occurs when an individual says something without considering whether the listener might want to hear or know it. The individual is usually unaware of the mistake, which has negative and unintended consequences (Baron-Cohen et al., 1999). After each vignette (i.e., story), up to eight questions are presented. The original answer format was either dichotomous (i.e., yes/no) or open-ended. We used the German adult version of the FPT (Ströbele, n.d.) and further revised the vignettes in order to better adapt them to both Austrian and German linguistic and cultural norms. We left the translated questions and the dichotomous yes/no answers in their original form. For the questions with an open-ended answer format (e.g., "Why do you think he/she said it?"), we developed a four-stage multiple-choice answer format based on a combination of theoretical and empirical considerations: First, a team consisting of five researchers collected the responses to the open-answer format from patients with MDD, PTSD, and BPD. Second, the team created answers for the four different response categories independently. The final responses were selected through group discussions. All answer options were formulated to be suitable for both Austrian and German-speaking individuals.

In the original and the German version of the FPT, several questions were formulated in a dichotomous yes/no answer format: Question 1 asks whether a faux-pas occurred or not, Question 5 asks whether the protagonist knew they committed a faux-pas, and three of the forty asked comprehension questions (Questions 7 and 8; i.e., "Did Sandra and Maria know each other?", "Did Svenja's story win anything?" and "Were there any seats available on the bus when she got on?"). One comprehension question (i.e., "Did Sarah remember that the party was a surprise party?") originally also had the dichotomous answer format. However, the results of our pilot test indicated that participants frequently responded with "not yet known" instead of selecting "yes" or "no". We therefore added "not yet known" as a third answer alternative.

Questions 2, 3, 4, and 6 were originally posed in an open-ended answer format. For Question 2, which asks respondents to identify the person who committed the faux-

pas, we developed a four-stage response format of different names based on the characters mentioned in the story. If fewer than four characters were mentioned in a vignette, we added a common German name as a distractor to create four alternative answers. For the other comprehension questions (7 and 8), we invented various distractors, such as different types of clothes, beverages, or colors. For the remaining three items, Question 3) why the protagonist should not have said/done something, Question 4) why they did it anyway, and Question 6) how the "victim" of the faux-pas felt, we created four multiple-choice answer alternatives which represent the correct response and three incorrect responses with different types of ToM impairments. As outlined in Section 1.2.3, ToM impairments, also known as error types, are based on previous research (Fretland et al., 2015; Vaskinn et al., 2015). The extent of these ToM impairments can range from no/reduced ToM to exceeding ToM. The former indicates a tendency to either ascribe no or too few mental states to others (Cortés-García et al., 2021; Dziobek et al., 2006; Quek et al., 2018), and the latter indicates the overinterpretation of the behavior of others in terms of mental states (Sharp & Vanwoerden, 2015; Vegni et al., 2021). We created one no ToM answer-category and two different exceeding ToM answer-categories, whereby we distinguished between positively or negatively overinterpreting the story character's intention.

2 Study 1: The Faux-Pas Recognition Test: Validation of the Dichotomous Answer Format

Despite the widespread use of ToM measures, significant concerns remain about their validity and reliability (Hayward & Homer, 2017). One commonly used ToM measurement, the FPT, has demonstrated good reliability, including high internal consistency (Şandor & İşcen, 2021). However, it has never been investigated whether the FPT actually measures a single, unidimensional construct (ToM). The present study aims to address this gap by applying Rasch analysis within the Item Response Theory (IRT) framework to assess the validity of the FPT.

To fully assess whether the FPT measures a single ability (ToM), IRT models, such as Rasch analysis, allow a more detailed examination of how individual items contribute to the measurement of the underlying construct. The one-dimensional Rasch or one-parameter logistic (1-PL) model (Rasch & Wright, 1980) is a probabilistic model that assumes that all items in a test measure a single dimension of ability. In other words, a key assumption of Rasch analysis is unidimensionality, i.e., all items should measure a single latent trait. Thus, Rasch analysis provides insights into item functioning by analyzing the probability of a correct response in relation to an individual's underlying ability (Baker & Kim, 2004).

According to the Rasch model, the probability of test takers answering a particular item correctly is influenced by two key factors: the person's ability and the difficulty of the item (Kubinger, 2005). This means that if a respondent's ability level exceeds the difficulty of an item, it is likely that they will answer the item correctly. The Rasch model can thus be regarded as a probabilistic analog of the Guttman scale analysis, which is based on the assumption that an individual with a specific ability level will (in accordance with the Rasch model) likely or (in accordance with the Guttman model) definitely respond correctly to items whose difficulty estimates are below their ability level (Wellman & Liu, 2004). The Rasch analysis is based on two fundamental assumptions: Firstly, more difficult items are less likely to be answered correctly. Secondly, individuals with a higher level of knowledge are more likely to answer questions correctly (Trakman et al., 2017).

These principles provide a robust framework for assessing the alignment of item difficulty with respondent ability, ensuring that the model measures a single, unidimensional construct. Therefore, items function consistently regardless of the

group being assessed (Baker & Kim, 2004). The probability is therefore dependent on the trait. If for some reason one group does not display the same probability of affirming the item, this would indicate a violation of the requirement of unidimensionality.

The present study has two main objectives: first, to validate the newly developed closed-answer format of the FPT to ensure it measures a single, underlying construct (ToM) and second, to estimate the difficulty of the items within this context. In this study, the underlying unidimensional ability being measured corresponds to the correct ToM. In other words, high person ability in the Rasch model refers to how effectively a respondent mentalizes. We tested whether the FPT maintains unidimensionality and explored its associations with the RMET to extend evidence of its validity. By fitting a unidimensional Rasch model to our data, we obtained estimates of ability and evaluated the overall effectiveness of the new answer format. Moreover, we correlated the FPT with the RMET and the MASC to extend the evidence of validity.

2.1   Method
2.1.1   Measures
4.1.1.1 The Faux-Pas Recognition Test (FPT). We developed a multiple-choice answer format (LINK) for the German adult version of the FPT (Ströbele, n.d.). The FPT is described in Section 1.4.2.3.

4.1.1.2 Reading the Mind in The Eyes – Test (RMET). We used the German version of the RMET (RMET; Pfaltz et al., 2013). This test is described in Section 1.4.2.4.

4.1.1.3 The Movie for the Assessment of Social Cognition (MASC). We also administered the MASC to assess ToM (Dziobek et al., 2006). The MASC is described in Section 1.4.2.5**.**

4.1.1.3 PTSD Screening. In order to exclude participants with PTSD from the first community sample, we utilized the Primary Care PTSD Screen for DSM-5 (PC-PTSD-5; Prins et al., 2016). The questionnaire assesses symptoms of PTSD within the past month. With five items, it measures the occurrence of nightmares, avoidance behavior, hyperarousal, emotional blunting, and alienation following an event. We excluded participants who scored ≥ 4 (Bovin et al., 2021). The instrument has previously demonstrated high internal consistency (α = .83; Cheng et al., 2021).

4.1.1.4 Sociodemographic Characteristics. We recorded the participants' self-reported gender, age, and education level. Gender was assessed polytomous

(female, male, diverse). The level of education was determined according to the Austrian education system and categorized into two groups: low level of education (compulsory school, middle school) and medium/high level of education (A-levels or university degree). In addition, the partnership status of the participants was coded into "living in a partnership" vs. "not living in a partnership".

2.1.2 Participants

In this study, we only included participants from the general population. Therefore, we excluded data from participants who reported currently receiving psychiatric, psychological, or psychotherapeutic treatment.

For the first community sample, we collected data from 109 participants from the general population. Forty-one data sets were incomplete. Additionally, we excluded four participants who reported currently receiving psychiatric, psychological, or psychotherapeutic treatment and five who had a positive screening for PTSD. Thus, the resulting sample consisted of 59 participants (23 women, 35 men, 1 diverse), with a mean age of 39.93 (SD = 15.88) years. A majority, 36 (61.02%), reported a medium or high education level, while 10 participants (16.95%) reported a low education level. Most participants (40, 67.80%) were in a partnership.

We recruited 57 participants from the general population for the second community sample. Three data sets were incomplete, three participants did not agree to data usage, and five participants were excluded as they were currently undergoing psychiatric, psychological, or psychotherapeutic treatment. Thus, the final sample consisted of 46 participants (32 women, 14 men). Their mean age was 27.48 years (SD = 14.07). A majority, 42 (91.30%), reported medium or high education levels, while 4 (8.70%) reported a low education level. Twenty-six participants (56.52%) had a partner.

The resulting sample comprised 105 participants (55 women, 49 men, 1 nonbinary). Their mean age was 34.48 (SD = 16.28) years. Seventy-two (72.38 %) participants stated that they had a medium/high education, which means the sample is comparatively well-educated (STATISTIK AUSTRIA, 2023). In the sample, 66 (62.86 %) were living in a partnership.

2.1.3 Procedure

The first community sample was gathered between June 2021 and September 2021. The second community sample was collected between April and July 2023. Both studies employed the SoSci Survey platform to administer the FPT. We employed a

snowball sampling method. The initial sample for the first community sample consisted of family members, relatives, and acquaintances from the personal environment of a master's student. The initial sample for the second community sample consisted of undergraduate students. The initial samples were then asked to tell other people about the study. Furthermore, participants in the first community sample were contacted via a variety of social media groups on Facebook, WhatsApp, and other online platforms. Following the provision of informed consent and completion of the mental health treatment item, participants proceeded to complete the FPT. The data was collected between June and September 2021. The participants in all samples provided written informed consent. The study was carried out in compliance with the ethical principles outlined in the Declaration of Helsinki (World Medical Association, 2013) received approval from the ethics committee of Karl Landsteiner University of Health Sciences (No: 1004/2021; 1010/2023).

2.1.4 Data Analysis

4.1.4.1 Data Preparation. As our aim was to assess ToM in terms of specific responses beyond yes-no answers, we limited our analysis to a subset of items from the FPT. First, as only three items per vignette assess ToM beyond a yes-no category, we thus, we selected three items per vignette. Secondly, we only included the vignettes that contained faux-pas (the targets), since the answers of the distractors (foils) do not allow the measurement of ToM errors. This resulted in 10 of the 20 vignettes being included. In total, 30 items were selected for the analysis. We dichotomized the 4-point answer format into the two categories correct (1) and incorrect (0). We thus defined the three ToM errors (positive exceeding ToM, negative exceeding ToM, and no ToM) as incorrect (0). The items were dichotomized using IBM SPSS (Version 27). If participants did not recognize a faux-pas, the three missing answers to the items of the vignette were coded with 0 (no ToM) because not recognizing that something inappropriate happened represents an insufficient ToM ability. In addition, correct answers to the two control questions per vignette were required to exclude factors not related to ToM.

4.1.4.2 Rasch Analysis. We fitted the Rasch model, a one-parameter logistic model (1-PL), to the data of the dichotomized items. We conducted the analyses using R 4.0.3 with the extended Rasch modeling (eRm; Mair & Hatzinger, 2007) and latent trait models (ltm; Rizopoulos, 2006) software packages. The objective of a Rasch

analysis is to develop a unidimensional questionnaire that accurately captures a single construct. The extent to which items contribute to a unidimensional construct (i.e., the fit of the Rasch model) can be evaluated by examining the fit indices of the items, which determine whether a specific item fulfills the assumptions of the model (Rasch & Wright, 1980). Fit indices are divided into infit and outfit mean-square statistics. The infit statistic is more sensitive to unexpected responses to items near the participants' ability level as predicted by their overall pattern of responses; the outfit statistic is most sensitive to unexpected responses far away from the participant's ability level (i.e., outlying scores). Because outfit is less threatening to measurement and easier to manage, infit scores are typically considered more informative. Reliability indices, such as the person separation reliability, can be employed to ascertain the dimensionality of the underlying ability. Separation reliability indicates the accuracy of the person or item separation at different ability or difficulty levels, i.e., how well the scale can differentiate between high and low ability levels. It ranges from 0 to 1, with a higher value indicating better differentiation (Verhavert et al., 2018).

We verified the fit of each of the questions to the model, taking into account the mean squares parameters (MNSQ) using the Rating Scale Model. Items with the infit MNSQ in boldface were removed due to misfit, as identified by values larger or smaller than 1 ± .20. This rather strict cut-off was used by the majority of reviewed studies. We also assessed the difficulty of the items (i.e., the probability of an item being answered correctly) to explain the score.

To extend the evidence of validity, we correlated the sum of the 30 dichotomous items of the FPT with other measures of social cognition: the RMET and the MASC.

2.2 Results

As previously outlined, we employed a Rasch model to analyze the data. This incorporated a discrimination parameter, a difficulty parameter, and item fit indices. Table 2 presents the results in detail. For each Vignette, the questions are in the same order and represent similar meanings. This means that each question of a vignette asks comparably the same thing: 3) why the protagonist should not have said/done something, 4) why they did it anyway, and 6) how the "victim" of the faux-pas felt.

All but four items had a good fit to the model (i.e., 0.80> *infit* mean-square statistic [MNSQ] <1.20; see Table 2). The following items were subsequently excluded:

Vignette 4: Item 4, Vignette 11: Item 6, Vignette 12: Item 6, Vignette 16: Item 4. The remaining 26 items were found to fit the Rasch model. Thus, the Rasch model demonstrated an excellent fit for 26 of the 30 items, as indicated by the infit statistics. The model demonstrated good separation reliability, r = .82. Overall, our findings suggest that the FPT shows good properties and that the items load on a single dimension. Item difficulties and person abilities are also plotted in the Wright Map in Figure 2. The wright map is a visual representation of the Rasch analysis, which displays person abilities on one side and item difficulties on the other. The vertical axis represents the latent trait being measured, in this case, ToM. Correlation analyses revealed strong associations between the FPT and the RMET, r = .54, $p < .001$, [.33; .70] and between the FPT and the MASC, r = .55, $p < .001$, [.31; .72].

Table 2

Results of the Rasch Analysis

No*	Vignette	Item	Diff.	MNSQ Outfit	Infit	χ2
2	Surprise Party	Item 3	-0.960	0.963	1.141	99.215
	Reveal	Item 4	-0.476	0.944	0.998	97.214
		Item 6	-0.884	0.760	0.942	78.284
4	Ugly Curtains	Item 3	-0.453	0.905	1.014	86.886
		Item 4	0.711	1.274	**1.207**	122.315
		Item 6	1.669	1.258	1.130	120.812
7	Lena is not a Boy	Item 3	-0.496	0.835	0.851	75.982
		Item 4	-0.427	0.821	0.849	74.730
		Item 6	1.042	1.153	1.135	104.900
11	Colleague is sick	Item 3	-0.572	0.865	0.983	87.391
		Item 4	-0.417	0.979	0.989	99.829
		Item 6	0.947	1.399	**1.291**	142.673
12	Football Changing	Item 3	0.575	1.722	1.194	175.681
	Room	Item 4	1.367	1.106	1.091	112.768
		Item 6	1.367	1.315	**1.226**	134.121
13	Apple Pie	Item 3	-0.476	0.785	0.946	80.852
		Item 4	-0.604	0.708	0.854	72.967
		Item 6	-0.476	0.787	0.925	81.062
14	Ugly Crystal Bowl	Item 3	-0.763	0.708	0.882	71.544
		Item 4	-1.079	0.721	0.934	72.792
		Item 6	-0.763	0.791	0.946	79.904
15	Storytelling	Item 3	1.132	0.755	0.817	77.056
	Competition	Item 4	0.432	1.240	1.169	126.502
		Item 6	-1.048	1.143	1.031	116.620
16	Not a Waiter	Item 3	-0.623	0.536	0.714	52.548
		Item 4	-0.766	0.468	**0.674**	45.894
		Item 6	-0.008	0.793	0.877	77.722

| No* | Vignette | Item | Diff. | MNSQ | | χ^2 |
				Outfit	Infit	
18	Lawyer	Item 3	1.019	0.769	0.838	76.926
		Item 4	-0.179	0.772	0.834	77.154
		Item 6	1.208	1.031	1.068	103.067

Note. *The numbers of the vignette correspond to the numbers of the original FPT, but we have only included targets; Diff. = difficulty; χ^2 = fit statistic; MNSQ = mean-squares statistic. Items with the infit MNSQ in boldface were removed due to misfit, as identified by values larger or smaller than $1 \pm .20$

Figure 2

Wright map showing person abilities and item difficulties
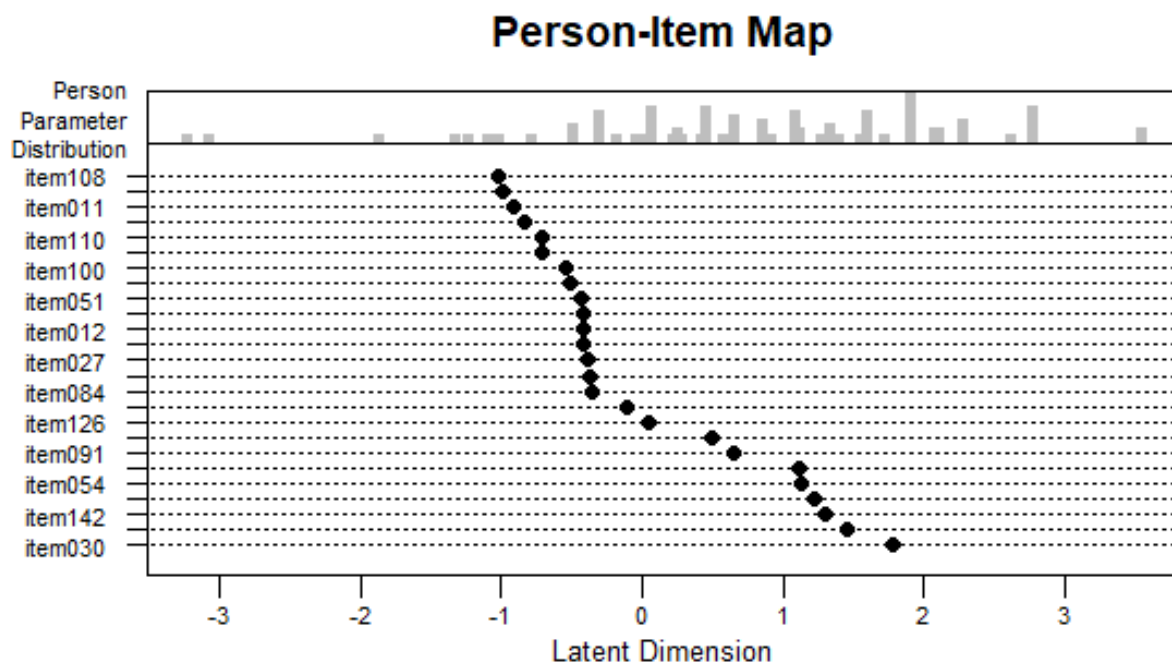
2.3 Discussion

The aim of the present study was to validate the accuracy of the closed answer format of the test in measuring a single underlying construct (ToM) and to estimate the difficulty of the items in this context.

The results of the Rasch analysis indicate that the FPT measures a single construct and that the dichotomous closed-answer format has been successfully operationalized. In light of these findings, the dichotomous version of the closed-answer format of the FPT might be an excellent instrument for detecting ToM deficits. To the best of our knowledge, this was the first study to employ the Rasch model for the assessment of the psychometric properties of the FPT. Rasch analysis enables the simultaneous comparison of item difficulty and persons' ability on the same logit scale. Overall, the criteria for reliable measurement were met, but four items were excluded from the analysis due to unsatisfactory infit and outfit indices.

The shortened 26-item FPT version demonstrated a unidimensional construct. This is demonstrated by the results, which indicate that despite the varying levels of difficulty, the items are not significantly different from one another. This indicates that the test items are homogenous among themselves, allowing us to attribute the participants' performance on the test to a single ability. We assume that the ability being measured is ToM, as previous literature indicates that the vignettes of the FPT measure ToM (Quesque & Rossetti, 2020). Our study demonstrates the great value of Rasch analysis in providing detailed item-level analysis and refining traditional psychometric methods. In conclusion, despite some items not performing as expected, the FPT performed well on most aspects of the assessment.

In addition to empirically testing the validity of the closed response format, the study also assessed convergent validity by correlating the FPT with another established ToM measure, the MASC. The results revealed a strong correlation, although it was closer to the moderate range. This indicates that while the two measures are related, they capture different aspects of ToM. The MASC assesses the ability to interpret social situations by integrating multimodal cues, such as facial expressions, gestures, and prosody. In contrast, the FPT evaluates the ability to identify social mistakes or faux-pas within structured narratives (Osterhaus & Bosacki, 2022). This highlights the validity of the FPT as a complementary tool for assessing ToM. This is particularly meaningful, as very high correlations might suggest

redundancy between the measures, potentially undermining the need for a new test. Instead, the overlap acknowledges the shared construct (ToM), while emphasizing the unique contributions of each measure. Given that ToM is a multifaceted construct, a correlation of this magnitude—while formally classified as strong—more closely reflects a moderate association. These findings underscore the importance of using multiple, diverse measures to capture the breadth of ToM abilities. Even though the RMET had been criticized as a ToM measure, the study also investigated associations with the RMET. The correlation with the RMET is also formally classified as strong. Again, however, the effect size is closer to the moderate range, reflecting the partial overlap and complementary focus of the two instruments in assessing different facets of ToM. this does not invalidate the FPT, as the two tests assess different aspects. Firstly, the RMET loads on a different factor of ToM than the FPT. Osterhaus et al. (2016) conducted an analysis of items from seven different ToM tasks (e.g., RMET, strange stories, FPT) with the objective of identifying the underlying factors. Through Rasch scaling and factor analyses, three distinct factors within ToM were identified: 1) social reasoning, 2) reasoning about ambiguity, and 3) recognizing transgressions of social norms. The FPT is primarily concerned with assessing instances of transgression of social norms. It is worth noting that the RMET is associated with a different factor: social reasoning. Secondly, research indicates that the RMET assesses decoding abilities, rather than reasoning about mental states (Maleki et al., 2020). In the RMET, participants are required to decode the emotions of the protagonists, which is why some studies have suggested that the RMET is a measure of emotion recognition (Kittel et al., 2022; Quesque & Rossetti, 2020). By contrast, the FPT requires participants to decode the mental states of both protagonists in order to identify a faux-pas and reason about the social norms that have been partially violated in these situations. Consequently, the tests assess different abilities, which explains why the associations between them are weaker.

2.4 Limitations and Future Directions

One limitation of the study is the inclusion of only three items per vignette, as only these three assess specific ToM responses beyond a yes/no category. It would be beneficial for future studies to validate the remaining items in the vignettes. A further limitation is that the sample mainly consisted of individuals from Western, Educated, Industrialized, Rich, and Democratic (WEIRD) backgrounds (Henrich et al., 2010),

which limits generalizability. This is, therefore, a common limitation in research. It is recommended that future studies include a more diverse sample and extend validation efforts to clinical groups with known ToM deficits. Such individuals may include those with autism spectrum disorders (Baron-Cohen, 2000), schizophrenia (Sprong et al., 2007) or personality disorders (McLaren et al., 2022). This should be used to draw conclusions about the entire population. The limitation of the highly educated sample is related to a further limitation, namely the correlations with the RMET. The RMET is a relatively easy test that is unable to differentiate between individuals with high levels of ability (Black, 2019). Therefore, it would be beneficial to use alternative ToM measures as an external criterion to further validate the response format in future studies. However, this limitation is mitigated in our study by the additional use of the MASC, a more complex ToM measure that does not exhibit ceiling effects (Yeung et al., 2024), thereby allowing for a more precise assessment of advanced ToM abilities.

2.5 Conclusion

In conclusion, the study validates the FPT and its dichotomous response format. The results demonstrate a strong alignment between the model and the data, indicating the FPT's psychometric suitability and accuracy in measurement. The FPT is therefore able to measure correct ToM. This advanced psychometric evaluation will not only enhance the objectivity and validity of the FPT but also contribute to the broader field of ToM assessment by informing the development of more refined and equitable measurement tools.

3 Study 2: The Faux-Pas Recognition Test: Validation of the Polytomous Answer Format

ToM deficits have been linked to a variety of mental disorders, including personality disorders and particularly BPD. BPD is marked by social and interpersonal difficulties (American Psychiatric Association, 2013; Lazarus et al., 2014) and often includes specific ToM impairments, such as exceeding ToM. This form of ToM is characterized by over-attributing mental states to others, often seeing intentions or motives that others might not find justifiable (Sharp & Vanwoerden, 2015). Other mental disorders, such as alcohol use disorder, are more commonly associated with reduced ToM, which refers to a diminished capacity to recognize or infer others' mental states (Onuoha et al., 2016). Recent research has challenged the notion that exceeding ToM is a distinctive feature of BPD. A meta-analysis conducted by McLaren et al. (2022) revealed evidence of exceeding ToM across a range of disorders (e.g., schizophrenia, autism spectrum disorder), concluding that it is associated with general psychopathology rather than specifically related to BPD.

In response to these findings, the present study seeks to develop a more nuanced measurement approach to ToM, focusing on the potential for ToM assessments to capture various levels of ToM ability rather than relying solely on binary responses. For this purpose, we developed a polytomous scoring format for the FPT and utilized the PCM (Masters, 1982). PCM, a type of Rasch model, extends the one-parameter logistic model (1-PL) in IRT by allowing for multiple scoring categories beyond simple correct/incorrect responses. This framework is particularly suited for items with polytomous (i.e., multi-level) scoring, where responses can capture gradations of correctness or understanding rather than a binary judgment.

The polytomous format enables a more comprehensive assessment by recording multiple response levels, which can more accurately reflect the complexity of ToM abilities. For example, individuals with greater ToM ability are expected to receive higher scores than those with lower ability, as the model assumes a partial credit pattern where intermediate responses reflect intermediate levels of understanding (De Ayala, 1995). This allows PCM to detect a broader range of ToM competencies, potentially capturing subtle distinctions in ToM impairments that a dichotomous model might overlook. Thus, the PCM provides a more precise and differentiated measurement of ToM abilities (Rasch & Wright, 1980).

The PCM might be more stable in a community sample than in a BPD sample because the PCM assumes that the responses to the items are distributed along a latent trait. As the literature has pointed out that BPD is predominantly marked by exceeding ToM (Sharp & Vanwoerden, 2015), responses may be less diverse or have a different distribution. This could violate the assumptions of the model and make model fitting more difficult. Thus, it is necessary to apply the PCM both in a community and in a clinical sample.

The present study aims to investigate the psychometric properties of the polytomous response format of the FPT, specifically examining its effectiveness in providing a more nuanced assessment of ToM beyond the traditional binary response format. By applying the partial credit model (PCM), this study seeks to enhance the discriminative power of ToM assessments, allowing for a more detailed evaluation of varying ToM levels rather than a simple correct/incorrect classification. Additionally, the study sought to examine if the FPT can capture subtle variations in ToM abilities across both a community sample and individuals with personality disorders.

3.1 Method

3.1.1 Participants

5.1.1.1 Community Sample.

In the first sample, we excluded data from participants who reported currently receiving psychiatric, psychological, or psychotherapeutic treatment, or who tested positive for PTSD symptoms using the Primary Care PTSD Screen for DSM-5 (see below). Of the original 68 participants, the resulting sample after data exclusion comprised 59 participants (23 women, 35 men, 1 nonbinary). Their mean age was 39.93 (SD = 15.88) years. Twenty-nine (49.15 %) participants stated that they had a college degree, which means the sample is comparatively well-educated (STATISTIK AUSTRIA, 2023). In the sample, 40 (67.80 %) were living in a partnership.

We recruited 57 participants from the general population for the second sample. Three data sets were incomplete, declined consent for data usage, and five participants were excluded due to current psychiatric, psychological, or psychotherapeutic treatment. The final sample therefore included 46 participants (32 women, 14 men). Their mean age was 27.48 years (SD = 14.07). Most participants (n=42; 91.30%), reported medium or high levels of education, while 4 participants

(8.70%) reported a low education level. Additionally, twenty-six participants (56.52%) reported being in a relationship.

The final sample included 105 participants (55 women, 49 men, 1 nonbinary individual). Their mean age was 34.48 (SD = 16.28) years. In total, 72 participants (72.38 %) reported having a medium or high level of education, indicating that the sample was comparatively well-educated (STATISTIK AUSTRIA, 2023). Furthermore, 66 participants (62.86 %) reported being in a partnership.

5.1.1.2 BPD sample. The BPD sample was recruited from a psychiatric-psychosomatic hospital in Austria. Patients were treated for BPD in a specialized treatment unit. Furthermore, the study excluded patients who did not currently meet the respective criteria for BPD as defined by the standardized self-report measure BSL-23. The hospital required patients to be at least 18 years of age and to have a basic level of conversational German skills. Individuals with acute psychotic symptoms, current suicidal behavior, or acute intoxication were not admitted to the hospital. The BPD sample originally consisted of 44 patients. Three data sets were incomplete and where thus excluded. The resulting sample consisted of 42 patients (35 women, 7 men) with a mean age of 34.14 (SD = 12.30) years. More than half (23, 54.76%) of the patients reported a medium or high education level and 18 (42.86%) reported a low education level. A minority of 12 patients (28.57%) had a partner.

3.1.2 Measures

5.1.2.1 The Faux-Pas Recognition Test (FPT). We developed a multiple-choice answer format (LINK) for the German adult version of the FPT (Ströbele, n.d.). The FPT is described in Section 1.4.2.3.

5.1.2.2 Reading the Mind in The Eyes – Test (RMET). We used the German version of the RMET (RMET; Pfaltz et al., 2013). This test is described in Section 1.4.2.4.

5.1.2.3 PTSD Screening. We used the PC-PTSD-5 (PC-PTSD-5; Prins et al., 2016) to screen for PTSD. With five items, this screening measures the occurrence of nightmares, avoidance behavior, hyperarousal, emotional blunting, and alienation after an event. We used a cut-off score of 4 (Prins et al., 2016). The PC-PTSD-5 recently showed high internal consistency (α = .83; Cheng et al., 2021).

5.1.2.4 Sociodemographic Characteristics. We recorded the participants' self-reported gender, age, and education level. Gender was assessed polytomous (female, male, diverse). The level of education was determined according to the

Austrian education system and categorized into two groups: low level of education (compulsory school, middle school) and medium/high level of education (A-levels or university degree). In addition, the partnership status of the participants was coded into "living in a partnership" vs. "not living in a partnership".

3.1.3 Procedure

A master's student in psychology approached a community sample by contacting relatives, friends, and acquaintances to collect data. The participants were then asked to recruit new participants. The random sample was obtained using the snowball technique. Participants completed the questionnaires online on the SoSci Survey platform. After informed consent, participants completed the PTSD screening and a mental health treatment item. They then completed the ToM measures. The data was collected between June and September 2021. The study was approved by the Ethics Committee of the Karl Landsteiner University of Health Sciences (No.: 1004/2021).

3.1.4 Data Analysis

5.1.4.1 Data Preparation. As our aim was to assess ToM in terms of specific responses beyond yes-no answers, we limited our analysis to a subset of items from the FPT. First, as only three items per vignette assess ToM beyond a yes-no category, we thus, we selected three items per vignette. Secondly, we only included the vignettes that contained a faux-pas (the targets) since the answers of the distractors (foils) do not allow for the measurement of ToM errors. This resulted in 10 of the 20 vignettes being included. In total, 30 items were selected for the analysis. We divided the 4-point answer format into the three categories no ToM (0), correct ToM (1), and exceeding ToM (2). As we assumed positive and negative exceeding ToM to be at the same ToM level, we summarized the ToM errors as positive and negative exceeding ToM. The items were coded using IBM SPSS (Version 27). If participants did not recognize a faux-pas, the three missing answers to the items of the vignette were coded with 0 (no ToM) because not recognizing that something inappropriate happened represents an insufficient ToM ability. In addition, correct answers to the two control questions per vignette were required to exclude factors not related to ToM.

5.1.4.2 Partial Credit Model Analysis. Given our assumption that ToM is a dimensional ability that individuals have to a greater or lesser degree, we applied a PCM within the IRT framework to the data (Masters, 1982). This model is suitable for the analysis of responses obtained from two or more ordinal categories (polytomous

items). This model assumes that the categories (i.e., no ToM, correct ToM, and exceeding ToM) are hierarchical and tests whether the items measure a single underlying concept and estimate individuals' abilities and item difficulties. We conducted the analyses using R 4.0.3 with the eRm (Mair & Hatzinger, 2007) and ltm (Rizopoulos, 2006) packages. The PCM enables the investigation of whether responses can be classified into three distinct categories with an increasing level of difficulty. It is important to note our assumptions in this context: responses in the lowest category (0) are indicative of a lack of ToM (i.e., no ToM), whereas those in the middle category (.5) suggest a level of ToM that is correct (i.e., correct ToM). The highest category (1), which is awarded full credit, indicates responses that exceed the standard level of ToM (i.e., exceeding ToM). The distinction in the underlying ability between the Rasch model and the PCM is a key differentiating factor. In the Rasch model, the underlying ability that is measured is "correct ToM," whereas in the PCM, the focus is on the intensity of ToM. In other words, in the Rasch model, high person ability refers to the effectiveness with which a respondent is able to mentalize. In contrast, in the PCM, it relates to the extent of ToM demonstrated by the respondent, which can be categorized as no ToM, correct ToM, or exceeding ToM. The study analyzed the quality of the items using the difficulty/adjustment measure of the model (infit and outfit), item difficulty, item discrimination, and separation reliability.

Fit indices are divided into infit and outfit mean-square statistics. The infit statistic is more sensitive to unexpected responses to items near the participants' ability level as predicted by their overall pattern of responses; the outfit statistic is most sensitive to unexpected responses far away from the participant's ability level (i.e., outlying scores). Because outfit is less threatening to measurement and easier to manage, infit scores are typically considered more informative. These parameters indicate the difficulty of the transitions between the categories for each item. This means that for each item the threshold values (i.e., the transitions between the categories) for the individual levels of the item (c_1, c_2, c_3) were estimated. A higher threshold means that it is more difficult for respondents to move from the lower to the higher category, whereas a lower threshold means that it is easier for the respondents to cross this threshold. Reliability indices, such as the person separation reliability, can be employed to ascertain the dimensionality of the underlying ability. Separation reliability indicates the accuracy of the person or item separation at different ability or

difficulty levels, i.e., how well the scale can differentiate between high and low ability levels. It ranges from 0 to 1, with a higher value indicating better differentiation (Verhavert et al., 2018).

We verified the fit of each of the questions to the model, taking into account the mean squares parameters (MNSQ) using the Rating Scale Model. Items with the infit MNSQ in boldface were removed due to misfit, as identified by values larger or smaller than 1 ± .20. This rather strict cut-off was used by the majority of reviewed studies. We also assessed the difficulty of the items (i.e., the probability of an item being answered correctly) and the item discrimination (the item's capacity to differentiate between individuals with varying levels of the trait) to explain the score.

To extend the evidence of validity, we correlated the sum of the 30 dichotomous items of the FPT with another measure of social cognition, the RMET.

3.2   Results

We employed a PCM model to the data, which incorporated both a discrimination and a difficulty parameter. The results are presented in Tables 1 and 2 in detail.

For each vignette, the questions are in the same order and stand for similar meanings. This means, that each question of a vignette asks comparably the same thing: 3) why the protagonist should not have said/done something, 4) why they did it anyway, and 6) how the "victim" of the faux-pas felt. In the PCM analysis of the community sample, all but four items had a good fit to the model (i.e., 0.80> *infit* mean-square statistic [MNSQ] <1.20; see Table 3). The following items were subsequently excluded: Vignette 12: Item 4, Vignette 13: Item 4, Vignette 15: Item 3, and Vignette 18: Item 3. The remaining 26 items were found to fit the PCM model. Thus, the PCM model demonstrated a good fit for 26 of the 30 items, as indicated by the infit statistics. The model demonstrated good separation reliability, r = .81. Overall, our findings suggest that the FPT shows good properties and that the items load on a single dimension. Figure 2 shows the Wright map, which depicts the location of the item (and threshold) parameters as well as the distribution of person parameters along the latent dimension. Person-item maps are useful for comparing the range and position of the distribution of item measures (lower panel) with the range and position of the distribution of person measures (upper panel). In the community sample, correlation

analysis of the polytomous response format of the FPT targets (i.e., the person abilities as estimated based on the PCM) showed a moderate association with the RMET, r = .35, p = .007, [.103; .556].

In the PCM analysis of the BPD sample, all but 12 items had a good fit to the model (i.e., 0.80> *infit* mean-square statistic [MNSQ] <1.20; see Table 4). The following items were subsequently excluded: Vignette 7: Item 4; Vignette 11: Item 4 and Item 6; Vignette 12: Item 4; Vignette 13: Item 3 and Item 4; Vignette 14: Item 3; Vignette 15: Item 4; Vignette 16: Item 4; and Vignette 18: Item 3, Item 4, and Item 6. The remaining 18 items were found to fit the PCM model. Thus, the PCM model demonstrated a good fit for 18 of the 30 items, as indicated by the infit statistics. The model demonstrated good separation reliability, r = .82. Overall, our findings suggest that the FPT shows good properties and that the items load on a single dimension. Figure 3 depicts the Wright map that depicts the location of the item (and threshold) parameters as well as the distribution of person parameters along the latent dimension.

Table 3

Results of the PCM analysis of the community sample

No*	Vignette Description	Item	Diff.	MNSQ		Discr.	$\chi 2$
				Outfit	*Infit*		
2	Surprise Party Reveal	Item 3 c2	1.228	0.931	0.965	0.306	95.879
		Item 4 c1	-1.430	1.003	1.052	0.276	103.300
		Item 4 c2	1.291				
		Item 6 c1	-1.774	0.922	1.002	0.296	94.981
		Item 6 c2	1.312				
4	Ugly Curtains	Item 3 c1	-1.052	0.909	1.009	0.313	87.244
		Item 3 c2	3.148				
		Item 4 c1	-0.614	1.141	1.045	0.300	109.491
		Item 4 c2	0.783				
		Item 6 c1	-0.210	0.954	0.957	0.469	91.606
		Item 6 c2	0.000				
7	Lena is not a Boy	Item 3 c1	-1.067	0.826	0.891	0.493	75.131
		Item 3 c2	3.900				
		Item 4 c1	-1.000	0.832	0.891	0.500	75.692
		Item 4 c2	3.955				
		Item 6 c1	-0.599	0.920	0.953	0.496	83.723
		Item 6 c2	0.209				
11	Colleague is sick	Item 3 c1	-1.252	0.915	0.936	0.340	92.376
		Item 3 c2	2.548				
		Item 4 c1	-1.828	0.983	0.994	0.241	100.241
		Item 4 c2	0.277				
		Item 6 c1	-1.027	1.030	1.042	0.345	105.048
		Item 6 c2	-0.321				
12	Football Changing Room	Item 3 c1	-0.212	1.291	1.189	0.101	131.717
		Item 3 c2	2.574				
		Item 4 c1	-1.706	1.239	**1.260**	0.196	126.350
		Item 4 c2	-1.718				
		Item 6 c1	-1.981	1.161	1.189	0.226	118.421
		Item 6 c2	-2.050				

No*	Vignette Description	Item	Diff.	MNSQ		Discr.	χ^2
				Outfit	Infit		
13	Apple Pie	Item 3 c1	-1.192	0.831	0.862	0.520	85.604
		Item 3 c2	2.276				
		Item 4 c1	-1.678	1.109	**1.235**	0.539	114.181
		Item 6 c1	-1.665	1.040	1.175	0.431	107.122
		Item 6 c2	1.842				
14	Ugly Crystal Bowl	Item 3 c1	-1.326	0.839	0.910	0.460	84.756
		Item 4 c1	-2.383	0.829	0.841	0.421	83.711
		Item 4 c2	0.305				
		Item 6 c1	-1.991	0.852	0.862	0.303	86.069
		Item 6 c2	0.290				
15	Storytelling Competition	Item 3 c1	-0.746	1.245	**1.226**	0.153	127.003
		Item 3 c2	-0.332				
		Item 4 c1	-1.461	0.946	0.954	0.408	96.475
		Item 4 c2	-0.265				
		Item 6 c1	-1.902	0.863	0.903	0.388	88.066
		Item 6 c2	1.680				
16	Not a Waiter	Item 3 c1	-1.205	0.885	0.855	0.538	86.683
		Item 3 c2	3.896				
		Item 4 c1	-1.345	0.838	0.822	0.570	82.162
		Item 4 c2	3.777				
		Item 6 c1	-0.782	0.839	0.840	0.525	82.264
		Item 6 c2	2.335				
18	Lawyer	Item 3 c1	0.077	0.727	**0.761**	0.609	72.653
		Item 3 c2	2.154				
		Item 4 c1	-1.087	0.829	0.907	0.631	82.901
		Item 6 c1	-0.145				
		Item 6 c2	1.355				

Notes. *The numbers of the vignette correspond to the numbers of the original FPT, but we have only included targets; Diff. = difficulty; Discr. = discriminancy; χ^2 = fit statistic; $^*p < .05$; MNSQ = mean-squares statistic. Items with the infit MNSQ in boldface were removed due to misfit, as identified by values larger or smaller than $1 \pm .20$.

Figure 3

Wright map showing person abilities and item difficulties
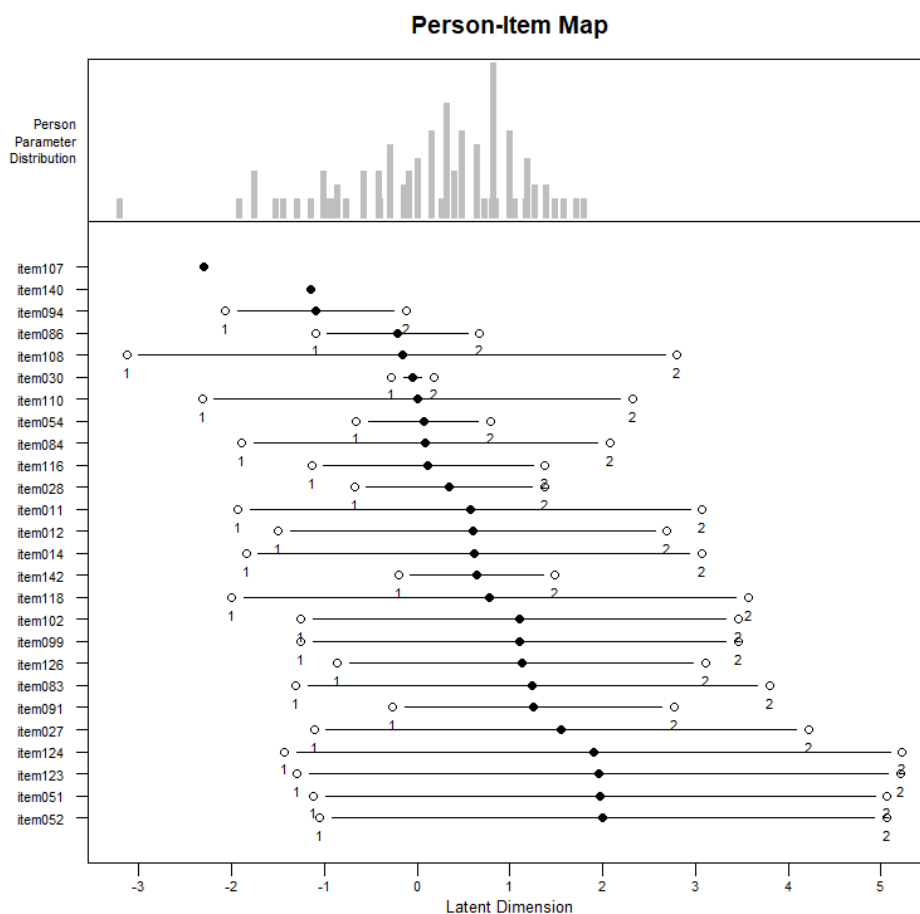
Table 4

Results of the PCM Analysis of the BPD Sample

No*	Vignette Description	Item	Diff.	MNSQ		Discr.	χ^2
				Outfit	*Infit*		
2	Surprise Party Reveal	Item 3 c2	3.685	1.064	1.023	0.323	44.675
		Item 4 c1	-0.981	1.199	1.146	0.247	50.373
		Item 4 c2	4.040				
		Item 6 c1	-1.125	0.901	0.901	0.361	37.833
		Item 6 c2	2.312				
4	Ugly Curtains	Item 3 c1	-1.899	0.890	0.866	0.364	36.492
		Item 3 c2	0.883				
		Item 4 c1	-0.975	1.053	1.066	0.278	43.162
		Item 4 c2	0.626				
		Item 6 c1	-1.704	1.014	1.088	0.346	41.568
		Item 6 c2	-1.566				
7	Lena is not a Boy	Item 3 c1	-0.192	0.855	0.877	0.550	31.639
		Item 4 c1	-0.034	0.642	**0.715**	0.726	23.767
		Item 6 c1	0.494	1.080	0.909	0.511	39.976
		Item 6 c2	1.492				
11	Colleague is sick	Item 3 c1	-0.563	1.275	1.166	0.289	49.735
		Item 4 c1	-1.582	0.716	**0.731**	0.478	27.935
		Item 4 c2	1.599				
		Item 6 c1	-0.118	0.679	**0.753**	0.601	26.493
		Item 6 c2	0.236				
12	Football Changing Room	Item 3 c1	-0.049	1.199	1.186	0.224	51.552
		Item 3 c2	3.738				
		Item 4 c1	-0.460	1.619	**1.408**	0.239	67.988
		Item 4 c2	-1.362				
		Item 6 c1	-1.300	0.905	1.003	0.409	38.903
		Item 6 c2	-1.912				

No*	Vignette Description	Item	Diff.	MNSQ		Discr.	χ^2
				Outfit	*Infit*		
13	Apple Pie	Item 3 c1	-1.461	1.861	**1.687**	0.259	76.301
		Item 4 c1	-0.983	1.448	**1.511**	0.026	59.374
		Item 4 c2	1.962				
		Item 6 c1	-0.493	0.983	0.966	0.401	40.311
		Item 6 c2	2.506				
14	Ugly Crystal Bowl	Item 3 c1	-1.944	1.807	**1.483**	0.473	74.098
		Item 4 c1	-2.163	0.931	1.010	0.489	38.157
		Item 4 c2	1.046				
		Item 6 c1	-1.443	0.906	0.887	0.472	37.148
		Item 6 c2	1.531				
15	Storytelling Competition	Item 3 c1	-0.971	0.977	0.913	0.322	0.467
		Item 3 c2	-0.299				
		Item 4 c1	-1.191	0.838	**0.787**	0.497	33.514
		Item 4 c2	-0.949				
		Item 6 c1	-2.130	1.142	1.069	0.123	45.672
		Item 6 c2	0.153				
16	Not a Waiter	Item 3 c1	-1.750	0.649	0.702	0.714	27.243
		Item 3 c2	3.235				
		Item 4 c1	-3.451	1.612	**3.202**	0.702	67.693
		Item 6 c1	-2.231	1.233	1.293	0.613	51.774
		Item 6 c2	0.979				
18	Lawyer	Item 3 c1	1.147	0.615	**0.693**	0.559	24.610
		Item 3 c2	2.825				
		Item 4 c1	0.248	0.698	**0.779**	0.625	27.916
		Item 6 c1	0.898	0.767	**0.781**	0.498	30.690
		Item 6 c2	3.027				

Notes. *The numbers of the vignette correspond to the numbers of the original FPT, but we have only included targets; Diff. = difficulty; Discr. = discriminancy; χ^2 = fit statistic; *p < .05; MNSQ = mean-squares statistics. Items with the infit MNSQ in boldface were removed due to misfit, as identified by values larger or smaller than 1 ± .20.

Figure 4
Wright map showing person abilities and item difficulties
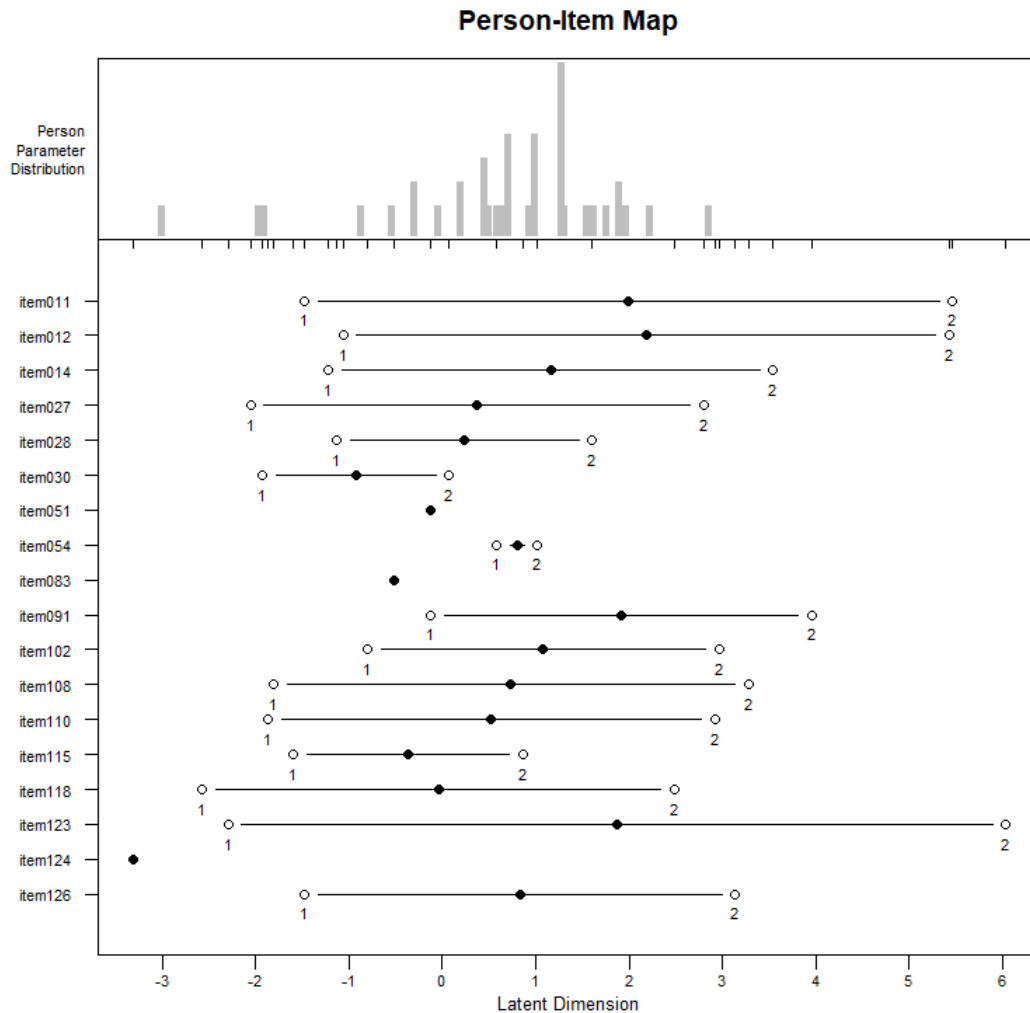
3.3   Discussion

This study aimed to investigate the FPT's ability to discriminate between degrees of ToM (i.e., ToM levels) and examine the psychometric properties of a polytomous response format. The implementation of a PCM allowed for a more nuanced assessment of ToM that went beyond the traditional binary classification of correct/incorrect responses. In addition, the study investigated whether the polytomous response format affected the validity of the FPT in the BPD sample compared to the community sample.

Our findings support the use of the polytomous response format, as evidenced by satisfactory PCM outcomes. The model demonstrated that the response categories effectively represented ascending levels of difficulty, suggesting that the FPT successfully differentiates between varying levels of ToM abilities. Thus, we were able to operationalize distinct ToM errors, affirming that the polytomous format captures subtle ToM distinctions.

This study marks the first validation of different types of ToM errors on the FPT, drawing inspiration from similar error distinctions applied in the MASC (e.g., Fretland et al., 2015; McLaren et al., 2022; Sharp & Vanwoerden, 2015). Through PCM, we mapped different levels of ToM, including no ToM, correct ToM, and exceeding ToM, to quantify the probability of participants demonstrating the highest ToM levels. This version of the FPT proved to be effective, indicating that ToM is indeed a variable construct that can be measured along a continuum. Three items (28, 91, and 115) were excluded in the community sample due to suboptimal fit, highlighting the value of retaining only well-fitting items to maintain reliability. Notably, more items showed a misfit in the BPD sample, leading to the conclusion that exceeding ToM responses in this group may have distorted the expected item-response patterns and thereby reduced overall model fit.

We further validated the FPT by correlating it with the RMET, a widely used ToM measure. The correlations between the RMET and FPT were moderate. Despite this association, the FPT remains valid due to the differing constructs it measures compared to the RMET. Previous factor analyses, such as Osterhaus et al. (2016) have shown that ToM encompasses multiple factors, with the RMET primarily measuring social reasoning and the FPT focusing on the violations of social norms. Additional research (Maleki et al., 2020) suggests that the RMET may primarily be decoding and

not reasoning as the FPT, as in the RMET, test participants are only required to decode the emotions of the protagonists, which is why some studies even suggest that the RMET is a measure of emotion recognition (Kittel et al., 2022; Quesque & Rossetti, 2020). The FPT, in contrast, requires participants to recognize and reason about faux-pas situations involving both mental state decoding and social norm understanding. Thus, the observed differences in associations are consistent with the distinct cognitive processes these tests engage.

3.4 Limitations and Future Directions

Even though this study supports the validity of the polytomous answer format of the FPT, some limitations must be addressed. First, only three items per vignette assessed ToM at levels beyond binary yes/no categories. Thus, the validation of the FPT cannot be expanded to the whole questionnaire. Future research should consider validating additional vignette items to further refine ToM measurement. Second, the community sample consisted of individuals from WEIRD backgrounds (Henrich et al., 2010), limiting generalizability. Future studies should aim to include a more diverse population when assessing community samples. Moreover, future studies should extend validation to clinical groups other than those with BPD, which includes known ToM deficits, such as individuals with autism spectrum conditions (Baron-Cohen, 2000) or schizophrenia (Sprong et al., 2007) to be able to draw conclusions about the entire population.

The limitation of the highly educated sample is related to a further limitation—the correlations with the RMET. The RMET is a relatively easy test that is unable to differentiate between individuals with high levels of ability (Black, 2019). Thus, it would be beneficial to use alternative ToM measures as an external criterion to further validate the response format in future studies. Additionally, future research could benefit from comparing the FPT with alternative ToM measures beyond the RMET, allowing for more comprehensive validation of the polytomous response format.

3.5 Conclusion

In summary, this study supports the validity of the FPT and its polytomous response format. The results demonstrate a satisfactory model fit, underscoring the FPT's psychometric robustness in capturing ToM at various levels. The FPT was shown to measure not only correct ToM but also distinguish among gradations of ToM ability. This study contributes to the literature on ToM by providing evidence for the FPT's

refined measurement capability, offering insights with implications for both theoretical understanding and applied ToM assessments.

4   Study 3: Differences in ToM Levels Across Community, MDD, and BPD Samples Using the Faux-Pas Recognition Test

To date, ToM deficits have been identified in various mental disorders, particularly in individuals with MDD (Berecz et al., 2016; Bora & Berk, 2016; Nestor et al., 2022; Pagnoni et al., 2022) and BPD (Bora, 2021; Németh et al., 2018), emphasizing their role as a transdiagnostic factor in psychopathology (Michelini et al., 2021; National Institutes of Mental Health, 2020). However, the literature presents inconsistencies - particularly in how ToM impairments are measured and the types of ToM deficits identified.

For instance, meta-analyses examining ToM deficits in MDD report moderate to severe ToM impairments, but the findings vary depending on the specific ToM measures used. MDD is characterized by persistent low mood, loss of pleasure and interest, impaired cognitive function, and vegetative symptoms, including disturbed sleep or appetite (Otte et al., 2016). Richman and Unoka (2015), who only included studies in their meta-analysis that employed the RMET, reported large ToM impairments. In contrast, Bora and Berk (2016) and Nestor et al. (2022) reported moderate ToM impairments in MDD in their meta-analyses but included studies that employed various ToM tasks such as the FPT, the RMET, and the MASC.

Furthermore, the existing literature on MDD demonstrates a lack of consistency in the specific levels of ToM impairments observed. ToM impairments can vary between no/less ToM and exceeding ToM. Mental states can therefore be either oversimplified (no/less ToM) or over-ascribed (exceeding ToM), which can result in attributing intentions to random behavior (e.g., Vegni et al., 2021). Some studies indicate that MDD is associated with less ToM (De Coninck et al., 2021; Safiye et al., 2023; Scandurra et al., 2020; Wolkenstein et al., 2011), while the studies included in the meta-analysis by McLaren et al. (2022) showed small effect sizes for exceeding ToM deficits, complicating the understanding of how MDD affects the interpretation of others' mental states. Thus, the current literature is unclear about whether individuals with MDD tend to over- or under-interpret the mental states of others. However, it should be noted that the aforementioned studies relied only on the RFQ and the MASC to assess ToM deficits. The validity of the RFQ as a measure of exceeding ToM has been called into question (Müller et al., 2022). Although the MASC is a reliable and valid instrument (Fossati et al., 2018), its specific ToM error types have yet to undergo

comprehensive validation. The questionable validity of the measurement instruments used could also explain the variability of the results. This variability of results again emphasizes the need for a validated instrument that can reliably capture the nuances of ToM impairments in different mental disorders.

Similar inconsistencies have been found in BPD research. BPD is characterized by significant challenges in social and interpersonal interactions (American Psychiatric Association, 2013; Lazarus et al., 2014). While individuals with BPD demonstrate a basic understanding of ToM, as assessed by the RMET (Baez et al., 2015; Németh et al., 2018; Pourmohammad et al., 2021; Zabihzadeh et al., 2017), they show deficits in AToM abilities, particularly when identifying emotional cues in complex social situations, as assessed by the FPT (e.g., Baez et al., 2015; Harari et al., 2010; Petersen et al., 2016; Zabihzadeh et al., 2017). One study observed that while BPD patients performed comparably to HC on basic FB tasks, they made significantly more ToM errors in faux-pas understanding than HC (Petersen et al., 2016). Such discrepancies might arise from the FPT's ability to assess the perception of others' emotions in nuanced, socially complex situations that mimic real-life contexts. These scenarios demand a deeper contextual understanding, which is particularly sensitive for detecting subtle ToM impairments in individuals with BPD.

A more detailed look at deficits of specific ToM levels reveals even greater inconsistency. Exceeding ToM is frequently cited as a core feature of BPD (Bora, 2021; Sharp & Vanwoerden, 2015), but a recent meta-analysis has shown that this phenomenon is not exclusive to BPD and is seen across a range of mental disorders (McLaren et al., 2022). Notably, effect sizes for exceeding ToM errors were larger in BPD patients than in those with MDD. The finding that exceeding ToM is associated with general psychopathology aligns with dimensional models of mental disorders, such as the RDoC (Michelini et al., 2021; National Institutes of Mental Health, 2020) or the HiTOP (Conway et al., 2019).

As individuals with BPD often suffer from comorbid symptoms of depression, ToM deficits have also been studied in BPD patients with comorbid MDD. However, the results are again inconsistent. An earlier meta-analysis revealed that BPD patients with comorbid MDD show better ToM abilities than patients with BPD alone (Richman & Unoka, 2015). This contrasts with a later study that reported lower ToM deficits in patients with BPD and MDD (Zabihzadeh et al., 2017).

These inconsistencies underscore the need for a clearer understanding of how MDD and BPD interact in relation to ToM deficits, particularly with regard to specific ToM levels, which may not be consistently captured by the existing ToM measures. Prior studies have often relied on general ToM abilities rather than distinguishing between various ToM levels. Furthermore, past research on ToM impairments in BPD, MDD, and non-clinical control groups has yielded inconsistent results, partly due to the use of inadequately validated measurement tools, such as the RFQ and the MASC. These inconsistencies in findings raise questions about the reliability of previous findings, suggesting the need for a more precise instrument capable of capturing the nuances of ToM deficits.

This study aims to clarify these inconsistencies by employing the FPT, which uniquely includes validated response categories for different ToM levels, categorized as no ToM, correct ToM, and exceeding ToM. It is the first study to compare the different ToM levels of the FPT across individuals with BPD or MDD and a community sample. By examining specific ToM levels, the study aims to clarify whether unique aspects of ToM are impaired in either BPD or MDD, as compared to a community sample. Such detailed comparisons could resolve contradictory findings in previous research, yielding more precise insights into ToM impairments specific to each disorder. The present study also aims to investigate the influence of depression severity when comparing ToM performance between individuals with MDD and BPD. Given that MDD is often a comorbid condition of BPD, examining the role of depression severity can help disentangle ToM impairments specific to BPD from those potentially attributable to depressive symptoms. This approach aims to clarify the overlapping and distinct social-cognitive deficits associated with BPD and MDD by examining both disorder-specific ToM impairments and the potential influence of depression severity. Thus, the study contributes to a deeper understanding of transdiagnostic mechanisms. These insights align with transdiagnostic frameworks, such as RDoC or HiTOP, which emphasize dimensional approaches to understanding psychopathology (Conway et al., 2019; Michelini et al., 2021). Moreover, we correlated the MASC's subscales with the FPT's subscales to gain a deeper understanding of previous findings and clarify the relationship between different dimensions of ToM assessed by these tools. Given prior findings that ToM abilities can vary by biological sex and age, showing that women may perform better than men (Baron-Cohen, 2002; Baron-Cohen et al., 2015) and that

ToM abilities might decline with age among healthy adults (Henry et al., 2013), we controlled for these demographic factors in the analyses.

4.1 Method

4.1.1 Participants

The study analyzed data from 150 participants (62 men, 88 women). Patients' ages ranged from 18 to 73 years, with an average of 39.89 (SD = 14.79). The total sample consisted of a community sample (N = 58) and a clinical sample (N = 92).

The clinical sample consisted of inpatients from a psychiatric-psychosomatic hospital in Austria. The patients were treated for MDD or BPD in specialized treatment units based on their psychiatric diagnosis. The hospital required a minimum age of 18 years and basic conversational skills in German for admission. Patients who exhibit acute psychotic symptoms, current suicidal behavior, or acute intoxication are not admitted to the hospital and are therefore not included in the study. Additionally, we excluded patients who did not currently meet the respective criteria for MDD or BPD according to standardized self-report measures.

Initially, the community sample consisted of 109 participants recruited from the general population. Forty-one data sets were incomplete and were therefore excluded. Moreover, we excluded participants who reported currently receiving psychiatric, psychological, or psychotherapeutic treatment and those who had a positive PTSD screening. Additionally, one participant who identified as nonbinary was excluded to avoid overestimation of sex-based effects and to ensure the integrity of comparisons based on sex. In total, the final dataset included 150 participants, with two statistically independent clinical samples (MDD and BPD) and a community sample. Comorbidities, however, were not excluded. Table 5 provides an overview of the demographic characteristics across the three groups. The gender distribution differed among the samples, with more women than men in both the BPD and MDD samples, while the community sample had a higher proportion of men. Education levels were generally lower in the clinical samples compared to the community sample. Regarding age, BPD patients were the youngest, while MDD patients were the oldest. Relationship status also varied: most participants in the community sample reported having a partner, whereas the majority of individuals in the BPD group reported being single.

We conducted a post-hoc power analysis with G*Power (Faul et al., 2007) to assess whether the study was adequately powered to detect the observed effect size. For the regression model, with a sample size of N=150 and 5 predictors and a significance level of α = 0.05, the statistical power was calculated as 0.250. This indicates a low probability (24.97%) of correctly rejecting the null hypothesis, with a high likelihood (75.03%) of committing a Type II error. The low power indicates that the study's design was inadequate for reliably detecting small effect sizes.

Table 5

Sociodemographic Variables of the Community Sample and Patients Diagnosed with BPD and MDD

	BPD diagnosis N = 44 N (%)	MDD diagnosis N = 48 N (%)	Community sample N = 58 N (%)	H (df)	*p*
Sex				18.31 (2)	< .001
Men	8 (18.18)	19 (39.58)	35 (60.34)		
Women	36 (81.82)	29 (60.42)	23 (39.65)		
Education level	a	b	c	7.49 (2)	.024
Low education	19 (43.18)	19 (39.58)	9 (15.52)		
Medium/high education	23 (52.27)	25 (52.08)	36 (62.07)		
Marital status	d	e		11.65 (2)	.003
Single	26 (59.09)	18 (37.50)	19 (32.76)		
In a partnership	12 (27.27)	21 (43.75)	39 (67.24)		
	M (SD)	**M (SD)**	**M (SD)**	**H (df)**	**P2**
Age	34.39 (12.52)	44.56 (13.94)	40.21 (15.88)	11.42 (2)	.003

Note. BPD = borderline personality disorder; MDD = major depressive disorder; ToM = Theory of Mind, 95% CI = 95% confidence interval; [a] 2 missing values, [b] 4 missing values, [c] 13 missing values, [d] 6 missing values, [e] 9 missing values; bold p-values are *p* < .05.

4.1.2 Procedure

6.1.2.1 Community Sample. We collected the data for the community sample between June 2021 and September 2021. The study employed the SoSci-Survey platform to administer the FPT. Furthermore, the MASC was presented to the community sample, using the Computer-based Health Evaluation Software (CHES; Holzner et al., 2012). The initial sample consisted of family members, relatives, and acquaintances from the personal environment of a master's student, who were then encouraged to disseminate the study further to other individuals via various social media groups on Facebook, WhatsApp, and other online platforms.

6.1.2.2 Clinical Samples. We collected data from the clinical samples between December 2020 and January 2021 as part of routine clinical care. All data presented

in this study were collected during routine outcome monitoring at the time of admission. Patients completed the questionnaires in a computer assessment room comprising eight separate cubicles. The assessment lasted two hours, divided into two one-hour sessions. The study used CHES, an electronic outcome monitoring software tool, to implement the questionnaires (Holzner et al., 2012).

The participants in all samples provided written informed consent. The study was carried out in compliance with the ethical principles outlined in the Declaration of Helsinki (World-Medical-Association, 2013) and received approval from the ethics committee of the Karl Landsteiner University of Health Sciences (No:1004/2021; 1010/2023).

4.1.3 Measures

6.1.3.1 The Faux-Pas Recognition Test (FPT). We have created a multiple-choice answer format (LINK) for the German adult version of the FPT (Ströbele, n.d.). The FPT is described in Section 1.4.1.3.

6.1.3.2 The Movie for the Assessment of Social Cognition (MASC). We also administered the MASC to assess ToM (Dziobek et al., 2006). The MASC is described in Section 1.4.2.5**.**

6.1.3.3 PTSD Screening. We used the PC-PTSD-5 (PC-PTSD-5; Prins et al., 2016) to screen for PTSD. The screening tool comprises five items and is designed to assess the occurrence of specific PTSD symptoms, including nightmares, avoidance behavior, hyperarousal, emotional blunting, and alienation, following an event. A cut-off score of 4 was used (Prins et al., 2016). The PC-PTSD-5 has recently demonstrated high internal consistency (α = .83; Cheng et al., 2021).

6.1.3.4 Borderline Symptoms. To ensure that patients from the specialized treatment unit exhibited at least moderate symptoms of BPD, we utilized the German version of the Borderline Symptom-List (BSL-23; Bohus et al., 2001). The BSL-23 is a 23-item assessment tool that measures borderline symptomatology in accordance with the DSM-IV criteria (Bohus et al., 2009). We selected patients with a mean symptom severity score of $\geq$ 1.07, representing at least moderate symptom severity (Kleindienst et al., 2020). The internal consistency of the BSL-23 has previously demonstrated high internal consistency (α = .94 -.97; Wolf et al., 2009).

6.1.3.5 Depression symptoms. To determine whether patients in the inpatient sample currently exhibited substantial symptoms of depression, we utilized the German version of the Patient Health Questionnaire (PHQ-9; Kroenke et al., 2001), which is based on the DSM-IV (American Psychiatric Association, 1994). In order to ensure the most accurate results, we used a PHQ-9 score of 10 or above (Kroenke et al., 2001) as our inclusion criteria. In previous studies, the German version of the PHQ-9 demonstrated robust psychometric properties, including high internal consistency (α = .88; Gräfe et al., 2004).

6.1.3.6 Sociodemographic characteristics. We recorded the participants' self-reported gender, age in years, current relationship status (living in a partnership, yes vs. no), and education level. The gender of participants was assessed polytomous with the options female, male, and diverse. The level of education was determined in accordance with the Austrian education system and categorized into two groups: low level of education (i.e., compulsory school) and medium/high level of education (A-levels or university degree).

4.1.4 *Data analysis*

All analyses for the study were performed using IBM SPSS (Version 29). First, to assess sociodemographic differences, we calculated the Kruskal-Wallis tests. Then, we calculated correlations to compare the subscales of the FPT with the subscales of the MASC to gain a better understanding of the results of previous studies. Afterwards, to examine associations between the three samples and ToM outcomes, we conducted multiple linear regression analyses. We chose this approach since the assumptions required for a multivariate analysis of covariance (MANCOVA) were not met. In the first regression analysis, we compared the BPD and the MDD sample with the community sample with regard to the ToM levels. We conducted three separate regression analyses with the same predictors but different dependent variables. The independent variables included diagnostic group (BPD vs. community sample and MDD vs. community sample), and age and gender as covariates. The dependent variables were the three ToM levels respectively (i.e., no ToM, correct ToM, and exceeding ToM). This method allowed for the investigation of specific predictors while controlling for potential confounding variables. In the second regression analysis, we compared the BPD with the MDD sample with regard to the ToM levels. The independent variables included the diagnostic group (BPD vs. MDD sample), age, gender, and depression severity as

covariates. We only employed the BPD and the MDD samples in this analysis, since we did not assess depression symptoms in the community sample. The dependent variables were again the three ToM levels respectively (i.e., no ToM, correct ToM, and exceeding ToM). The alpha level was set to p = 0.05 for all tests.

4.2 Results

Table 6 summarizes the correlations between the subscales of the MASC and the FPT. Interestingly, the FPT's no ToM subscale showed no association with either the no ToM or less ToM subscale of the MASC. However, significant relationships were observed between the correct and exceeding ToM subscales of both measures. Table 7 presents the findings from the first analyses. The first regression analysis examined the associations between BPD and MDD diagnoses versus a community sample and different ToM levels, controlling for sex and age. It explored associations between BPD and MDD diagnoses with no ToM, controlling for demographic factors. Results showed no significant association between either diagnosis and no ToM, suggesting that neither BPD nor MDD was predictive of a complete lack of ToM ability. The second regression analysis investigated associations with correct ToM, again controlling for age and sex. This analysis revealed a significant negative association between MDD diagnosis and correct ToM performance, with individuals diagnosed with MDD significantly underperforming the community sample. However, BPD diagnosis showed no such association, suggesting that individuals with BPD performed similarly to the community sample in terms of correct ToM responses. The third regression analysis examined the association between BPD and MDD diagnoses and exceeding ToM levels. Results indicated a significant positive association between BPD diagnosis and exceeding ToM, with individuals diagnosed with BPD making significantly more exceeding ToM errors than the community sample. No such association was found for MDD. Across all three analyses, neither age nor sex had a significant impact on the outcomes.

The second regression analysis examined associations between BPD versus MDD diagnoses and different ToM levels, controlling for sex, age, and depression severity. Table 8 displays the regression analyses. The regression showed that BPD was associated with exceeding ToM, even after controlling for depression severity. Additionally, sex was negatively associated with correct ToM, both with and without

controlling for depressive symptoms. Thus, female participants in the BPD and MDD samples were more likely to provide correct ToM responses compared to males.

Table 6

Correlation coefficients (Pearson) for no ToM and exceeding ToM errors and correct responses of the MASC and the FPT in the community sample

	MASC no ToM	MASC less ToM	MASC correct ToM	MASC exceeding ToM
FPT no ToM	.15	-.01	-.11	.14
FPT correct ToM	-.41	-.33*	.53**	.48**
FPT exceeding ToM	.24	.21	-.45*	.49**

Note. *p < .05 **p < .01

Table 7

Regressions of Diagnoses, Sex, and Age on ToM Correct Responses, No ToM and Exceeding ToM Errors

Predictors	*B*	*SE*	*t*	*p*	95% CI
			exceeding ToM		
Regression 1					
Age in years	.01	.02	.35	.724	[-.03, .04]
Sex (0 = female, 1 = male)	.05	.51	.10	.919	[-.95, 1.05]
BPD diagnosis vs. community sample	**1.56**	**.61**	**2.58**	**.011**	**[.36, 2.76]**
MDD diagnosis vs. community sample	.18	.53	.33	.740	[-.87,1.23]
			correct ToM		
Regression 2					
Age in years	-.00	.04	-.12	.902	[-.08, .07]
Sex (0 = female, 1 = male)	-1.55	1.10	-1.40	.163	[-3.73, .63]
BPD diagnosis vs. community sample	1.66	1.32	1.25	.212	[-.95, 4.27]
MDD diagnosis vs. community sample	**-3.61**	**1.16**	**-3.11**	**.002**	**[-5.89, -1.32]**
			no ToM		
Regression 3					
Age in years	.01	.01	1.90	.059	[-.00, .03]
Sex (0 = female, 1 = male)	.39	.21	1.84	.067	[-.03, .82]
BPD diagnosis vs. community sample	.29	.26	1.14	.258	[-.22, .80]
MDD diagnosis vs. community sample	.14	.23	.63	.530	[-.30, .59]

Note. BPD = borderline personality disorder; MDD = major depressive disorder; ToM = Theory of Mind, CI = confidence interval; bold p-values are *p* < .05.

Table 8

Regressions of diagnosis (BPD vs MDD), depression severity, sex, and age on ToM correct responses, no ToM, and exceeding ToM errors

| Predictors | exceeding ToM | | | | |
	B	*SE*	*t*	*p*	95% CI
Regression 1 (enter)					
Age in years	.02	.03	.60	.552	[-.04, .07]
Sex (0 = female, 1 = male)	.01	.75	.01	.994	[-1.48, 1.50]
BPD vs. MDD diagnosis	1.70	.73	2.32	.023	[.24, 3.15]
Regression 2 (enter)					
Age in years	.02	.03	.61	.543	[-.04, .07]
Sex (0 = female, 1 = male)	-05	.76	-.07	.948	[-1.55, 1.45]
Depression severity	.06	.08	.69	.494	[-.11, .22]
BPD vs. MDD diagnosis	**1.60**	**.75**	**2.15**	**.035**	**[-.11, .22]**
	correct ToM				
	B	*SE*	*t*	*p*	95% CI
Regression 1 (enter)					
Age in years	-.03	.05	-.51	.613	[-.13, .07]
Sex (0 = female, 1 = male)	**-3.94**	**1.49**	**-2.65**	**.010**	**[-6.90, -.98]**
BPD vs MDD diagnosis	1.50	1.45	1.03	.304	[-1.39, 4.39]
Regression 2 (enter)					
Age in years	-.03	.05	-.51	.608	[-.13, .07]
Sex (0 = female, 1 = male)	**-3.88**	**1.51**	**-2.58**	**.012**	**[-6.88, -.89]**
Depression severity	-.06	.16	-.39	.701	[-.38, .26]
BPD vs MDD diagnosis	1.61	1.49	1.08	.282	[-1.35, 4.56]
	no ToM				
	B	*SE*	*t*	*p*	95% CI
Regression 1 (enter)					
Age in years	.01	.01	1.25	.216	[-.01, .03]
Sex (0 = female, 1 = male)	.30	.30	.99	.325	[-.30, .90]
BPD vs MDD diagnosis	.26	.29	.88	.381	[-.33, .84]
Regression 2 (enter)					
Age in years	.01	.01	1.24	.220	[-.01, .03]
Sex (0 = female, 1 = male)	.30	.31	.99	.327	[-.31, .91]
Depression severity	-.00	.03	-.06	.950	[-.07, .06]
BPD vs MDD diagnosis	.26	.30	.87	.385	[-.34, .86]

Note. BPD = borderline personality disorder; MDD = major depressive disorder; ToM = Theory of Mind, CI = confidence interval; bold p-values are *p* < .05.

4.3 Discussion

This study is the first to examine differences in ToM levels among individuals with BPD, MDD, and a community sample, using a newly validated response format of the FPT. By incorporating various ToM levels, this study aimed to clarify previous inconsistencies in ToM findings across these mental disorders. Results indicated that individuals with BPD were significantly more likely to exhibit exceeding ToM, compared to the community sample. In contrast, MDD was not associated with this type of ToM error. Individuals with MDD underperformed the community sample on correct ToM.

First, analyses of the current study show that BPD is associated with exceeding ToM. Exceeding ToM refers to the tendency to over-interpret behaviors by attributing intentions that others typically would not perceive. The results support previous research that has established exceeding ToM as a core feature of BPD (Bora, 2021; Sharp & Vanwoerden, 2015) and extends this association by employing a new, validated measure to examine this association. As expected, the study found no significant differences between individuals with BPD and the community sample in correct ToM or no ToM, which supports prior research suggesting that the ToM difficulties of individuals with BPD are specific to exceeding ToM and not to all aspects of ToM (Bora, 2021). Further analyses even indicated a unique relationship between BPD and exceeding ToM, independent of depression severity. The meta-analysis by Németh et al. (2018) did not explicitly assess exceeding ToM; however, their findings support the notion that overall ToM differences between BPD and HC are not significantly influenced by the current depression rate within individuals with BPD. This finding is consistent with the results of the present study, which suggests that ToM characteristics associated with BPD may reflect unique traits of the disorder. The reasons for exceeding ToM in BPD are probably due to several interacting factors. Individuals with BPD often exhibit heightened sensitivity to social cues (Frick et al., 2012) and a negativity or anger bias in emotion recognition (Domes et al., 2009). These tendencies can lead individuals with BPD to perceive hostility or intent where none exists (i.e., exceeding ToM). Additionally, difficulties in emotion regulation, common in BPD, have been shown to partially mediate the relationship between exceeding ToM and BPD traits (Sharp et al., 2011). Consequently, exceeding ToM may exacerbate interpersonal challenges and difficulties of individuals with BPD, as they often struggle

to accurately interpret others' intentions, leading to misunderstandings and conflict in social interactions (Sharp & Vanwoerden, 2015).

Secondly, the current study did not find differences between MDD and the community sample regarding exceeding ToM. Therefore, the findings do not support the idea that exceeding ToM is broadly linked to psychopathology (McLaren et al., 2022). It is important to note, however, that in the meta-analysis by McLaren et al. (2022), the effect size for exceeding ToM was smaller for MDD than for BPD. This suggests that individuals with MDD show less exceeding ToM than individuals with BPD. Nevertheless, the present study has identified ToM impairments in individuals with MDD. Regarding impaired overall ToM ability (correct ToM), the results of the present study confirm that community samples outperform individuals with MDD regarding correct ToM. The results thus are in line with previous studies' findings that overall ToM ability is impaired in MDD patients (Bora & Berk, 2016; Nestor et al., 2022; Richman & Unoka, 2015). A number of factors may be responsible for this impairment. First, individuals with MDD often disengage from social interactions with others, whether due to low mood, fatigue, or other depressive symptoms (Porcelli et al., 2019). This may, in turn, result in a reduction in the frequency of interpersonal interactions. As a result, individuals with MDD may become less proficient in identifying the thoughts and feelings of others, due to a lack of practice. Nestor et al. (2022) posit that participation in social interactions may enhance one's capacity to comprehend the emotional and mental states of others. Secondly, research indicates that individuals with MDD often display a tendency towards negative self-focus or self-directed attention (Mor & Winquist, 2002). This inclination towards introspection and negative reflection on one's own thoughts and feelings can diminish or hinder one's capacity for empathy and engagement in understanding the mental and emotional states of others (Nestor et al., 2022). The predominantly negative outlook of individuals with MDD can contribute to misinterpretations and inaccuracies in recognizing others' cognitive and emotional states. Furthermore, individuals with MDD may perceive the world as aversive due to the presence of negative preconceptions. For instance, individuals with MDD may misjudge the mental states of others due to their tendency to view others through a lens characterized by negativity.

The study did not replicate prior findings that suggest individuals with MDD tend to under-interpret others' mental states (De Coninck et al., 2021; Safiye et al., 2023;

Scandurra et al., 2020; Wolkenstein et al., 2011). However, it is important to note that these previous studies used the RFQ or the MASC, which focus on different aspects of ToM than the FPT. The RFQ asks respondents to endorse or reject statements related to mentalizing processes (Müller et al., 2022). While the MASC tests the ability to interpret social situations by integrating multimodal cues such as facial expression, gestures, and prosody (Dziobek et al., 2006), the FPT measures the ability to recognize social mistakes or faux-pas within structured narratives (Osterhaus & Bosacki, 2022). Given that MDD is associated with cognitive slowing and reduced sensitivity to social cues (Jin et al., 2015), individuals with MDD may struggle more with tasks like the MASC, which require rapid processing of complex, dynamic social information. In contrast, the FPT relies on logical inference and verbal reasoning, which may explain why MDD patients do not exhibit the same under-interpretation on this task. Furthermore, according to Study 2, the no ToM level of the FPT is not associated with either the no or the less ToM level of the MASC, which also suggests that these two tests measure different aspects of ToM. This distinction highlights the task-dependent nature of ToM impairments in MDD and suggests that deficits may vary based on the specific cognitive demands of the ToM measure used.

The observed negative association between gender and correct ToM also provides important insights. Male participants in the BPD and MDD groups were less accurate in correctly identifying mental states compared to females, aligning with research that suggests females generally perform better on ToM tasks (Baron-Cohen, 2002; Henry et al., 2013b, p. 201). This finding underscores the need for further investigation into gender differences in ToM abilities within clinical populations and highlights gender as a potential moderating factor in social-cognitive impairments.

Notably, the FPT uniquely assesses 'no ToM' errors, which do not align with the 'no ToM' or 'less ToM' subscales of the MASC. This suggests that the FPT may capture distinct aspects of under-attributing mental states to others compared to the MASC. Specifically, the FPT focuses on the recognition and interpretation of social blunders, emphasizing the ability to detect and understand unintentional social missteps. In contrast, the MASC evaluates the comprehension of complex social interactions by presenting nuanced social cues through film, incorporating both verbal and non-verbal elements such as facial expressions and tone of voice. This distinction implies that the MASC might identify forms of under-ascribing mental states that are not addressed by

the FPT. This finding might explain why studies using the MASC indicated that exceeding ToM is associated with general psychopathology Understanding these differences could provide deeper insights into the findings of previous studies, underscoring the need to consider task-specific characteristics when interpreting ToM-related impairments.

Moreover, the implications for clinical practice are substantial. A clearer understanding of the distinct ToM deficits in MDD and BPD could significantly enhance treatment planning by allowing clinicians to focus on specific social-cognitive interventions. Personalized treatment strategies that target these areas could ultimately improve outcomes for individuals with these disorders. This is particularly relevant in the context of mentalization-based therapy (MBT), which has proven effective in treating BPD and holds promise for treating other disorders like MDD (Luyten et al., 2020; Malda-Castillo et al., 2019). By better understanding the specific ToM deficits in these disorders, targeted interventions could be developed to address the specific needs of these clinical populations, ultimately improving treatment outcomes.

4.4   Conclusion

The present study reveals distinct differences in ToM levels, taking into account age and gender, between individuals with BPD, MDD, and a community sample and represents the first comparison of specific ToM levels with the FPT. The results indicate that exceeding ToM is associated with BPD regardless of depression severity, whereas MDD is characterized by impairments in correct ToM. By demonstrating these disorder-specific ToM deficits, this study emphasizes the value of using validated ToM measures, such as the FPT, to accurately assess and differentiate ToM deficits across clinical populations. Given the relatively small sample sizes, future studies should consider increasing the sample size to improve power and reduce the risk of missing significant effects.

5   General Discussion

This dissertation aimed to contribute to the growing body of ToM research by addressing key gaps in the field — particularly in the measurement of ToM and its impairments in clinical populations. Three studies extended ToM understanding with the FPT by validating the new closed response formats to obtain a differentiated understanding of ToM deficits in individuals with BPD and MDD and a community sample. The results of these studies have implications for the measurement of ToM and the understanding of ToM deficits in mental disorders.

5.1   Summary of the Main Findings

5.1.1   Summary of Study 1 - Validation of the Dichotomous Answer Format

Study 1 investigated the validity of the newly developed dichotomous answer format of the FPT in a community sample. Specifically, the study assessed the ability of the FPT to measure a single underlying construct (ToM) and estimated the difficulty of its items. The results confirmed that the dichotomous answer format of the FPT has been successfully operationalized to measure a single construct. Moreover, the findings showed a strong alignment between the model and the data, indicating the FPT's psychometric suitability and accuracy in measurement. The FPT is therefore able to measure correct ToM reliably and accurately. The FPT demonstrated strong correlations with both the MASC and the RMET. This level of correlation suggests that while the tests measure related constructs, they also capture distinct aspects of ToM, emphasizing its value as a unique tool for ToM assessment. The findings confirm the FPT as a valuable instrument for assessing ToM and provide meaningful contributions to the theoretical understanding and practical application of ToM measurements. Considering these results, the dichotomous version of the FPT represents a robust instrument for capturing ToM abilities, complementing existing measures, and enhancing research into ToM constructs across various contexts.

5.1.2   Summary of Study 2 - Validation of the Polytomous Answer Format

Study 2 validated the new polytomous answer format of the FPT in both a community and a clinical sample of individuals with BPD. Moreover, the study assessed correlations between the FPT and the RMET. The innovative polytomous answer format classifies responses into three distinct ToM levels, no ToM, correct ToM, and exceeding ToM, allowing for a more nuanced assessment than traditional binary measures. Results confirmed that the polytomous format effectively captures

ascending levels of difficulty, establishing the FPT as a reliable tool for differentiating ToM abilities. Results for the BPD sample were slightly worse, consistent with the prevalence of exceeding ToM errors in this group, which reduced response variability. By validating different ToM levels of the FPT, Study 2 contributes to a deeper understanding of ToM deficits. This research represents the first effort to validate various ToM levels, which until now have primarily been examined using either insufficiently validated questionnaires or the MASC (e.g., Fretland et al., 2015; McLaren et al., 2022; Sharp & Vanwoerden, 2015), a tool that has not yet been validated for ToM errors. The FPT's polytomous format captures a broader range of ToM abilities, thereby enhancing the assessment of ToM and advancing our understanding of related deficits. While the FPT demonstrated only moderate correlations with the RMET, this difference does not detract from its validity, as the two measures target distinct aspects of ToM.

5.1.3 Summary of Study 3 - Differences in ToM Abilities Across Groups

Study 3 investigated ToM levels (no ToM, correct ToM, and exceeding ToM) using the newly validated polytomous answer format of the FPT. Specifically, the study investigated differences in ToM levels among individuals with BPD or MDD and a community sample, as well as differences in ToM levels between individuals with BPD and MDD, taking depression symptoms into account. By incorporating distinct ToM levels, the study aimed to clarify previous inconsistencies in ToM research by exploring distinct ToM impairments in these mental disorders. Results indicated that exceeding ToM was uniquely associated with BPD, independent of depression severity. Conversely, MDD was not associated with exceeding ToM. Individuals with MDD did, however, exhibit deficits in correct ToM. The results align with prior research on a general ToM impairment in MDD; however, they challenge the notion that exceeding ToM is associated with general psychopathology (McLaren et al., 2022). These findings highlight the task-dependent nature of ToM deficits and underscore the FPT's value in detecting disorder-specific ToM impairments. Notably, the FPT's correct ToM and exceeding ToM subscales showed strong correlations with the MASC's corresponding subscales, supporting its convergent validity. However, the FPT's no ToM subscale was not associated with the MASC's no or less ToM subscales, reflecting differences in the constructs each test measures. In sum, the findings emphasize the necessity of using validated measurement tools and underscore its potential in guiding targeted

interventions to address the unique social-cognitive challenges faced by individuals with BPD and MDD.

5.2 Discussion of the Findings of the FPT

The FPT was a central measure in this research, and its findings contribute significantly to understanding how ToM can be operationalized and assessed. This section discusses the implications of the FPT in its dichotomous and polytomous response formats, its comparison with other ToM measures, and its broader implications as a tool for assessing ToM.

5.2.1 Implications for the Dichotomous Format of the FPT

The validation of the dichotomous answer format of the FPT in Study 1 represents a key advancement in its utility as a ToM measure. The dichotomous format simplifies the response process, making it suitable when distinguishing between basic ToM responses (correct/incorrect) is sufficient. The design of the dichotomous answer format of the FPT reduces cognitive load, making it particularly useful for time-efficient assessments in research settings and clinical contexts. In clinical contexts, the dichotomous response format improves applicability, as more complex response formats can be challenging in some populations (e.g., individuals with cognitive impairment or high symptom severity). In these populations, the simplified format retains its reliability in capturing ToM abilities without the need for more complex categorizations. Moreover, the dichotomous format enhances the accessibility and practicality of the FPT in large-scale studies or community samples, where a simpler answer format is sufficient to measure ToM abilities effectively. Taken together, the findings of Study 1 emphasize the use of specific formats of ToM tests, depending on the sample.

5.2.2 Discussion of the Findings on Different ToM levels

The polytomous format that was validated in Study 2 has been specifically developed for use in clinical populations. Inspired by the MASC (Dziobek et al., 2006), this format allows for a more nuanced assessment of ToM abilities, classified as no ToM, correct ToM, and exceeding ToM. This refinement moves beyond the binary approach of the dichotomous format, providing deeper insights into the dimensional nature of ToM. To validate the nuanced assessment of ToM deficits, Study 2 conceptualized ToM as a dimensional construct, despite distinguishing between response categories such as no ToM, correct ToM, and exceeding ToM. For instance,

correct ToM reflects an intermediate ToM ability, while exceeding ToM represents a higher level along the continuum that is characterized by over-interpretation of mental states. No ToM, in turn, represents a lower level along the continuum that is characterized by under-interpretation of mental states. This approach aligns with the RDoC framework, which emphasizes dimensional constructs cutting across traditional diagnostic boundaries. The ToM level categories are not rigid classifications but serve as practical markers for interpreting positions along a continuum, with response categories representing thresholds along this dimension. In summary, the ability of the FPT to differentiate these levels provides valuable insights into the heterogeneity of ToM impairments within clinical groups, making it an indispensable tool for both research and clinical diagnostics.

5.3 Comparison of the FPT with other ToM measures

As the FPT is part of the five most relevant AToM operationalizations, identified by Osterhaus and Bosacki (2022), this section aims to compare the FPT with the four tests of the other main operationalizations to evaluate its unique position within ToM research. As already mentioned in the introduction, the most common operationalization of AToM involves the recognition of nonliteral speech such as irony, jokes, or sarcasm, followed by drawing inferences from nonverbal cues (e.g., the eyes), higher-order FB understanding, interpreting social situations, and detecting faux-pas. Five tasks dominate ToM research and account for more than 60 % of ToM measures used: second-order FB tasks (Perner & Wimmer, 1985), Strange Stories (Happé, 1994), the RMET (Baron-Cohen et al., 2001), the FPT (Stone et al., 1998), and the Frith-Happé Triangles test (Abell et al., 2000). While the Triangles test focuses on ascribing mental states to non-human objects (a less frequent operationalization), the MASC (Dziobek et al., 2006) is included for its widespread use in clinical ToM assessment in the following comparison.

The Strange Stories task is a frequently used task within the most common operationalization, the recognition of nonliteral speech (Osterhaus & Bosacki, 2022). The task shares a lot of similarities with the FPT test, as both involve complex second-order reasoning and present social situations where something strange happens. In addition, both tasks fulfill the criteria proposed by Quesque and Rossetti (mentalizing and non-merging criterion; 2020). However, there are still differences between these two tests. While the Strange Stories task is about a character doing or saying something unusual, the FPT is about whether a character commits a faux-pas (Bianco

et al., 2021). Despite these similarities, the FPT and Strange Stories task differ in their primary focus and operationalization of ToM. The Strange Stories task emphasizes reasoning about nonliteral speech such as irony or sarcasm (Osterhaus & Bosacki, 2022). Moreover, following the results of the factor analysis of several ToM measures by Osterhaus et al. (2016), the Strange Stories task loads on a different factor (i.e., social reasoning) than the FPT (i.e., recognizing transgressions of social norms). In other words, the emphasis of the FPT is to assess the ability of participants to identify and interpret instances where one person unintentionally causes social harm, capturing the interplay between social reasoning and norm understanding. In contrast, the Strange Stories task provides a rich narrative context to evaluate the comprehension of social interactions. While both tasks involve reasoning about complex social scenarios, the Strange Stories task adopts a broader focus on understanding social situations, whereas the FPT specifically targets the recognition of social mistakes and violations of norms. This targeted approach enables the FPT to capture the interaction between mental state reasoning and social understanding in a unique way.

The second most common operationalization of AToM tests is drawing inferences from nonverbal cues, such as the eyes (Osterhaus & Bosacki, 2022). One test within this operationalization is the RMET. Like the FPT, the RMET is a widely recognized instrument for assessing ToM, but they target different aspects and serve different purposes in research and clinical settings. The RMET focuses primarily on the affective facet of ToM, requiring participants to infer emotional states from subtle facial expressions, particularly the eye region (Vellante et al., 2013). This task involves social-perceptual processes and emphasizes emotion recognition, which is an essential component of ToM but does not necessarily extend to understanding complex social interactions, as it is necessary in the FPT. Some studies have even criticized that the RMET is a measure of emotion recognition and not ToM (Kittel et al., 2022; Quesque & Rossetti, 2020). The moderate correlation observed between the FPT and the RMET in Study 2 reflect their differing emphases. This divergence underscores the FPT's unique contribution to ToM measurement, particularly in its ability to capture higher-order cognitive processes related to social norm violations. The FPT's focus on recognizing violations of social norms is consistent with the factor analysis conducted by Osterhaus et al. (2016) which found that the FPT loads on the factor of recognizing social norm violations, whereas the RMET loads on the factor of social reasoning. In

addition, the FPT includes more extensive considerations of intentions and contextual information and thus addresses cognitive aspects of ToM that the RMET may not capture. In contrast to the FPT, the REMT does not capture either of the two criteria to qualify as a ToM measure, as proposed by Quesque and Rossetti (2020).

Second-order FB tasks focus on the ability to understand that one person can hold a belief about another person's belief, operationalized as higher-order FB reasoning (Osterhaus & Bosacki, 2022). In other words, second-order FB tasks assess a person's ability to recognize that another person has a different mental state, as well as the ability to think about another person's mental state. Therefore, these tasks also fulfil the mentalizing and non-merging criteria (Quesque & Rossetti, 2020). These tasks assess a foundational aspect of ToM (although they belong to AToM tasks), but lack the complexity and social contextuality that characterize tasks like the FPT. Among other questions, the FPT contained questions on second-order FB. This capacity is foundational for understanding complex social interactions, such as those involving misunderstandings, irony, or faux-pas. By incorporating these questions, the FPT evaluates a critical aspect of ToM that is necessary for identifying social norm violations and reasoning about the mental states of multiple individuals simultaneously. Second-order FB tasks load on the social reasoning factor (Osterhaus et al., 2016), which reflects the ability to interpret and reason about mental states and intentions in others. While second-order FB tasks in isolation, such as those by Perner and Wimmer (1985) are effective for assessing fundamental ToM reasoning, the FPT expands on this by embedding such reasoning within richer social narratives. This makes the FPT particularly suited for capturing real-world social-cognitive challenges, especially in clinical populations where impairments in these processes are more pronounced. Furthermore, the FPT's combination of second-order reasoning with faux-pas recognition allows for a multidimensional assessment of ToM that complements other measures, such as Strange Stories or the MASC.

The MASC is to be found within the operationalization of interpretations of social situations (Osterhaus & Bosacki, 2022). In contrast to the RMET, Strange Stories, second-order FB tasks and the FPT, the MASC is not assessed in a paper-and-pencil format. Instead, the MASC is assessed via videos, which necessitates the presence of speakers and monitors (Dziobek et al., 2006). Despite the fact that both the MASC and the FPT meet the mentalizing and non-merging criteria (Quesque & Rossetti, 2020), each task provides unique insights into distinct aspects of ToM. The correlations

between the FPT and the MASC, as found in Studies 1 and 2, also suggest that the two measures capture overlapping but distinct constructs. The MASC places emphasis on interpreting multimodal social cues (e.g., facial expressions, gestures, and prosody), whilst the FPT prioritizes reasoning about social mistakes. The format of the MASC captures ToM as it unfolds in real-time, requiring participants to integrate information across various modalities. Conversely, the FPT focuses on the detection and reasoning about social norm violations within contextualized, narrative-based scenarios (Şandor & İşcen, 2021).

5.3.1 Clinical Relevance and Practical Applications of the FPT

The findings of this dissertation underscore the clinical relevance of the FPT as a tool for assessing nuanced ToM abilities across various populations. Both the dichotomous and the polytomous response formats of the FPT demonstrate significant potential for practical applications in clinical and research contexts, particularly in understanding and addressing ToM deficits in mental disorders such as BPD and MDD.

5.4 Implications for Understanding Theory of Mind in Clinical Samples

Impairments in ToM have been consistently investigated across various mental disorders, reflecting their profound impact on interpersonal functioning. For instance, the systematic review of meta-analyses by Cotter et al. (2018) identified ToM impairments in 30 different clinical populations, with the most pronounced deficits observed in neurodegenerative disorders (including ASD) and psychotic disorders (such as schizophrenia).

Interestingly, the effect sizes in ToM deficits in individuals with BPD were almost comparable to HC (Cotter et al., 2018). The meta-analysis by Hanegraaf et al. (2021) found medium effect sizes in ToM impairments for BPD studies. A meta-analysis conducted by Németh et al. (2018) reported a small effect size in ToM impairments when comparing BPD and HC. Their comparison of performance in specific task types revealed that BPD patients performed significantly worse in verbal than in visual tasks. Likewise, Bora et al. (2021) also found small effects for ToM deficits in BPD compared to HC. Notably, separate investigations revealed that while individuals with BPD showed significantly poorer ToM performance in reasoning tasks, this difference did not extend to ToM-decoding tasks, emphasizing the variability of ToM impairments depending on the type of task.

In MDD, two meta-analyses that included studies that employed various ToM tasks, such as the FPT, the RMET, and the MASC, to measure ToM reported moderate

ToM impairments (Bora & Berk, 2016; Nestor et al., 2022). This contradicts the findings of large effects of another meta-analysis by Richman and Unoka (2015), which only included one measure (i.e., the RMET). Hence, the choice of the ToM measure has a significant influence on the results. Bora and Berk (2016) and Nestor et al. (2022) included several ToM measures, including verbal and visual tasks, in their analyses. They found ToM deficits in cognitive and affective facets, evident in both verbal and visual tasks. This variability supports the notion that ToM is not a monolithic, but a multifaceted construct.

With regard to specific ToM impairments, the meta-analysis by McLaren (2022) found that individuals with ASD, schizophrenia, and MDD demonstrated small effect sizes in exceeding ToM, while individuals with BPD exhibited small to large effect sizes in this ToM level. These findings suggest that while exceeding ToM is present across clinical populations, its degree varies considerably between disorders. However, direct comparisons of effect sizes across clinical conditions remain methodologically challenging due to differences in study designs. Despite these limitations, ToM deficits appear to be a consistent feature across various mental disorders, suggesting that difficulties in these social-cognitive domains transcend diagnostic categories. The degree and nature of these deficits show substantial variability, reflecting the complexity of ToM impairments in clinical populations.

5.4.1 Theory of Mind Deficits Within the RDoC Framework

The consistent observation of ToM impairments across different mental disorders and the potential of ToM as a transdiagnostic feature are of such significance that ToM has been included in integrative frameworks. The RDoC (National Institutes of Mental Health, 2020) and HiTOP (Conway et al., 2019) frameworks were developed with the objective of overcoming the limitations of traditional categorical diagnostic systems. This dimensional approach is consistent with the view that ToM impairments are not restricted to specific diagnostic categories. Rather, they are better conceptualized as a transdiagnostic phenomenon, reflecting shared mechanisms across a range of mental disorders. For example, Study 2 revealed less variability in the different ToM levels within the BPD sample, which can be attributed to their tendency to make exceeding ToM errors. These findings further support the conceptualization of ToM impairments as a dimensional construct that cuts across traditional diagnostic boundaries, offering a more nuanced understanding of social-cognitive deficits (Insel et al., 2010). Study 3 extended these findings by demonstrating

disorder-specific ToM profiles: individuals with BPD exhibited exceeding ToM, while those with MDD displayed deficits in correct ToM. These results illustrate how ToM impairments manifest differently across diagnostic categories but reflect a common underlying mechanism (ToM). The findings thus challenge the notion that exceeding ToM is associated with general psychopathology (McLaren et al., 2022). Since McLaren et al. (2022) only included studies that investigated ToM deficits using the MASC in their meta-analysis, the findings of Study 3 emphasize the task-dependent and disorder-specific nature of ToM impairments. As the various disorders were identified using established cut-off values for symptom manifestations, the results of the disorder-specific profiles show how different patterns of ToM deficits are related to specific clinical symptoms, which in turn supports the broader dimensional framework of RDoC.

The findings of Studies 2 and 3 thus underscore the importance of task-dependent and dimensional assessments in capturing the complexity of ToM deficits. Specifically, the findings of Study 3 not only advance the understanding of ToM impairments in BPD and MDD but also highlight the need for continued research within the RDoC framework. By employing dimensional assessments such as the FPT, future studies can better delineate the shared and unique mechanisms underlying ToM deficits.

5.4.2 *Theory of Mind Profiles in Borderline Personality Disorder*

Studies 2 and 3 provided additional insights into the association between exceeding ToM and BPD, as previously investigated in the research conducted by Bora et al., (2021), Sharp and Vanwoerden (2015) and McLaren et al. (2022). Exceeding ToM reflects an over-interpretation of social cues, where individuals attribute intentions or mental states that are not present. Although McLaren et al. (2022) suggested that exceeding ToM may be associated with general psychopathology rather than a specific feature of BPD, Study 3 is consistent with the other studies that have identified exceeding ToM as a hallmark of BPD (Bora, 2021; Sharp & Vanwoerden, 2015). This was also exemplified by the PCM analysis in Study 2 which revealed less variability in the different ToM levels within the BPD sample, which can be attributed to their tendency to make exceeding ToM errors. Notably, ToM errors in BPD patients were not observed in correct ToM or no ToM responses, emphasizing the specificity of exceeding ToM as a distinct impairment. As individuals with BPD often suffer from comorbid symptoms of depression (Richman & Unoka, 2015), the study also

investigated the impact of depression severity. Results demonstrated that BPD was associated with exceeding ToM regardless of depression severity, reinforcing its unique relevance to BPD. This notion is further supported by a meta-analysis by Németh et al. (2018). Even though they did not explicitly assess exceeding ToM, their findings support the notion that overall ToM differences between BPD and HC are not significantly influenced by the current depression rate within individuals with BPD. Consequently, the findings of the present study align with those of the aforementioned study, which posits that ToM characteristics associated with BPD may reflect unique traits of the disorder. The prevalence of exceeding ToM errors in BPD patients may stem from heightened sensitivity to social cues and biases, such as negativity or anger (Domes et al., 2009; Frick et al., 2012), or difficulties in emotion regulation (Sharp et al., 2011). These factors can lead individuals with BPD to perceive hostility or intent where none exists, resulting in interpersonal challenges and frequent conflicts (Sharp & Vanwoerden, 2015). -

However, ToM deficits have not been consistently observed in BPD. Instead, research suggests that it depends on the specific ToM task used, highlighting the task-dependent nature of ToM impairments in BPD. This is also evidenced by studies that found that individuals with BPD demonstrate a basic understanding of ToM when it is assessed by the RMET (Baez et al., 2015; Németh et al., 2018; Pourmohammad et al., 2021; Zabihzadeh et al., 2017). However, when ToM is assessed with the FPT, BPD patients demonstrate impaired AToM abilities, particularly in the identification of emotional cues in complex social situations (e.g., Baez et al., 2015; Harari et al., 2010; Petersen et al., 2016; Zabihzadeh et al., 2017). Specifically, one study observed that while BPD patients performed at a comparable level to HC on basic FB tasks, they made significantly more ToM errors in understanding faux-pas than HC (Petersen et al., 2016). These discrepancies may account for the observation that the ToM deficits of individuals with BPD exhibited effect sizes that were nearly equivalent to those of HC in the systematic review of meta-analyses conducted by Cotter et al. (2018). In their review, they grouped ToM tasks together, noting that these include a relatively heterogeneous array of measures.

Some studies argue that ToM deficits in BPD become more pronounced in complex tasks requiring contextual processing and the integration of multiple mental state perspectives (Baez et al., 2015; Petersen et al., 2016). Scenarios that mimic real-life contexts, such as in the FPT, demand a deeper contextual understanding, which is

particularly sensitive for detecting subtle ToM impairments in individuals with BPD. However, the Strange Stories task by Happé (1994) also includes the integration of multiple mental state perspectives and is therefore also considered a complex AToM task. Still, Arntz et al. (2009) found no evidence for ToM impairments in BPD patients, regardless of whether intelligence, social reasoning, or mood were controlled for. A key distinction, however, lies in the type of ToM error being assessed. Arntz et al. (2009) did not specifically examine exceeding ToM, which Study 3 identified as uniquely associated with BPD. Moreover, Study 3 found no significant differences in correct ToM abilities between individuals with BPD and the community sample, suggesting that ToM impairments in BPD are not generalized but rather specific to exceeding ToM errors. This specificity may explain discrepancies across studies and highlights the importance of task selection in ToM research. Tasks like the FPT, that are sensitive to nuanced social-cognitive impairments, are particularly valuable for uncovering the specific ToM deficits that characterize BPD.

5.4.3 Theory of Mind Profiles in Major Depressive Disorder

Study 3 revealed that individuals with MDD displayed deficits in correct ToM, aligning with prior research reporting that overall ToM abilities are impaired in MDD (e.g., Bora & Berk, 2016; Richman & Unoka, 2015). Despite the growing body of evidence for ToM deficits, current research offers conflicting results regarding the specific direction of ToM errors (i.e., whether mental states are more likely to be over- or under-interpreted in MDD). Some studies have indicated that MDD is associated with less ToM (De Coninck et al., 2021; Safiye et al., 2023; Scandurra et al., 2020; Wolkenstein et al., 2011). However, studies included in the meta-analysis by McLaren et al. (2022) have shown that MDD is associated with exceeding ToM. Addressing this gap, Study 3 presented findings on the association between MDD and specific ToM levels.

The observed inconsistencies emphasize the necessity for validated ToM measures, given that previous studies have employed the RFQ and the MASC to assess ToM deficits. The validity of the RFQ as a ToM measure has been questioned (Müller et al., 2022), and although the MASC is a reliable and valid instrument (Fossati et al., 2018), its specific ToM error categories lack thorough validation. Notably, the FPT assesses 'no ToM' errors, which do not correspond with the 'no' or 'less ToM' subscales of the MASC. Thus the FPT might capture different aspects of under-attributing mental states to others than the MASC. The FPT focuses on recognizing

social blunders, assessing an individual's capacity to detect and interpret unintentional social missteps. In contrast, the MASC evaluates the understanding of complex social interactions through film. This method assesses the participant's capacity to understand nuanced social cues presented through both verbal and non-verbal channels, such as facial expressions and tone of voice. Therefore, the MASC might identify forms of under-ascribing mental states that the FPT does not specifically measure, highlighting the importance of selecting appropriate tools to capture the multifaceted nature of ToM impairments. In MDD, deficits in correct ToM may arise from factors such as reduced social engagement, cognitive slowing, and a tendency toward self-focused attention (Nestor et al., 2022; Porcelli et al., 2019). The absence of exceeding ToM in MDD could reflect the disorder's association with cognitive and emotional withdrawal, limiting opportunities for over-interpretation of social cues.

While no such association was found between MDD and exceeding ToM, Study 3 challenges the notion that exceeding ToM is a general feature of psychopathology, as indicated by McLaren et al. (2022). In summary, the findings emphasize the distinct ToM profiles associated with various mental disorders and underscore the importance of considering ToM as a dimensional construct that manifests differently across disorders.

5.5 Limitations and Implications for Future Research

5.5.1 Limitations

While this dissertation advances the field of ToM research and provides valuable insights into the assessment and understanding of ToM impairments, there are several limitations that require further investigations that must be addressed. One overall limitation affecting the research presented across studies is that the studies only validated three items per vignette. These items were selected because they assess ToM responses that extend beyond binary yes/no categories. While these items demonstrated robust psychometric properties, the studies did not assess the remaining items in the FPT vignettes. As a result, the findings cannot be generalized to the full questionnaire. To provide a more comprehensive assessment of ToM, future studies should validate additional vignette items, ensuring that all components of the FPT contribute to measuring ToM effectively. Moreover, even though the FPT showed a good overall fit, some items had to be excluded due to suboptimal fit, highlighting the need for further refinement of the measure.

The formation of the clinical and community samples in this study represents a general shortcoming of this dissertation, as neither was clearly diagnostically defined. The clinical groups were not based on structured diagnostic interviews to confirm clinical diagnoses. However, evidence suggests that the screening instruments used provide reasonable diagnostic accuracy. For instance, a meta-analysis by Moriarty et al. (2015), demonstrated that the PHQ-9 has acceptable diagnostic properties for detecting MDD. Similarly, according to a study by Kleindienst et al. (2020) the BSL-23 can reliably detect moderate symptom severity in individuals with BPD. Similarly, the community samples excluded participants currently undergoing psychiatric, psychological, or psychotherapeutic treatment, and one sample was additionally screened negative for PTSD. However, depression severity and other clinical markers were not assessed, leaving the potential for undetected symptoms. Consequently, this lacks a true healthy control group, which limits the ability to compare clinical groups to a fully non-clinical baseline in Study 3. Moreover, for Studies 1 and 2, this may have introduced variability into the community sample.

Another limitation of the studies comprising this dissertation is the homogeneous community sample. The community samples mainly consisted of individuals from WEIRD backgrounds (Henrich et al., 2010). While this is a common limitation in psychological research, it restricts the generalizability of the findings to more diverse populations. Moreover, research suggests that this test is particularly susceptible to ceiling effects in homogenous, highly educated samples (Yeung et al., 2024). Ceiling effects indicate that a task cannot generate enough variance to study individual differences effectively.

In this dissertation, the polytomous version of the FPT overcame this issue by allowing for more a nuanced differentiation of ToM levels, particularly among high-achieving participants. This increased complexity raises the difficulty of the test, requiring participants to demonstrate deeper comprehension and recognize finer distinctions. This limitation is closely tied to a further limitation regarding associations with certain ToM measures in this study, such as the RMET. As the test primarily measures basic emotion recognition, it is relatively easy to administer, lacking sufficient sensitivity to differentiate between individuals with higher levels of ToM abilities (Black, 2019) and therefore often leads to ceiling effects, with a significant proportion of participants achieving near-perfect scores. This compresses the range of variability and reduces the sensitivity of the test to detect individual differences in ToM abilities.

However, the inclusion of the MASC in this study is a significant strength. The MASC does not exhibit ceiling effects (Yeung et al., 2024) and can thus differentiate between individuals with low and high ToM abilities. Moreover, it provides a more sensitive and nuanced assessment of ToM abilities, particularly in identifying specific ToM levels. By using the MASC alongside other ToM measures, the studies are able to generate more robust findings. Nevertheless, expanding the diversity of participant samples in future research would further enhance the generalizability of these results and allow for a more comprehensive understanding of how ToM abilities vary across different populations and contexts.

A further shortcoming is the relatively small sample sizes of the clinical samples in Studies 2 and 3. While the findings offer valuable insights, the limited sample size may restrict the generalizability of the results to broader populations, particularly given the variability in ToM abilities across individuals within the same diagnostic category. Larger sample sizes would increase statistical power, allow for more robust subgroup analyses, and provide greater confidence in the stability of the observed patterns of ToM impairments. Furthermore, Study 3 did not include a sufficiently diverse range of clinical conditions. ToM impairments are a well-documented feature of several other psychopathologies, such as autism spectrum disorder (Baron-Cohen, 2000) or schizophrenia (Sprong et al., 2007), yet these groups were not represented. The absence of these conditions limits the conclusions that can be drawn about the specificity and generalizability of the findings. From an RDoC (2020) perspective, ToM is a core construct within the Social Processes domain, emphasizing the importance of understanding its shared and unique patterns across different psychopathologies. By limiting the clinical groups to BPD and MDD, the current study captures a narrow spectrum of ToM impairments, which may overlook transdiagnostic mechanisms or ToM deficits unique to other disorders. Including a broader range of clinical groups in future research would facilitate comparisons across disorders, providing deeper insights into both shared ToM impairments and disorder-specific profiles. This transdiagnostic approach would align with the RDoC framework and contribute to a more comprehensive understanding of the dimensional nature of ToM deficits across psychopathologies.

Another potential limitation of our study is the inability to account fully for the influence of general cognitive abilities on the observed ToM performance. ToM tasks rely on various underlying higher-order abilities, such as causal inference, as well as

lower-order abilities like gaze tracking (Schaafsma et al., 2015). Due to time constraints, it was not possible to assess cognitive abilities in a clinical context. However, all three studies excluded items based on performance on the FPT control questions to reduce the influence of general cognitive ability. Despite these precautions, the potential influence of cognitive factors still cannot be entirely ruled out. Prior research indicates that ToM performance is related to several cognitive abilities, such as executive functions (EF; Baez et al., 2015; Németh et al., 2020) and intelligence (Baker et al., 2014; Dodell-Feder et al., 2013). EF, which supports cognitive processes such as attentional control, cognitive flexibility, and inhibition, likely contributed to the observed group differences, as EF predict performance specifically in cognitively more demanding tasks, like the FPT (Baez et al., 2015; Gregory et al., 2002; Stone et al., 1998b).

Furthermore, intelligence, particularly verbal comprehension, may have influenced FPT outcomes, given the strong relationship between ToM and language-based tasks (Milligan et al., 2007; Navarro et al., 2021). Importantly, the differences in ToM abilities between groups remain evident, regardless of the potential influence of EF. This highlights that ToM impairments in clinical populations are meaningful and distinct, even if EF and/or intelligence play a contributory or mediating role. In BPD, for example, the heightened sensitivity to social cues (Frick et al., 2012) may interact with emotional dysregulation to produce exceeding ToM. In MDD, cognitive slowing and self-focused attention (Jin et al., 2015) may impair correct ToM. Understanding these interactions could provide a more holistic view of how ToM deficits arise and how they might be addressed.

5.5.2 Implications for Future Research

Building on the promising results of this dissertation, future research should aim to extend the further development and validation of the FPT. During the analysis of the data, certain items had to be removed due to misfit with the model. Misfitting items can indicate that the content or structure of these items did not align well with the latent construct being measured or that they failed to perform consistently across participants. The removal of such items is known to enhance the overall validity and reliability of the measure, but it may also have implications for the breadth and representativeness of the construct. Future studies should therefore carefully revise or replace these items to address their shortcomings while preserving their theoretical

relevance. This highlights the need for continuous refinement of assessment tools to ensure their robustness and applicability across diverse samples.

Moreover, future research should aim to improve its applicability and accuracy across a wider range of clinical populations and contexts. To achieve this, future studies could apply adaptive testing methods. Adaptive versions of the FPT could dynamically adjust item difficulty based on individual performance, ensuring more accurate assessment across varying levels of ToM ability. This approach would ensure that the FPT remains accessible while maintaining its diagnostic value (Dodd et al., 1995). This is particularly relevant in populations with lower cognitive abilities, such as individuals with cognitive impairment due to high symptom severity (Pan et al., 2019). In these populations, the introduction of adaptive testing could enhance the utility of the FPT. By incorporating easier items, adaptive testing could improve the sensitivity of the FPT, ensuring that it captures a broader range of ToM abilities. This adjustment is particularly relevant because cognitive abilities such as EF and intelligence have been shown to predict ToM reasoning, as assessed by the FPT (Németh et al., 2020). Including a wider range of item difficulties in future versions of the FPT would better account for individual variability in cognitive abilities. Moreover, it would ensure more accurate assessments across various populations that have reliably been associated with ToM deficits, e.g., people with ASC, schizophrenia, or PTSD. By addressing these differences, the FPT could maintain its diagnostic value even in populations with significant cognitive limitations. Additionally, future research should investigate the potential role of EF and intelligence in explaining ToM differences to further refine the FPT's application across diverse populations. Further, this approach would enable personalized treatment planning. By identifying specific ToM impairments, the FPT could guide interventions that target the unique social-cognitive challenges faced by individuals with different disorders. For example, as Study 3 has shown that individuals with BPD exhibit deficits in exceeding ToM, interventions could specifically target their exacerbated interpersonal conflicts and misunderstandings. Additionally, future studies could use the FPT to track changes in ToM abilities over time in order to evaluate treatment efficacy.

The emphasis on refining and expanding testing contexts aligns with the conclusions of Arntz et al. (2009), who found no evidence of ToM deficits in BPD when assessed in neutral, nonpersonal settings. Arntz and colleagues (2009) highlighted that while BPD may not be associated with gross ToM impairments, ToM and other social-

cognitive functions could be disturbed under specific circumstances, such as emotional or personal situations. This underscores the importance of studying ToM in emotionally charged or contextually relevant conditions, as these may reveal deficits that remain undetected in neutral contexts. Subsequent studies should therefore prioritize testing ToM and related capacities in socially or emotionally complex scenarios to capture the full spectrum of impairments.

Further research is also needed to explore the cognitive and affective facets of the FPT. As studies diverge on whether the FPT assesses solely cognitive ToM (Maleki et al., 2020; Zabihzadeh et al., 2017) or both cognitive and affective ToM (Németh et al., 2020; Pluta et al., 2018; Pourmohammad et al., 2021), future validation studies should aim to clarify the ability of the FPT to differentiate between cognitive and affective ToM components. Such differentiation would enhance the theoretical and clinical utility of the FPT by providing greater sensitivity to domain-specific impairments.

By pursuing further development of the FPT in these directions and emphasizing contextualized testing, researchers may gain a deeper understanding of ToM deficits in various clinical populations. Detecting disorder-specific ToM profiles would help to identify the specific social-cognitive challenges faced by people with mental disorders and provide targeted interventions tailored to individual needs.

6 Conclusion

This dissertation advances the field of ToM research by validating new dichotomous and polytomous response formats of the FPT and applying them to clinical and community samples. The findings highlight the FPT's utility in measuring ToM as a multidimensional construct, providing valuable insights into specific ToM impairments and their relevance to psychopathology. The validation of the dichotomous format confirmed the validity of the FPT and its ability to measure a single underlying construct, while the polytomous format captured distinct ToM levels (i.e., no ToM, correct ToM, and exceeding ToM), offering a more nuanced assessment of ToM deficits. Study 2 therefore enhances the precision of ToM measurement, addressing longstanding concerns about the limitations of traditional ToM instruments.

The importance of employing valid ToM measures cannot be overstated. Many widely used tools, such as the RMET, have been criticized for their limited scope or lack of adherence to the core criteria of ToM assessments (i.e., the ability to represent others' mental states and to distinguish these from one's own; Quesque & Rossetti,

2020). While other tools, like the Strange Stories and the MASC, align with these criteria, they face practical challenges, including open-ended formats or resource-intensive administration requirements. The newly validated closed-answer formats of the FPT overcome these issues, offering a reliable, easy-to-administer, and objective alternative. By capturing ToM across multiple levels, the FPT establishes itself as a robust instrument that complements existing measures and provides more nuanced insights into ToM impairments.

The studies presented in this dissertation underscore the important role of valid ToM measures in advancing both theoretical and clinical understanding. The FPT's ability to differentiate between disorder-specific ToM profiles has significant implications for understanding how ToM impairments manifest in conditions like BPD and MDD. For instance, the finding that individuals with BPD predominantly exhibit exceeding ToM errors while those with MDD show deficits in correct ToM demonstrates the necessity of distinguishing between different types of ToM impairments. These findings challenge the notion that exceeding ToM is a transdiagnostic marker of psychopathology and suggest that it is uniquely associated with BPD, emphasizing the need for tools that can reliably identify such patterns.

The validated FPT formats also align with the goals of the RDoC framework by operationalizing ToM as a multidimensional construct relevant across traditional diagnostic boundaries. The ability to reliably measure distinct ToM levels supports the development of targeted interventions tailored to specific impairments, enhancing the potential for personalized treatment approaches. Moreover, using valid ToM measures like the FPT ensures that future research can build on robust, reliable data to explore how ToM deficits evolve over time, interact with other cognitive and emotional processes, and respond to therapeutic interventions.

According to Quesque and Rossetti (2020), the fundamental components of ToM include the ability to represent others' mental states and to distinguish these from one's own. The FPT used in this dissertation aligns with these principles, while additionally offering a nuanced assessment of ToM levels by examining participants' ability to identify and interpret social blunders. While many ToM measures are criticized for their limited ecological validity or reliance, the FPT stands out as a versatile tool that captures complex social interactions in an accessible format. Furthermore, the new dichotomous and polytomous answer formats introduced and validated in this research enhance the test's objectivity and usability. However, to solidify the FPT's

role as a reliable and ecologically valid ToM measure, future studies should aim to replicate these findings across diverse populations and explore its applicability in emotionally charged scenarios. Comparative analyses with other validated ToM tools would further strengthen its validity and contribute to a deeper understanding of ToM impairments across clinical and non-clinical groups.

In conclusion, this dissertation represents a significant contribution to ToM research, addressing critical gaps in the field by providing validated and innovative measurement tools. The findings advance the theoretical understanding of ToM as a complex, multidimensional construct and offer practical solutions for its assessment in both clinical and community settings.

References

Abell, F., Happé, F., & Frith, U. (2000). Do triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development. *Cognitive Development*, *15*(1), 1–16. https://doi.org/10.1016/S0885-2014(00)00014-9

Abu-Akel, A., & Shamay-Tsoory, S. (2013). *Characteristics of Theory of Mind Impairments in Schizophrenia*.

Adegboye, D., Sterr, A., Lin, J.-P., & Owen, T. J. (2017). Theory of mind, emotional and social functioning, and motor severity in children and adolescents with dystonic cerebral palsy. *European Journal of Paediatric Neurology*, *21*(3), 549–556. https://doi.org/10.1016/j.ejpn.2017.01.013

Adolphs, R. (1999). Social cognition and the human brain. *Trends in Cognitive Sciences*, *3*(12), 469–479. https://doi.org/10.1016/S1364-6613(99)01399-6

American Psychiatric Association (Ed.). (1994). *Diagnostic criteria from DSM-IV* (Reprint.). American Psychiatric Association.

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders (5th ed.)*. American Psychiatric Association.

Apperly, I. A. (2012). What is "theory of mind"? Concepts, cognitive processes and individual differences. *Quarterly Journal of Experimental Psychology*, *65*(5), 825–839. https://doi.org/10.1080/17470218.2012.676055

Arioli, M., Crespi, C., & Canessa, N. (2018). Social Cognition through the Lens of Cognitive and Clinical Neuroscience. *BioMed Research International*, *2018*, 1–18. https://doi.org/10.1155/2018/4283427

Arntz, A., Bernstein, D., Oorschot, M., & Schobre, P. (2009). Theory of Mind in Borderline and Cluster-C Personality Disorder. *Journal of Nervous & Mental Disease*, *197*(11), 801–807. https://doi.org/10.1097/NMD.0b013e3181be78fb

Baez, S., Marengo, J., Perez, A., Huepe, D., Font, F. G., Rial, V., Gonzalez-Gadea, M. L., Manes, F., & Ibanez, A. (2015). Theory of mind and its relationship with executive functions and emotion recognition in borderline personality disorder. *Journal of Neuropsychology*, *9*(2), 203–218. https://doi.org/10.1111/jnp.12046

Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences*, *14*(3), 110–118. https://doi.org/10.1016/j.tics.2009.12.006

Baker, C. A., Peterson, E., Pulos, S., & Kirkland, R. A. (2014). Eyes and IQ: A meta-analysis of the relationship between intelligence and "Reading the Mind in the Eyes." *Intelligence*, *44*, 78–92. https://doi.org/10.1016/j.intell.2014.03.001

Baker, F. B., & Kim, S.-H. (Eds.). (2004). *Item Response Theory* (0 ed.). CRC Press. https://doi.org/10.1201/9781482276725

Balogh, N., Égerházi, A., Berecz, R., & Csukly, G. (2014). Investigating the state-like and trait-like characters of social cognition in schizophrenia: A short term follow-up study. *Schizophrenia Research*, *159*(2–3), 499–505. https://doi.org/10.1016/j.schres.2014.08.027

Banerjee, R., Watling, D., & Caputi, M. (2011). Peer Relations and the Understanding of Faux-pas: Longitudinal Evidence for Bidirectional Associations: Peer Relations Faux-pas. *Child Development*, *82*(6), 1887–1905. https://doi.org/10.1111/j.1467-8624.2011.01669.x

Baron-Cohen, S. (2000). Theory of mind and autism: A review. In *International Review of Research in Mental Retardation* (Vol. 23, pp. 169–184). Academic Press. https://doi.org/10.1016/S0074-7750(00)80010-5

Baron-Cohen, S. (2002). The extreme male brain theory of autism. *Trends in Cognitive Sciences*, *6*(6), 248–254. https://doi.org/10.1016/S1364-6613(02)01904-6

Baron-Cohen, S. (2009). Autism: The Empathizing–Systemizing (E-S) Theory. *Annals of the New York Academy of Sciences*, *1156*(1), 68–80. https://doi.org/10.1111/j.1749-6632.2009.04467.x

Baron-Cohen, S., Bowen, D. C., Holt, R. J., Allison, C., Auyeung, B., Lombardo, M. V., Smith, P., & Lai, M.-C. (2015). The "Reading the Mind in the Eyes" Test: Complete Absence of Typical Sex Difference in ~400 Men and Women with Autism. *PLOS ONE*, *10*(8), e0136521. https://doi.org/10.1371/journal.pone.0136521

Baron-Cohen, S., Campbell, R., Karmiloff-Smith, A., Grant, J., & Walker, J. (1995). Are children with autism blind to the mentalistic significance of the eyes? *British Journal of Developmental Psychology*, *13*(4), 379–398. https://doi.org/10.1111/j.2044-835X.1995.tb00687.x

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, *21*(1), 37–46. https://doi.org/10.1016/0010-0277(85)90022-8

Baron-Cohen, S., O'Riordan, M., Stone, V., Jones, R., & Plaisted, K. (1999). Recognition of Faux-pas by Normally Developing Children and Children with Asperger Syndrome or High-Functioning Autism. *Journal of Autism and*

Developmental Disorders, *29*(5), 407–418. https://doi.org/10.1023/a:1023035012436

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, and Allied Disciplines*, *42*(2), 241–251. https://doi.org/10.1111/1469-7610.00715

Baron-Cohen, S., Wheelwright, S., & Jolliffe, T. (1997). Is There a "Language of the Eyes"? Evidence from Normal Adults, and Adults with Autism or Asperger Syndrome. *Visual Cognition*, *4*(3), 311–331. https://doi.org/10.1080/713756761

Baxter, A. J., Patton, G., Scott, K. M., Degenhardt, L., & Whiteford, H. A. (2013). Global Epidemiology of Mental Disorders: What Are We Missing? *PLoS ONE*, *8*(6), e65514. https://doi.org/10.1371/journal.pone.0065514

Berecz, H., Tényi, T., & Herold, R. (2016). Theory of Mind in Depressive Disorders: A Review of the Literature. *Psychopathology*, *49*(3), 125–134. https://doi.org/10.1159/000446707

Bernstein, D. M., Thornton, W. L., & Sommerville, J. A. (2011). Theory of Mind Through the Ages: Older and Middle-Aged Adults Exhibit More Errors Than Do Younger Adults on a Continuous False Belief Task. *Experimental Aging Research*, *37*(5), 481–502. https://doi.org/10.1080/0361073X.2011.619466

Berthelot, N., Lemieux, R., Garon-Bissonnette, J., Lacharité, C., & Muzik, M. (2019). The protective role of mentalizing: Reflective functioning as a mediator between child maltreatment, psychopathology and parental attitude in expecting parents.

Child Abuse & Neglect, *95*, 104065. https://doi.org/10.1016/j.chiabu.2019.104065

Bianco, F., Lombardi, E., Lecce, S., Marchetti, A., Massaro, D., Valle, A., & Castelli, I. (2021). Supporting Children's Second-order Recursive Thinking and Advanced ToM Abilities: A Training Study. *Journal of Cognition and Development*, *22*(4), 561–584. https://doi.org/10.1080/15248372.2021.1901712

Black, J. E. (2019). An IRT Analysis of the Reading the Mind in the Eyes Test. *Journal of Personality Assessment*, *101*(4), 425–433. https://doi.org/10.1080/00223891.2018.1447946

Bohus, M., Kleindienst, N., Limberger, M. F., Stieglitz, R.-D., Domsalla, M., Chapman, A. L., Steil, R., Philipsen, A., & Wolf, M. (2009). The Short Version of the Borderline Symptom List (BSL-23): Development and Initial Data on Psychometric Properties. *Psychopathology*, *42*(1), 32–39. https://doi.org/10.1159/000173701

Bohus, M., Limberger, M. F., Frank, U., Sender, I., Gratwohl, T., & Stieglitz, R.-D. (2001). Entwicklung der Borderline- Symptom-Liste1. *PPmP - Psychotherapie · Psychosomatik · Medizinische Psychologie*, *51*(5), 201–211. https://doi.org/10.1055/s-2001-13281

Bora, E. (2021). A meta-analysis of theory of mind and 'mentalization' in borderline personality disorder: A true neuro-social-cognitive or meta-social-cognitive impairment? *Psychological Medicine*, *51*(15), 2541–2551. https://doi.org/10.1017/S0033291721003718

Bora, E., & Berk, M. (2016). Theory of mind in major depressive disorder: A meta-analysis. *Journal of Affective Disorders*, *191*, 49–55. https://doi.org/10.1016/j.jad.2015.11.023

Bora, E., & Pantelis, C. (2013). Theory of mind impairments in first-episode psychosis, individuals at ultra-high risk for psychosis and in first-degree relatives of schizophrenia: Systematic review and meta-analysis. *Schizophrenia Research*, *144*(1), 31–36. https://doi.org/10.1016/j.schres.2012.12.013

Bora, E., Yucel, M., & Pantelis, C. (2009). Theory of mind impairment in schizophrenia: Meta-analysis. *Schizophrenia Research*.

Bosco, F. M., Gabbatore, I., & Tirassa, M. (2014). A broad assessment of theory of mind in adolescence: The complexity of mindreading. *Consciousness and Cognition*, *24*, 84–97. https://doi.org/10.1016/j.concog.2014.01.003

Bovin, M. J., Kimerling, R., Weathers, F. W., Prins, A., Marx, B. P., Post, E. P., & Schnurr, P. P. (2021). Diagnostic Accuracy and Acceptability of the Primary Care Posttraumatic Stress Disorder Screen for the *Diagnostic and Statistical Manual of Mental Disorders* (Fifth Edition) Among US Veterans. *JAMA Network Open*, *4*(2), e2036733. https://doi.org/10.1001/jamanetworkopen.2020.36733

Brewer, N., Young, R. L., & Barnett, E. (2017). Measuring Theory of Mind in Adults with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *47*(7), 1927–1941. https://doi.org/10.1007/s10803-017-3080-x

Brewin, C. R., Andrews, B., & Valentine, J. D. (2000). Meta-analysis of risk factors for posttraumatic stress disorder in trauma-exposed adults. *Journal of Consulting and Clinical Psychology*, *68*(5), 748–766. https://doi.org/10.1037/0022-006X.68.5.748

Brooks, R., & Meltzoff, A. N. (2015). Connecting the dots from infancy to childhood: A longitudinal study connecting gaze following, language, and explicit theory of mind. *Journal of Experimental Child Psychology*, *130*, 67–78. https://doi.org/10.1016/j.jecp.2014.09.010

Buhlmann, U., Wacker, R., & Dziobek, I. (2015). Inferring other people's states of mind: Comparison across social anxiety, body dysmorphic, and obsessive–compulsive disorders. *Journal of Anxiety Disorders*, *34*, 107–113. https://doi.org/10.1016/j.janxdis.2015.06.003

Burghardt, J., Gradl, S., Knopp, M., & Sprung, M. (2023). Psychopathology and Theory of Mind in patients with personality disorders. *Borderline Personality Disorder and Emotion Dysregulation*, *10*(1), 18. https://doi.org/10.1186/s40479-023-00224-1

Buttelmann, F., Suhrke, J., & Buttelmann, D. (2015). What you get is what you believe: Eighteen-month-olds demonstrate belief understanding in an unexpected-identity task. *Journal of Experimental Child Psychology*, *131*, 94–103. https://doi.org/10.1016/j.jecp.2014.11.009

Callenmark, B., Kjellin, L., Rönnqvist, L., & Bölte, S. (2014). Explicit versus implicit social cognition testing in autism spectrum disorder. *Autism*, *18*(6), 684–693. https://doi.org/10.1177/1362361313492393

Carpendale, J. I. M., & Lewis, C. (2004). Constructing an understanding of mind: The development of children's social understanding within social interaction. *Behavioral and Brain Sciences*, *27*(01). https://doi.org/10.1017/S0140525X04000032

Cavallini, E., Lecce, S., Bottiroli, S., Palladino, P., & Pagnin, A. (2013). Beyond False Belief: Theory of Mind in Young, Young-Old, and Old-Old Adults. *International Journal of Aging & Human Development*, *76*, 181–198. https://doi.org/10.2190/AG.76.3.a

Charernboon, T., & Lerthattasilp, T. (2017). The Reading the Mind in the Eyes Test: Validity and Reliability of the Thai Version. *Cognitive and Behavioral Neurology*, *30*(3), 98–101. https://doi.org/10.1097/WNN.0000000000000130

Cheng, P., Jasinski, N., Zheng, W., Yadava, A., Wang, L., Li, L., Xu, L., Zhou, Y., Zhang, L., & Li, W. (2021). Psychometric Properties of the Primary Care PTSD Screen for DSM-5: Findings From Family Members of Chinese Healthcare Workers During the Outbreak of COVID-19. *Frontiers in Psychiatry*, *12*, 695678. https://doi.org/10.3389/fpsyt.2021.695678

Chung, Y. S., Barch, D., & Strube, M. (2014). A Meta-Analysis of Mentalizing Impairments in Adults With Schizophrenia and Autism Spectrum Disorder. *Schizophrenia Bulletin*, *40*(3), 602–616. https://doi.org/10.1093/schbul/sbt048

Clements, W. A., & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development*, *9*(4), 377–395. https://doi.org/10.1016/0885-2014(94)90012-4

Conway, C. C., Forbes, M. K., Forbush, K. T., Fried, E. I., Hallquist, M. N., Kotov, R., Mullins-Sweatt, S. N., Shackman, A. J., Skodol, A. E., South, S. C., Sunderland, M., Waszczuk, M. A., Zald, D. H., Afzali, M. H., Bornovalova, M. A., Carragher, N., Docherty, A. R., Jonas, K. G., Krueger, R. F., … Eaton, N. R. (2019). A Hierarchical Taxonomy of Psychopathology Can Transform Mental Health Research. *Perspectives on Psychological Science: A Journal of the Association*

for Psychological Science, *14*(3), 419–436. https://doi.org/10.1177/1745691618810696

Cortés-García, L., McLaren, V., Vanwoerden, S., & Sharp, C. (2021). Attachment, mentalizing, and eating disorder symptoms in adolescent psychiatric inpatients and healthy controls: A test of a mediational model. *Eating and Weight Disorders - Studies on Anorexia, Bulimia and Obesity*, *26*(4), 1159–1168. https://doi.org/10.1007/s40519-020-01017-z

Cotter, J., Granger, K., Backx, R., Hobbs, M., Looi, C. Y., & Barnett, J. H. (2018). Social cognitive dysfunction as a clinical marker: A systematic review of meta-analyses across 30 clinical conditions. *Neuroscience & Biobehavioral Reviews*, *84*, 92–99. https://doi.org/10.1016/j.neubiorev.2017.11.014

Couette, M., Mouchabac, S., Bourla, A., Nuss, P., & Ferreri, F. (2020). Social cognition in post-traumatic stress disorder: A systematic review. *British Journal of Clinical Psychology*, *59*(2), 117–138. https://doi.org/10.1111/bjc.12238

Davis, H. L., & Pratt, C. (1995). The development of children's theory of mind: The working memory explanation. *Australian Journal of Psychology*, *47*(1), 25–31. https://doi.org/10.1080/00049539508258765

De Ayala, R. J. (1995). The Influence of Dimensionality on Estimation in the Partial Credit Model. *Educational and Psychological Measurement*, *55*(3), 407–422. https://doi.org/10.1177/0013164495055003004

De Coninck, D., Matthijs, K., & Luyten, P. (2021). Depression in Belgian first-year university students: A longitudinal study of self-definition, interpersonal relatedness, mentalizing, and integration. *Journal of Clinical Psychology*, *77*(7), 1715–1731. https://doi.org/10.1002/jclp.23149

Devine, R. T. (2021). Individual differences in theory of mind in middle childhood and adolescence. In R. T. Devine & S. Lecce (Eds.), *Theory of Mind in Middle Childhood and Adolescence* (1st ed., pp. 55–76). Routledge. https://doi.org/10.4324/9780429326899-5

Devine, R. T., & Hughes, C. (2013). Silent Films and Strange Stories: Theory of Mind, Gender, and Social Experiences in Middle Childhood. *Child Development*, *84*(3), 989–1003. https://doi.org/10.1111/cdev.12017

Devine, R. T., & Hughes, C. (2016). Measuring theory of mind across middle childhood: Reliability and validity of the Silent Films and Strange Stories tasks. *Journal of Experimental Child Psychology*, *149*, 23–40. https://doi.org/10.1016/j.jecp.2015.07.011

Doba, K., Saloppé, X., Choukri, F., & Nandrino, J.-L. (2022). Childhood trauma and posttraumatic stress symptoms in adolescents and young adults: The mediating role of mentalizing and emotion regulation strategies. *Child Abuse & Neglect*, *132*, 105815. https://doi.org/10.1016/j.chiabu.2022.105815

Dodd, B. G., De Ayala, R. J., & Koch, W. R. (1995). Computerized Adaptive Testing With Polytomous Items. *Applied Psychological Measurement*, *19*(1), 5–22. https://doi.org/10.1177/014662169501900103

Dodell-Feder, D., Lincoln, S. H., Coulson, J. P., & Hooker, C. I. (2013). Using Fiction to Assess Mental State Understanding: A New Task for Assessing Theory of Mind in Adults. *PLoS ONE*, *8*(11), e81279. https://doi.org/10.1371/journal.pone.0081279

Domes, G., Schulze, L., & Herpertz, S. C. (2009). Emotion Recognition in Borderline Personality Disorder—A Review of the Literature. *Journal of Personality Disorders*, *23*(1), 6–19. https://doi.org/10.1521/pedi.2009.23.1.6

Dumontheil, I., Apperly, I. A., & Blakemore, S. (2010). Online usage of theory of mind continues to develop in late adolescence. *Developmental Science*, *13*(2), 331–338. https://doi.org/10.1111/j.1467-7687.2009.00888.x

Dvash, J., & Shamay-Tsoory, S. G. (2014). Theory of Mind and Empathy as Multidimensional Constructs: Neurological Foundations. *Topics in Language Disorders*, *34*(4), 282–295. https://doi.org/10.1097/TLD.0000000000000040

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. https://doi.org/10.1007/s10803-006-0107-0

Eilers, R., & Rosner, R. (2021). Die einfache und komplexe Posttraumatische Belastungsstörung in der Praxis: Eine Übersicht und Einordnung der neuen ICD-11 Kriterien in Bezug auf Kinder und Jugendliche. *Kindheit und Entwicklung*, *30*(3), 144–153. https://doi.org/10.1026/0942-5403/a000342

Erskine, H. E., Moffitt, T. E., Copeland, W. E., Costello, E. J., Ferrari, A. J., Patton, G., Degenhardt, L., Vos, T., Whiteford, H. A., & Scott, J. G. (2015). A heavy burden on young minds: The global burden of mental and substance use disorders in children and youth. *Psychological Medicine*, *45*(7), 1551–1563. https://doi.org/10.1017/S0033291714002888

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical

sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Faux-pas Test (Adult). (2023, November 20). Autism Research Centre. https://www.autismresearchcentre.com/tests/faux-pas-test-adult/

Fernández-Modamio, M., Arrieta-Rodríguez, M., Bengochea-Seco, R., Santacoloma-Cabero, I., Gómez De Tojeiro-Roce, J., García-Polavieja, B., González-Fraile, E., Martín-Carrasco, M., Griffin, K., & Gil-Sanz, D. (2018). Faux-Pas Test: A Proposal of a Standardized Short Version. *Clinical Schizophrenia & Related Psychoses*, *aop*, CSRP.FEAR.061518. https://doi.org/10.3371/CSRP.FEAR.061518

Fett, A.-K. J., Viechtbauer, W., Dominguez, M.-G., Penn, D. L., van Os, J., & Krabbendam, L. (2011). The relationship between neurocognition and social cognition with functional outcomes in schizophrenia: A meta-analysis. *Neuroscience & Biobehavioral Reviews*, *35*(3), 573–588. https://doi.org/10.1016/j.neubiorev.2010.07.001

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

Freed, J., McBean, K., Adams, C., Lockton, E., Nash, M., & Law, J. (2015). Performance of children with social communication disorder on the Happé Strange Stories: Physical and mental state responses and relationship to language ability. *Journal of Communication Disorders*, *55*, 1–14. https://doi.org/10.1016/j.jcomdis.2015.03.002

Fretland, R. A., Andersson, S., Sundet, K., Andreassen, O. A., Melle, I., & Vaskinn, A. (2015). Theory of mind in schizophrenia: Error types and associations with symptoms. *Schizophrenia Research*, *162*(1–3), 42–46. https://doi.org/10.1016/j.schres.2015.01.024

Frick, C., Lang, S., Kotchoubey, B., Sieswerda, S., Dinu-Biringer, R., Berger, M., Veser, S., Essig, M., & Barnow, S. (2012). Hypersensitivity in Borderline Personality Disorder during Mindreading. *PLoS ONE*, *7*(8), e41650. https://doi.org/10.1371/journal.pone.0041650

Frith, C. D., & Frith, U. (2008). Implicit and Explicit Processes in Social Cognition. *Neuron*, *60*(3), 503–510. https://doi.org/10.1016/j.neuron.2008.10.032

Frith, C. D., & Frith, U. (2012). Mechanisms of Social Cognition. *Annual Review of Psychology*, *63*(1), 287–313. https://doi.org/10.1146/annurev-psych-120710-100449

Fuhrmann, D., Knoll, L. J., & Blakemore, S.-J. (2015). Adolescence as a Sensitive Period of Brain Development. *Trends in Cognitive Sciences*, *19*(10), 558–566. https://doi.org/10.1016/j.tics.2015.07.008

Gräfe, K., Zipfel, S., Herzog, W., & Löwe, B. (2004). Screening psychischer Störungen mit dem "Gesundheitsfragebogen für Patienten (PHQ-D)". *Diagnostica*, *50*(4), 171–181. https://doi.org/10.1026/0012-1924.50.4.171

Green, M. F., Bearden, C. E., Cannon, T. D., Fiske, A. P., Hellemann, G. S., Horan, W. P., Kee, K., Kern, R. S., Lee, J., Sergi, M. J., Subotnik, K. L., Sugar, C. A., Ventura, J., Yee, C. M., & Nuechterlein, K. H. (2012). Social Cognition in Schizophrenia, Part 1: Performance Across Phase of Illness. *Schizophrenia Bulletin*, *38*(4), 854–864. https://doi.org/10.1093/schbul/sbq171

Green, M. F., Horan, W. P., & Lee, J. (2015). Social cognition in schizophrenia. *Nature Reviews Neuroscience*, *16*(10), 620–631. https://doi.org/10.1038/nrn4005

Green, M. F., Penn, D. L., Bentall, R., Carpenter, W. T., Gaebel, W., Gur, C., Kring, A. M., Park, S., Silverstein, M., & Heinssen, R. (2008). Social Cognition in Schizophrenia: An NIMH Workshop on Definitions, Assessment, and Research Opportunities. *Social Cognition in Schizophrenia*.

Gregory, C., Lough, S., Stone, V., Erzinclioglu, S., Martin, L., Baron-Cohen, S., & Hodges, J. R. (2002). Theory of mind in patients with frontal variant frontotemporal dementia and Alzheimer's disease: Theoretical and practical implications. *Brain*, *125*(4), 752–764. https://doi.org/10.1093/brain/awf079

Grosse Wiesmann, C., Friederici, A. D., Singer, T., & Steinbeis, N. (2017). Implicit and explicit false belief development in preschool children. *Developmental Science*, *20*(5), e12445. https://doi.org/10.1111/desc.12445

Gunderson, J. G. (2007). Disturbed Relationships as a Phenotype for Borderline Personality Disorder. *American Journal of Psychiatry*, *164*(11), 1637–1640. https://doi.org/10.1176/appi.ajp.2007.07071125

Hanegraaf, L., Van Baal, S., Hohwy, J., & Verdejo-Garcia, A. (2021). A systematic review and meta-analysis of 'Systems for Social Processes' in borderline personality and substance use disorders. *Neuroscience & Biobehavioral Reviews*, *127*, 572–592. https://doi.org/10.1016/j.neubiorev.2021.04.013

Hanson, J. L., Nacewicz, B. M., Sutterer, M. J., Cayo, A. A., Schaefer, S. M., Rudolph, K. D., Shirtcliff, E. A., Pollak, S. D., & Davidson, R. J. (2015). Behavioral Problems After Early Life Stress: Contributions of the Hippocampus and

Amygdala. *Biological Psychiatry*, *77*(4), 314–323. https://doi.org/10.1016/j.biopsych.2014.04.020

Happé, F., Cook, J. L., & Bird, G. (2017). The Structure of Social Cognition: In(ter)dependence of Sociocognitive Processes. *Annual Review of Psychology*, *68*(1), 243–267. https://doi.org/10.1146/annurev-psych-010416-044046

Happé, F., & Frith, U. (1996). The neuropsychology of autism. *Brain*, *119*(4), 1377–1400. https://doi.org/10.1093/brain/119.4.1377

Happé, F. G. E. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders*, *24*(2), 129–154. https://doi.org/10.1007/BF02172093

Happé, F. G. E. (1995). The Role of Age and Verbal Ability in the Theory of Mind Task Performance of Subjects with Autism. *Child Development*, *66*(3), 843. https://doi.org/10.2307/1131954

Happé, F. G. E., Winner, E., & Brownell, H. (1998). The getting of wisdom: Theory of mind in old age. *Developmental Psychology*, *34*(2), 358–362. https://doi.org/10.1037/0012-1649.34.2.358

Harari, H., Shamay-Tsoory, S. G., Ravid, M., & Levkovitz, Y. (2010). Double dissociation between cognitive and affective empathy in borderline personality disorder. *Psychiatry Research*, *175*(3), 277–279. https://doi.org/10.1016/j.psychres.2009.03.002

Harkness, K., Sabbagh, M., Jacobson, J., Chowdrey, N., & Chen, T. (2005). Enhanced accuracy of mental state decoding in dysphoric college students. *Cognition & Emotion*, *19*(7), 999–1025. https://doi.org/10.1080/02699930541000110

Harwood, M. D., & Farrar, M. J. (2006). Conflicting emotions: The connection between affective perspective taking and theory of mind. *British Journal of Developmental Psychology*, *24*(2), 401–418. https://doi.org/10.1348/026151005X50302

Hasson-Ohayon, I., Avidan-Msika, M., Mashiach-Eizenberg, M., Kravetz, S., Rozencwaig, S., Shalev, H., & Lysaker, P. H. (2015). Metacognitive and social cognition approaches to understanding the impact of schizophrenia on social quality of life. *Schizophrenia Research*, *161*(2–3), 386–391. https://doi.org/10.1016/j.schres.2014.11.008

Hayward, E. O., & Homer, B. D. (2017). Reliability and validity of advanced theory-of-mind measures in middle childhood and adolescence. *British Journal of Developmental Psychology*, *35*(3), 454–462. https://doi.org/10.1111/bjdp.12186

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Henry, J. D., Phillips, L. H., Ruffman, T., & Bailey, P. E. (2013). A meta-analytic review of age differences in theory of mind. *Psychology and Aging*, *28*(3), 826–839. https://doi.org/10.1037/a0030677

Herpertz, S. C., & Bertsch, K. (2014). The social-cognitive basis of personality disorders. *Current Opinion in Psychiatry,* *27*(1), 73–77. https://doi.org/10.1097/YCO.0000000000000026

Hollebrandse, B., Van Hout, A., & Hendriks, P. (2014). Children's first and second-order false-belief reasoning in a verbal and a low-verbal task. *Synthese*, *191*(3), 321–333. https://doi.org/10.1007/s11229-012-0169-9

Holzner, B., Giesinger, J. M., Pinggera, J., Zugal, S., Schöpf, F., Oberguggenberger, A. S., Gamper, E. M., Zabernigg, A., Weber, B., & Rumpold, G. (2012). The Computer-based Health Evaluation Software (CHES): A software for electronic patient-reported outcome monitoring. *BMC Medical Informatics and Decision Making*, *12*, 126. https://doi.org/10.1186/1472-6947-12-126

Huang, Y. L., Fonagy, P., Feigenbaum, J., Montague, P. R., Nolte, T., & and London Personality and Mood Disorder Research Consortium. (2020). Multidirectional Pathways between Attachment, Mentalizing, and Posttraumatic Stress Symptomatology in the Context of Childhood Trauma. *Psychopathology*, *53*(1), 48–58. https://doi.org/10.1159/000506406

Hughes, C., Adlam, A., Happé, F., Jackson, J., Taylor, A., & Caspi, A. (2000). Good Test-Retest Reliability for Standard and Advanced False-Belief Tasks across a Wide Range of Abilities. *Journal of Child Psychology and Psychiatry*, *41*(4), 483–490. https://doi.org/10.1111/1469-7610.00633

Hughes, C., & Devine, R. T. (2015). Individual Differences in Theory of Mind From Preschool to Adolescence: Achievements and Directions. *Child Development Perspectives*, *9*(3), 149–153. https://doi.org/10.1111/cdep.12124

Humphrey, G., & Dumontheil, I. (2016). Development of Risk-Taking, Perspective-Taking, and Inhibitory Control During Adolescence. *Developmental Neuropsychology*, *41*(1–2), 59–76. https://doi.org/10.1080/87565641.2016.1161764

Im-Bolter, N., Agostino, A., & Owens-Jaffray, K. (2016). Theory of mind in middle childhood and early adolescence: Different from before? *Journal of Experimental Child Psychology*, *149*, 98–115. https://doi.org/10.1016/j.jecp.2015.12.006

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., Sanislow, C., & Wang, P. (2010). Research Domain Criteria (RDoC): Toward a New Classification Framework for Research on Mental Disorders. *American Journal of Psychiatry, 167*(7), 748–751. https://doi.org/10.1176/appi.ajp.2010.09091379

Janssen, P. G. J., Van Est, L. A. C., Hilbink, M., Gubbels, L., Egger, J., Cillessen, A. H. N., & Van Ee, E. (2022). Social cognitive performance in posttraumatic stress disorder: A meta-analysis. *Journal of Affective Disorders*, *297*, 35–44. https://doi.org/10.1016/j.jad.2021.09.082

Jin, A. B., Steding, L. H., & Webb, A. K. (2015). Reduced emotional and cardiovascular reactivity to emotionally evocative stimuli in major depressive disorder. *International Journal of Psychophysiology*, *97*(1), 66–74. https://doi.org/10.1016/j.ijpsycho.2015.04.014

Kaltefleiter, L. J., Sodian, B., Kristen-Antonow, S., Grosse Wiesmann, C., & Schuwerk, T. (2021). Does syntax play a role in Theory of Mind development before the age of 3 years? *Infant Behavior and Development*, *64*, 101575. https://doi.org/10.1016/j.infbeh.2021.101575

Karam, E. G., Friedman, M. J., Hill, E. D., Kessler, R. C., McLaughlin, K. A., Petukhova, M., Sampson, L., Shahly, V., Angermeyer, M. C., Bromet, E. J., De Girolamo, G., De Graaf, R., Demyttenaere, K., Ferry, F., Florescu, S. E., Haro, J. M., He, Y., Karam, A. N., Kawakami, N., … Koenen, K. C. (2014). CUMULATIVE TRAUMAS

AND RISK THRESHOLDS: 12-MONTH PTSD IN THE WORLD MENTAL HEALTH (WMH) SURVEYS: Cumulative Trauma and PTSD. *Depression and Anxiety*, *31*(2), 130–142. https://doi.org/10.1002/da.22169

Kessler, R. C., Aguilar-Gaxiola, S., Alonso, J., Benjet, C., Bromet, E. J., Cardoso, G., Degenhardt, L., De Girolamo, G., Dinolova, R. V., Ferry, F., Florescu, S., Gureje, O., Haro, J. M., Huang, Y., Karam, E. G., Kawakami, N., Lee, S., Lepine, J.-P., Levinson, D., … Koenen, K. C. (2017). Trauma and PTSD in the WHO World Mental Health Surveys. *European Journal of Psychotraumatology*, *8*(sup5), 1353383. https://doi.org/10.1080/20008198.2017.1353383

Kieling, C., Buchweitz, C., Caye, A., Silvani, J., Ameis, S. H., Brunoni, A. R., Cost, K. T., Courtney, D. B., Georgiades, K., Merikangas, K. R., Henderson, J. L., Polanczyk, G. V., Rohde, L. A., Salum, G. A., & Szatmari, P. (2024). Worldwide Prevalence and Disability From Mental Disorders Across Childhood and Adolescence: Evidence From the Global Burden of Disease Study. *JAMA Psychiatry*, *81*(4), 347. https://doi.org/10.1001/jamapsychiatry.2023.5051

Kimhi, Y. (2014). Theory of Mind Abilities and Deficits in Autism Spectrum Disorders. *Topics in Language Disorders*, *34*(4), 329–343. https://doi.org/10.1097/TLD.0000000000000033

Kittel, A. F. D., Olderbak, S., & Wilhelm, O. (2022). Sty in the Mind's Eye: A Meta-Analytic Investigation of the Nomological Network and Internal Consistency of the "Reading the Mind in the Eyes" Test. *Assessment*, *29*(5), 872–895. https://doi.org/10.1177/1073191121996469

Kleindienst, N., Jungkunz, M., & Bohus, M. (2020). A proposed severity classification of borderline symptoms using the borderline symptom list (BSL-23). *Borderline*

Personality Disorder and Emotion Dysregulation, *7*(1), 11. https://doi.org/10.1186/s40479-020-00126-6

Kloo, D., Kristen-Antonow, S., & Sodian, B. (2020). Progressing from an implicit to an explicit false belief understanding: A matter of executive control? *International Journal of Behavioral Development*, *44*(2), 107–115. https://doi.org/10.1177/0165025419850901

Knopp, M., Burghardt, J., Oppenauer, C., Meyer, B., Moritz, S., & Sprung, M. (2024). Affective and cognitive Theory of Mind in patients with alcohol use disorder: Associations with symptoms of depression, anxiety, and somatization. *Journal of Substance Use and Addiction Treatment*, *157*, 209227. https://doi.org/10.1016/j.josat.2023.209227

Koenen, K. C., Ratanatharathorn, A., Ng, L., McLaughlin, K. A., Bromet, E. J., Stein, D. J., Karam, E. G., Meron Ruscio, A., Benjet, C., Scott, K., Atwoli, L., Petukhova, M., Lim, C. C. W., Aguilar-Gaxiola, S., Al-Hamzawi, A., Alonso, J., Bunting, B., Ciutan, M., De Girolamo, G., … Kessler, R. C. (2017). Posttraumatic stress disorder in the World Mental Health Surveys. *Psychological Medicine*, *47*(13), 2260–2274. https://doi.org/10.1017/S0033291717000708

Koerber, S., & Osterhaus, C. (2020). Some but not all aspects of (advanced) theory of mind predict loneliness. *British Journal of Developmental Psychology*, *38*(1), 144–148. https://doi.org/10.1111/bjdp.12302

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9. *Journal of General Internal Medicine*, *16*(9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Kubinger, K. D. (2005). Psychological Test Calibration Using the Rasch Model—Some Critical Suggestions on Traditional Approaches. *International Journal of Testing*, *5*(4), 377–394. https://doi.org/10.1207/s15327574ijt0504_3

Kvarstein, E. H., Folmo, E., Antonsen, B. T., Normann-Eide, E., Pedersen, G., & Wilberg, T. (2020). Social Cognition Capacities as Predictors of Outcome in Mentalization-Based Treatment (MBT). *Frontiers in Psychiatry*, *11*, 691. https://doi.org/10.3389/fpsyt.2020.00691

Lahera, G., Boada, L., Pousa, E., Mirapeix, I., Morón-Nozaleda, G., Marinas, L., Gisbert, L., Pamiàs, M., & Parellada, M. (2014). Movie for the Assessment of Social Cognition (MASC): Spanish Validation. *Journal of Autism and Developmental Disorders*, *44*(8), 1886–1896. https://doi.org/10.1007/s10803-014-2061-6

Lazarus, S. A., Cheavens, J. S., Festa, F., & Zachary Rosenthal, M. (2014). Interpersonal functioning in borderline personality disorder: A systematic review of behavioral and laboratory-based assessments. *Clin Psychol Rev*, *34*(3), 193–205. https://doi.org/10.1016/j.cpr.2014.01.007

Lecce, S., Bianco, F., Demicheli, P., & Cavallini, E. (2014). Training Preschoolers on First-Order False Belief Understanding: Transfer on Advanced ToM Skills and Metamemory. *Child Development*, *85*(6), 2404–2418. https://doi.org/10.1111/cdev.12267

Lecce, S., Caputi, M., Pagnin, A., & Banerjee, R. (2017). Theory of mind and school achievement: The mediating role of social competence. *Cognitive Development*, *44*, 85–97. https://doi.org/10.1016/j.cogdev.2017.08.010

Lecce, S., & Devine, R. T. (2021). Theory of mind at school: Academic outcomes and the influence of the school context. *Infant and Child Development*, *31*(1), e2274. https://doi.org/10.1002/icd.2274

Li, E. T., Carracher, E., & Bird, T. (2020). Linking childhood emotional abuse and adult depressive symptoms: The role of mentalizing incapacity. *Child Abuse & Neglect*, *99*, 104253. https://doi.org/10.1016/j.chiabu.2019.104253

Lieb, K., Zanarini, M. C., Schmahl, C., Linehan, M. M., & Bohus, M. (2004). Borderline personality disorder. *The Lancet*, *364*(9432), 453–461. https://doi.org/10.1016/S0140-6736(04)16770-6

Longobardi, E., Spataro, P., & Rossi-Arnaud, C. (2019). Direct and Indirect Associations of Empathy, Theory of Mind, and Language with Prosocial Behavior: Gender Differences in Primary School Children. *The Journal of Genetic Psychology*, *180*(6), 266–279. https://doi.org/10.1080/00221325.2019.1653817

Luke, N., & Banerjee, R. (2013). Differentiated associations between childhood maltreatment experiences and social understanding: A meta-analysis and systematic review. *Developmental Review*, *33*(1), 1–28. https://doi.org/10.1016/j.dr.2012.10.001

Lupien, S. J., McEwen, B. S., Gunnar, M. R., & Heim, C. (2009). Effects of stress throughout the lifespan on the brain, behaviour and cognition. *Nature Reviews Neuroscience*, *10*(6), 434–445. https://doi.org/10.1038/nrn2639

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*(1), 297–325. https://doi.org/10.1146/annurev-clinpsy-071919-015355

Mair, P., & Hatzinger, R. (2007). Extended Rasch Modeling: The eRm Package for the Application of IRT Models in R. *Journal of Statistical Software*, *20*, 1–20. https://doi.org/10.18637/jss.v020.i09

Malda-Castillo, J., Browne, C., & Perez-Algorta, G. (2019). Mentalization-based treatment and its evidence-base status: A systematic literature review. *Psychology and Psychotherapy: Theory, Research and Practice*, *92*(4), 465–498. https://doi.org/10.1111/papt.12195

Maleki, G., Zabihzadeh, A., Richman, M. J., Demetrovics, Z., & Mohammadnejad, F. (2020). Decoding and reasoning mental states in major depression and social anxiety disorder. *BMC Psychiatry*, *20*(1), 463. https://doi.org/10.1186/s12888-020-02873-w

Martino, D. J., Bucay, D., Butman, J. T., & Allegri, R. F. (2007). Neuropsychological frontal impairments and negative symptoms in schizophrenia. *Psychiatry Research*, *152*(2–3), 121–128. https://doi.org/10.1016/j.psychres.2006.03.002

Masters, G. N. (1982). A rasch model for partial credit scoring. *Psychometrika*, *47*(2), 149–174. https://doi.org/10.1007/BF02296272

Mathersul, D., McDonald, S., & Rushby, J. A. (2013). Understanding advanced theory of mind and empathy in high-functioning adults with autism spectrum disorder. *Journal of Clinical and Experimental Neuropsychology*, *35*(6), 655–668. https://doi.org/10.1080/13803395.2013.809700

Mayes, L. C., Klin, A., Tercyak, K. P., Cicchetti, D. V., & Cohen, D. J. (1996). Test-Retest Reliability for False-Belief Tasks. *Journal of Child Psychology and Psychiatry*, *37*(3), 313–319. https://doi.org/10.1111/j.1469-7610.1996.tb01408.x

Mazza, M., Giusti, L., Albanese, A., Mariano, M., Pino, M. C., & Roncone, R. (2012). Social cognition disorders in military police officers affected by posttraumatic stress disorder after the attack of An-Nasiriyah in Iraq 2006. *Psychiatry Research*, *198*(2), 248–252. https://doi.org/10.1016/j.psychres.2011.11.027

Mazza, M., Tempesta, D., Pino, M. C., Nigri, A., Catalucci, A., Guadagni, V., Gallucci, M., Iaria, G., & Ferrara, M. (2015). Neural activity related to cognitive and emotional empathy in post-traumatic stress disorder. *Behavioural Brain Research*, *282*, 37–45. https://doi.org/10.1016/j.bbr.2014.12.049

McCrory, E. J., & Viding, E. (2015). The theory of latent vulnerability: Reconceptualizing the link between childhood maltreatment and psychiatric disorder. *Development and Psychopathology*, *27*(2), 493–505. https://doi.org/10.1017/S0954579415000115

McLaren, V., Gallagher, M., Hopwood, C. J., & Sharp, C. (2022). Hypermentalizing and Borderline Personality Disorder: A Meta-Analytic Review. *American Journal of Psychotherapy*, *75*(1), 21–31. https://doi.org/10.1176/appi.psychotherapy.20210018

Meinhardt-Injac, B., Daum, M. M., & Meinhardt, G. (2020). Theory of mind development from adolescence to adulthood: Testing the two-component model. *British Journal of Developmental Psychology*, *38*(2), 289–303. https://doi.org/10.1111/bjdp.12320

Michelini, G., Palumbo, I. M., DeYoung, C. G., Latzman, R. D., & Kotov, R. (2021). Linking RDoC and HiTOP: A new interface for advancing psychiatric nosology and neuroscience. *Clinical Psychology Review*, *86*, 102025. https://doi.org/10.1016/j.cpr.2021.102025

Miller, S. A. (2009). Children's understanding of second-order mental states. *Psychological Bulletin*, *135*(5), 749–773. https://doi.org/10.1037/a0016854

Miller, S. A. (2013). Children's Understanding of Second-Order False Belief: Comparisons of Content and Method of Assessment. *Infant and Child Development*, *22*(6), 649–658. https://doi.org/10.1002/icd.1810

Milligan, K., Astington, J. W., & Dack, L. A. (2007). Language and Theory of Mind: Meta-Analysis of the Relation Between Language Ability and False-belief Understanding. *Child Development*, *78*(2), 622–646. https://doi.org/10.1111/j.1467-8624.2007.01018.x

Molenberghs, P., Johnson, H., Henry, J. D., & Mattingley, J. B. (2016). Understanding the minds of others: A neuroimaging meta-analysis. *Neuroscience & Biobehavioral Reviews*, *65*, 276–291. https://doi.org/10.1016/j.neubiorev.2016.03.020

Montag, C., Dziobek, I., Richter, I. S., Neuhaus, K., Lehmann, A., Sylla, R., Heekeren, H. R., Heinz, A., & Gallinat, J. (2011). Different aspects of theory of mind in paranoid schizophrenia: Evidence from a video-based assessment. *Psychiatry Research*, *186*(2–3), 203–209. https://doi.org/10.1016/j.psychres.2010.09.006

Montag, C., Ehrlich, A., Neuhaus, K., Dziobek, I., Heekeren, H. R., Heinz, A., & Gallinat, J. (2010). Theory of mind impairments in euthymic bipolar patients. *Journal of*

Affective Disorders, *123*(1–3), 264–269. https://doi.org/10.1016/j.jad.2009.08.017

Montag, C., Neuhaus, K., Lehmann, A., Krüger, K., Dziobek, I., Heekeren, H. R., Heinz, A., & Gallinat, J. (2012). Subtle deficits of cognitive theory of mind in unaffected first-degree relatives of schizophrenia patients. *European Archives of Psychiatry and Clinical Neuroscience*, *262*(3), 217–226. https://doi.org/10.1007/s00406-011-0250-2

Mor, N., & Winquist, J. (2002). Self-focused attention and negative affect: A meta-analysis. *Psychological Bulletin*, *128*(4), 638–662. https://doi.org/10.1037/0033-2909.128.4.638

Moran, J. M. (2013). Lifespan development: The effects of typical aging on theory of mind. *Behavioural Brain Research*, *237*, 32–40. https://doi.org/10.1016/j.bbr.2012.09.020

Moriarty, A. S., Gilbody, S., McMillan, D., & Manea, L. (2015). Screening and case finding for major depressive disorder using the Patient Health Questionnaire (PHQ-9): A meta-analysis. *General Hospital Psychiatry*, *37*(6), 567–576. https://doi.org/10.1016/j.genhosppsych.2015.06.012

Müller, S., Wendt, L. P., Spitzer, C., Masuhr, O., Back, S. N., & Zimmermann, J. (2022). A Critical Evaluation of the Reflective Functioning Questionnaire (RFQ). *Journal of Personality Assessment*, *104*(5), 613–627. https://doi.org/10.1080/00223891.2021.1981346

Mundy, P., Sigman, M., Ungerer, J., & Sherman, T. (1986). DEFINING THE SOCIAL DEFICITS OF AUTISM: THE CONTRIBUTION OF NON-VERBAL

COMMUNICATION MEASURES. *Journal of Child Psychology and Psychiatry*, *27*(5), 657–669. https://doi.org/10.1111/j.1469-7610.1986.tb00190.x

Murray, C. J. L. (2022). The Global Burden of Disease Study at 30 years. *Nature Medicine*, *28*(10), 2019–2026. https://doi.org/10.1038/s41591-022-01990-1

Murray, K., Johnston, K., Cunnane, H., Kerr, C., Spain, D., Gillan, N., Hammond, N., Murphy, D., & Happé, F. (2017). A new test of advanced theory of mind: The "Strange Stories Film Task" captures social processing differences in adults with autism spectrum disorders. *Autism Research*, *10*(6), 1120–1132. https://doi.org/10.1002/aur.1744

National Institutes of Mental Health. (2020, September 25). *NIMH » About RDoC*. Research Domain Criteria (RDoC). https://www.nimh.nih.gov/research/research-funded-by-nimh/rdoc/about-rdoc.shtml

Navarro, E., Goring, S. A., & Conway, A. R. A. (2021). The Relationship between Theory of Mind and Intelligence: A Formative g Approach. *Journal of Intelligence*, *9*(1), 11. https://doi.org/10.3390/jintelligence9010011

Nelson, P. B., Adamson, L. B., & Bakeman, R. (2008). Toddlers' joint engagement experience facilitates preschoolers' acquisition of theory of mind. *Developmental Science*, *11*(6), 847–852. https://doi.org/10.1111/j.1467-7687.2008.00733.x

Németh, N., Mátrai, P., Hegyi, P., Czéh, B., Czopf, L., Hussain, A., Pammer, J., Szabó, I., Solymár, M., Kiss, L., Hartmann, P., Szilágyi, Á. L., Kiss, Z., & Simon, M. (2018). Theory of mind disturbances in borderline personality disorder: A meta-analysis. *Psychiatry Research*, *270*, 143–153. https://doi.org/10.1016/j.psychres.2018.08.049

Németh, N., Péterfalvi, Á., Czéh, B., Tényi, T., & Simon, M. (2020). Examining the Relationship Between Executive Functions and Mentalizing Abilities of Patients With Borderline Personality Disorder. *Frontiers in Psychology*, *11*, 1583. https://doi.org/10.3389/fpsyg.2020.01583

Nestor, B. A., Sutherland, S., & Garber, J. (2022). Theory of mind performance in depression: A meta-analysis. *Journal of Affective Disorders*, *303*, 233–244. https://doi.org/10.1016/j.jad.2022.02.028

Nettle, D., & Liddle, B. (2008). Agreeableness is related to social-cognitive, but not social-perceptual, theory of mind. *European Journal of Personality*, *22*(4), 323–335. https://doi.org/10.1002/per.672

Nietlisbach, G., Maercker, A., Rösler, W., & Haker, H. (2010). Are Empathic Abilities Impaired in Posttraumatic Stress Disorder? *Psychological Reports*, *106*(3), 832–844. https://doi.org/10.2466/pr0.106.3.832-844

Normann-Eide, E., Antonsen, B. T., Kvarstein, E. H., Pedersen, G., Vaskinn, A., & Wilberg, T. (2020). Are Impairments in Theory of Mind Specific to Borderline Personality Disorder? *Journal of Personality Disorders*, *34*(6), 827–841. https://doi.org/10.1521/pedi_2019_33_417

O'Brien, M., Miner Weaver, J., Nelson, J. A., Calkins, S. D., Leerkes, E. M., & Marcovitch, S. (2011). Longitudinal associations between children's understanding of emotions and theory of mind. *Cognition & Emotion*, *25*(6), 1074–1086. https://doi.org/10.1080/02699931.2010.518417

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*. https://doi.org/10.3389/fpsyg.2015.01503

Onuoha, R. C., Quintana, D. S., Lyvers, M., & Guastella, A. J. (2016). A Meta-analysis of Theory of Mind in Alcohol Use Disorders. *Alcohol and Alcoholism*, *51*(4), 410–415. https://doi.org/10.1093/alcalc/agv137

Osterhaus, C., & Bosacki, S. L. (2022). Looking for the lighthouse: A systematic review of advanced theory-of-mind tests beyond preschool. *Developmental Review*, *64*, 101021. https://doi.org/10.1016/j.dr.2022.101021

Osterhaus, C., & Koerber, S. (2021). The Development of Advanced Theory of Mind in Middle Childhood: A Longitudinal Study From Age 5 to 10 Years. *Child Development*, *92*(5), 1872–1888. https://doi.org/10.1111/cdev.13627

Osterhaus, C., & Koerber, S. (2023). The complex associations between scientific reasoning and advanced theory of mind. *Child Development*, *94*(1). https://doi.org/10.1111/cdev.13860

Osterhaus, C., Koerber, S., & Sodian, B. (2016). Scaling of Advanced Theory-of-Mind Tasks. *Child Development*, *87*(6), 1971–1991. https://doi.org/10.1111/cdev.12566

Otte, C., Gold, S. M., Penninx, B. W., Pariante, C. M., Etkin, A., Fava, M., Mohr, D. C., & Schatzberg, A. F. (2016). Major depressive disorder. *Nature Reviews Disease Primers*, *2*(1), 16065. https://doi.org/10.1038/nrdp.2016.65

Ozer, E. J., Best, S. R., Lipsey, T. L., & Weiss, D. S. (2003). Predictors of posttraumatic stress disorder and symptoms in adults: A meta-analysis. *Psychological Bulletin*, *129*(1), 52–73. https://doi.org/10.1037/0033-2909.129.1.52

Pagnoni, I., Gobbi, E., Alaimo, C., Campana, E., Rossi, R., Manenti, R., Balconi, M., & Cotelli, M. (2022). The relationship between theory of mind and executive functions in major depressive disorders: A review. *Frontiers in Psychiatry*, *13*, 980392. https://doi.org/10.3389/fpsyt.2022.980392

Pan, Z., Park, C., Brietzke, E., Zuckerman, H., Rong, C., Mansur, R. B., Fus, D., Subramaniapillai, M., Lee, Y., & McIntyre, R. S. (2019). Cognitive impairment in major depressive disorder. *CNS Spectrums*, *24*(1), 22–29. https://doi.org/10.1017/S1092852918001207

Penn, D. L., Sanna, L. J., & Roberts, D. L. (2008). Social Cognition in Schizophrenia: An Overview. *Schizophrenia Bulletin*, *34*(3), 408–411. https://doi.org/10.1093/schbul/sbn014

Perner, J., Leekam, S. R., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology*, *5*(2), 125–137. https://doi.org/10.1111/j.2044-835X.1987.tb01048.x

Perner, J., & Wimmer, H. (1985). "John thinks that Mary thinks that…" attribution of second-order beliefs by 5- to 10-year-old children. *Journal of Experimental Child Psychology*, *39*(3), 437–471. https://doi.org/10.1016/0022-0965(85)90051-7

Peskin, J., & Ardino, V. (2003). Representing the Mental World in Children's Social Behavior: Playing Hide-and-Seek and Keeping a Secret. *Social Development*, *12*(4), 496–512. https://doi.org/10.1111/1467-9507.00245

Petersen, R., Brakoulias, V., & Langdon, R. (2016). An experimental investigation of mentalization ability in borderline personality disorder. *Comprehensive Psychiatry*, *64*, 12–21. https://doi.org/10.1016/j.comppsych.2015.10.004

Peterson, C. C., Wellman, H. M., & Liu, D. (2005). Steps in Theory-of-Mind Development for Children With Deafness or Autism. *Child Development*, *76*(2), 502–517. https://doi.org/10.1111/j.1467-8624.2005.00859.x

Peyroux, E., Prost, Z., Danset-Alexandre, C., Brenugat-Herne, L., Carteau-Martin, I., Gaudelus, B., Jantac, C., Attali, D., Amado, I., Graux, J., Houy-Durand, E., Plasse, J., & Franck, N. (2019). From "under" to "over" social cognition in schizophrenia: Is there distinct profiles of impairments according to negative and positive symptoms? *Schizophrenia Research: Cognition*, *15*, 21–29. https://doi.org/10.1016/j.scog.2018.10.001

Pfaltz, M. C., McAleese, S., Saladin, A., Meyer, A. H., Stoecklin, M., Opwis, K., Dammann, G., & Martin-Soelch, C. (2013). *The reading the mind in the eyes test: Test-retest reliability and preliminary psychometric properties of the German version*. https://doi.org/10.5167/UZH-87335

Pickup, G. J., & Frith, C. D. (2001). Theory of mind impairments in schizophrenia: Symptomatology, severity and specificity. *Psychological Medicine*, *31*(2), 207–220. https://doi.org/10.1017/S0033291701003385

Pijnenborg, G. H. M., Spikman, J. M., Jeronimus, B. F., & Aleman, A. (2013). Insight in schizophrenia: Associations with empathy. *European Archives of Psychiatry and Clinical Neuroscience*, *263*(4), 299–307. https://doi.org/10.1007/s00406-012-0373-0

Pinkham, A. E., Penn, D. L., Green, M. F., Buck, B., Healey, K., & Harvey, P. D. (2013). The Social Cognition Psychometric Evaluation Study: Results of the Expert Survey and RAND Panel. *Schizophrenia Bulletin*, *40*(4), 813–823. https://doi.org/10.1093/schbul/sbt081

Pitman, R. K., Rasmusson, A. M., Koenen, K. C., Shin, L. M., Orr, S. P., Gilbertson, M. W., Milad, M. R., & Liberzon, I. (2012). Biological studies of post-traumatic stress disorder. *Nature Reviews Neuroscience*, *13*(11), 769–787. https://doi.org/10.1038/nrn3339

Plana, I., Lavoie, M.-A., Battaglia, M., & Achim, A. M. (2014). A meta-analysis and scoping review of social cognition performance in social phobia, posttraumatic stress disorder and other anxiety disorders. *Journal of Anxiety Disorders*, *28*(2), 169–177. https://doi.org/10.1016/j.janxdis.2013.09.005

Pluta, A., Kulesza, M., Grzegorzewski, P., & Kucharska, K. (2018). Assessing advanced theory of mind and alexithymia in patients suffering from enduring borderline personality disorder. *Psychiatry Research*, *261*, 436–441. https://doi.org/10.1016/j.psychres.2018.01.003

Porcelli, S., Van Der Wee, N., Van Der Werff, S., Aghajani, M., Glennon, J. C., Van Heukelum, S., Mogavero, F., Lobo, A., Olivera, F. J., Lobo, E., Posadas, M., Dukart, J., Kozak, R., Arce, E., Ikram, A., Vorstman, J., Bilderbeck, A., Saris, I., Kas, M. J., & Serretti, A. (2019). Social brain, social dysfunction and social withdrawal. *Neuroscience & Biobehavioral Reviews*, *97*, 10–33. https://doi.org/10.1016/j.neubiorev.2018.09.012

Pourmohammad, P., Imani, M., Goodarzi, M. A., & Sarafraz, M. R. (2021). Impaired complex theory of mind and low emotional self-awareness in outpatients with

borderline personality disorder compared to healthy controls: A cross-sectional study. *Journal of Psychiatric Research*, *143*, 445–450. https://doi.org/10.1016/j.jpsychires.2021.10.009

Preißler, S., Dziobek, I., Ritter, K., Heekeren, H. R., & Roepke, S. (2010). Social Cognition in Borderline Personality Disorder: Evidence for Disturbed Recognition of the Emotions, Thoughts, and Intentions of others. *Frontiers in Behavioral Neuroscience*, *4*, 182–182. https://doi.org/10.3389/fnbeh.2010.00182

Premack, D., & Woodruff, G. (1978). Does a chimpanzee have a theory of mind. *Behavioral and Brain Sciences*, *1*, 515–526. https://doi.org/10.1017/S0140525X00076512

Prins, A., Bovin, M. J., Smolenski, D. J., Marx, B. P., Kimerling, R., Jenkins-Guarnieri, M. A., Kaloupek, D. G., Schnurr, P. P., Kaiser, A. P., Leyva, Y. E., & Tiet, Q. Q. (2016). The Primary Care PTSD Screen for DSM-5 (PC-PTSD-5): Development and Evaluation Within a Veteran Primary Care Sample. *Journal of General Internal Medicine*, *31*(10), 1206–1211. https://doi.org/10.1007/s11606-016-3703-5

Quek, J., Bennett, C., Melvin, G. A., Saeedi, N., Gordon, M. S., & Newman, L. K. (2018). An investigation of the mentalization-based model of borderline pathology in adolescents. *Comprehensive Psychiatry*, *84*, 87–94. https://doi.org/10.1016/j.comppsych.2018.04.005

Quesque, F., Apperly, I., Baillargeon, R., Baron-Cohen, S., Becchio, C., Bekkering, H., Bernstein, D., Bertoux, M., Bird, G., Bukowski, H., Burgmer, P., Carruthers, P., Catmur, C., Dziobek, I., Epley, N., Erle, T. M., Frith, C., Frith, U., Galang, C. M.,

… Brass, M. (2024). Defining key concepts for mental state attribution. *Communications Psychology*, *2*(1), 29. https://doi.org/10.1038/s44271-024-00077-6

Quesque, F., & Rossetti, Y. (2020). What Do Theory-of-Mind Tasks Actually Measure? Theory and Practice. *Perspectives on Psychological Science*, *15*(2), 384–396. https://doi.org/10.1177/1745691619896607

Rasch, G., & Wright, B. D. (1980). *Probabilistic models for some intelligence and attainment tests*. Univ. of Chicago Press.

Richman, M. J., & Unoka, Z. (2015). Mental state decoding impairment in major depression and borderline personality disorder: Meta-analysis. *British Journal of Psychiatry*, *207*(6), 483–489. https://doi.org/10.1192/bjp.bp.114.152108

Rizopoulos, D. (2006). ltm: An *R* Package for Latent Variable Modeling and Item Response Theory Analyses. *Journal of Statistical Software*, *17*(5). https://doi.org/10.18637/jss.v017.i05

Roepke, S., Vater, A., Preißler, S., Heekeren, H. R., & Dziobek, I. (2013). Social cognition in borderline personality disorder. *Frontiers in Neuroscience*, *6*. https://doi.org/10.3389/fnins.2012.00195

Ronald, A., Happé, F., Hughes, C., & Plomin, R. (2005). Nice and Nasty Theory of Mind in Preschool Children: Nature and Nurture. *Social Development*, *14*(4), 664–684. https://doi.org/10.1111/j.1467-9507.2005.00323.x

Ruffman, T., Garnham, W., & Rideout, P. (2001). Social Understanding in Autism: Eye Gaze as a Measure of Core Insights. *Journal of Child Psychology and Psychiatry*, *42*(8), 1083–1094. https://doi.org/10.1111/1469-7610.00807

Sabbagh, M. A. (2004). Understanding orbitofrontal contributions to theory-of-mind reasoning: Implications for autism. *Brain and Cognition*, *55*(1), 209–219. https://doi.org/10.1016/j.bandc.2003.04.002

Safiye, T., Gutić, M., Dubljanin, J., Stojanović, T. M., Dubljanin, D., Kovačević, A., Zlatanović, M., Demirović, D. H., Nenezić, N., & Milidrag, A. (2023). Mentalizing, Resilience, and Mental Health Status among Healthcare Workers during the COVID-19 Pandemic: A Cross-Sectional Study. *International Journal of Environmental Research and Public Health*, *20*(8), 5594. https://doi.org/10.3390/ijerph20085594

Şandor, S., & İşcen, P. (2021). Faux-Pas Recognition Test: A Turkish adaptation study and a proposal of a standardized short version. *Applied Neuropsychology: Adult*, 1–9. https://doi.org/10.1080/23279095.2021.1909030

Sanefuji, W., Wada, K., Yamamoto, T., Mohri, I., & Taniike, M. (2014). Development of preference for conspecific faces in human infants. *Developmental Psychology*, *50*(4), 979–985. https://doi.org/10.1037/a0035205

Savla, G. N., Vella, L., Armstrong, C. C., Penn, D. L., & Twamley, E. W. (2013). Deficits in Domains of Social Cognition in Schizophrenia: A Meta-Analysis of the Empirical Evidence. *Schizophrenia Bulletin*.

Scandurra, C., Dolce, P., Vitelli, R., Esposito, G., Testa, R. J., Balsam, K. F., & Bochicchio, V. (2020). Mentalizing stigma: Reflective functioning as a protective factor against depression and anxiety in transgender and gender-nonconforming people. *Journal of Clinical Psychology*, *76*(9), 1613–1630. https://doi.org/10.1002/jclp.22951

Schaafsma, S. M., Pfaff, D. W., Spunt, R. P., & Adolphs, R. (2015). Deconstructing and reconstructing theory of mind. *Trends in Cognitive Sciences*, *19*(2), 65–72. https://doi.org/10.1016/j.tics.2014.11.007

Scheeren, A. M., De Rosnay, M., Koot, H. M., & Begeer, S. (2013). Rethinking theory of mind in high-functioning autism spectrum disorder. *Journal of Child Psychology and Psychiatry*, *54*(6), 628–635. https://doi.org/10.1111/jcpp.12007

Schneider, D., Bayliss, A. P., Becker, S. I., & Dux, P. E. (2012). Eye movements reveal sustained implicit processing of others' mental states. *Journal of Experimental Psychology: General*, *141*(3), 433–438. https://doi.org/10.1037/a0025458

Schurz, M., Radua, J., Tholen, M. G., Maliske, L., Margulies, D. S., Mars, R. B., Sallet, J., & Kanske, P. (2021). Toward a hierarchical model of social cognition: A neuroimaging meta-analysis and integrative review of empathy and theory of mind. *Psychological Bulletin*, *147*(3), 293–327. https://doi.org/10.1037/bul0000303

Schuwerk, T., Jarvers, I., Vuori, M., & Sodian, B. (2016). Implicit Mentalizing Persists beyond Early Childhood and Is Profoundly Impaired in Children with Autism Spectrum Condition. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01696

Schuwerk, T., Vuori, M., & Sodian, B. (2015). Implicit and explicit Theory of Mind reasoning in autism spectrum disorders: The impact of experience. *Autism*, *19*(4), 459–468. https://doi.org/10.1177/1362361314526004

Shamay-Tsoory, S. G. (2011). The Neural Bases for Empathy. *The Neuroscientist*, *17*(1), 18–24. https://doi.org/10.1177/1073858410379268

Shamay-Tsoory, S. G., & Aharon-Peretz, J. (2007). Dissociable prefrontal networks for cognitive and affective theory of mind: A lesion study. *Neuropsychologia*, *45*(13), 3054–3067. https://doi.org/10.1016/j.neuropsychologia.2007.05.021

Sharp, C., Fonagy, P., & Allen, J. G. (2012). Posttraumatic stress disorder: A social-cognitive perspective. *Clinical Psychology: Science and Practice*, *19*(3), 229–240. https://doi.org/10.1111/cpsp.12002

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.e1. https://doi.org/10.1016/j.jaac.2011.01.017

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

Shur, S., Shamay-Tsoory, S. G., & Levkovitz, Y. (2008). Integration of emotional and cognitive aspects of theory of mind in schizophrenia and its relation to prefrontal neurocognitive performance. *Cognitive Neuropsychiatry*, *13*(6), 472–490. https://doi.org/10.1080/13546800802490034

Slaughter, V., Imuta, K., Peterson, C. C., & Henry, J. D. (2015). Meta-Analysis of Theory of Mind and Peer Popularity in the Preschool and Early School Years. *Child Development*, *86*(4), 1159–1174. https://doi.org/10.1111/cdev.12372

Sodian, B., & Kristen-Antonow, S. (2015). Declarative joint attention as a foundation of theory of mind. *Developmental Psychology*, *51*(9), 1190–1200. https://doi.org/10.1037/dev0000039

Sodian, B., Kristen-Antonow, S., & Kloo, D. (2020). How Does Children's Theory of Mind Become Explicit? A Review of Longitudinal Findings. *Child Development Perspectives*, *14*(3), 171–177. https://doi.org/10.1111/cdep.12381

Spenser, K., Bull, R., Betts, L., & Winder, B. (2020). Underpinning Prosociality: Age related performance in theory of mind, empathic understanding, and moral reasoning. *Cognitive Development*, *56*, 100928. 9https://doi.org/10.1016/j.cogdev.2020.100928

Sprong, M., Schothorst, P., Vos, E., Hox, J., & Engeland, H. V. (2007). Theory of mind in schizophrenia: Meta-analysis. *The British Journal of Psychiatry*, *191*(1), 5–13. https://doi.org/10.1192/bjp.bp.107.035899

Sprung, M. (2010). Clinically Relevant Measures of Children's Theory of Mind and Knowledge about Thinking: Non-Standard and Advanced Measures. *Child and Adolescent Mental Health*, *15*(4), 204–216. https://doi.org/10.1111/j.1475-3588.2010.00568.x

STATISTIK AUSTRIA. (2023, June 27). *Bildungsstand der Bevölkerung*. STATISTIK AUSTRIA. https://www.statistik.at/statistiken/bevoelkerung-und-soziales/bildung/bildungsstand-der-bevoelkerung

Steinberg, L. (2005). Cognitive and affective development in adolescence. *Trends in Cognitive Sciences*, *9*(2), 69–74. https://doi.org/10.1016/j.tics.2004.12.005

Stone, V. E., Baron-Cohen, S., & Knight, R. T. (1998a). *Faux-pas Recognition Test(Adult Version)*. Retrieved August 24, 2024, from https://doi.org/10.1037/t73330-000

Stone, V. E., Baron-Cohen, S., & Knight, R. T. (1998b). Frontal Lobe Contributions to Theory of Mind. *Journal of Cognitive Neuroscience*, *10*(5), 640–656. https://doi.org/10.1162/089892998562942

Ströbele, A. (n.d.). *Faux-pas Test (Adult)—Deutsch*. Retrieved August 22, 2024, from https://www.autismresearchcentre.com/tests/faux-pas-test-adult/

Symeonidou, I., Dumontheil, I., Chow, W.-Y., & Breheny, R. (2016). Development of online use of theory of mind during adolescence: An eye-tracking study. *Journal of Experimental Child Psychology*, *149*, 81–97. https://doi.org/10.1016/j.jecp.2015.11.007

Tager-Flusberg, H. (2007). Evaluating the Theory-of-Mind Hypothesis of Autism. *Current Directions in Psychological Science*, *16*(6), 311–315. https://doi.org/10.1111/j.1467-8721.2007.00527.x

Tousignant, B., Sirois, K., Achim, A. M., Massicotte, E., & Jackson, P. L. (2017). A comprehensive assessment of social cognition from adolescence to adulthood. *Cognitive Development*, *43*, 214–223. https://doi.org/10.1016/j.cogdev.2017.05.001

Trakman, G. L., Forsyth, A., Hoye, R., & Belski, R. (2017). The nutrition for sport knowledge questionnaire (NSKQ): Development and validation using classical test theory and Rasch analysis. *Journal of the International Society of Sports Nutrition*, *14*(1), 26. https://doi.org/10.1186/s12970-017-0182-y

Valle, A., Massaro, D., Castelli, I., & Marchetti, A. (2015). Theory of Mind Development in Adolescence and Early Adulthood: The Growing Complexity of Recursive Thinking Ability. *Europe's Journal of Psychology*, *11*(1), 112–124. https://doi.org/10.5964/ejop.v11i1.829

Van Donkersgoed, R. J. M., Wunderink, L., Nieboer, R., Aleman, A., & Pijnenborg, G. H. M. (2015). Social Cognition in Individuals at Ultra-High Risk for Psychosis: A Meta-Analysis. *PLOS ONE*, *10*(10), e0141075. https://doi.org/10.1371/journal.pone.0141075

Vaskinn, A., Antonsen, B. T., Fretland, R. A., Dziobek, I., Sundet, K., & Wilberg, T. (2015). Theory of mind in women with borderline personality disorder or schizophrenia: Differences in overall ability and error patterns. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01239

Vegni, N., D'Ardia, C., & Torregiani, G. (2021). Empathy, Mentalization, and Theory of Mind in Borderline Personality Disorder: Possible Overlap With Autism Spectrum Disorders. *Frontiers in Psychology*, *12*, 626353. https://doi.org/10.3389/fpsyg.2021.626353

Vellante, M., Baron-Cohen, S., Melis, M., Marrone, M., Petretto, D. R., Masala, C., & Preti, A. (2013). The "Reading the Mind in the Eyes" test: Systematic review of psychometric properties and a validation study in Italy. *Cognitive Neuropsychiatry*, *18*(4), 326–354. https://doi.org/10.1080/13546805.2012.721728

Velloso, R. D. L., Duarte, C. P., & Schwartzman, J. S. (2013). Evaluation of the theory of mind in autism spectrum disorders with the Strange Stories test. *Arquivos de Neuro-Psiquiatria*, *71*(11), 871–876. https://doi.org/10.1590/0004-282X20130171

Verhavert, S., De Maeyer, S., Donche, V., & Coertjens, L. (2018). Scale Separation Reliability: What Does It Mean in the Context of Comparative Judgment?

Applied Psychological Measurement, *42*(6), 428–445. https://doi.org/10.1177/0146621617748321

Vetter, N. C., Leipold, K., Kliegel, M., Phillips, L. H., & Altgassen, M. (2013). Ongoing development of social cognition in adolescence. *Child Neuropsychology*, *19*(6), 615–629. https://doi.org/10.1080/09297049.2012.718324

Warnell, K. R., & Redcay, E. (2019). Minimal coherence among varied theory of mind measures in childhood and adulthood. *Cognition*, *191*, 103997. https://doi.org/10.1016/j.cognition.2019.06.009

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-Analysis of Theory-of-Mind Development: The Truth about False Belief. *Child Development*, *72*(3), 655–684. https://doi.org/10.1111/1467-8624.00304

Wellman, H. M., & Liu, D. (2004). Scaling of Theory-of-Mind Tasks. *Child Development*, *75*(2), 523–541. https://doi.org/10.1111/j.1467-8624.2004.00691.x

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103–128. https://doi.org/10.1016/0010-0277(83)90004-5

Winsper, C., Marwaha, S., Lereya, S. T., Thompson, A., Eyden, J., & Singh, S. P. (2015). Clinical and psychosocial outcomes of borderline personality disorder in childhood and adolescence: A systematic review. *Psychological Medicine*, *45*(11), 2237–2251. https://doi.org/10.1017/S0033291715000318

Wolf, M., Limberger, M. F., Kleindienst, N., Stieglitz, R.-D., Domsalla, M., Philipsen, A., Steil, R., & Bohus, M. (2009). Kurzversion der Borderline-Symptom-Liste (BSL-23): Entwicklung und Überprüfung der psychometrischen Eigenschaften. *PPmP*

- Psychotherapie · Psychosomatik · Medizinische Psychologie, *59*(08), 321–324. https://doi.org/10.1055/s-0028-1104598

Wolkenstein, L., Schönenberg, M., Schirm, E., & Hautzinger, M. (2011). I can see what you feel, but I can't deal with it: Impaired theory of mind in depression. *Journal of Affective Disorders*, *132*(1), 104–111. https://doi.org/10.1016/j.jad.2011.02.010

World-Medical-Association. (2013). World medical association declaration of Helsinki: Ethical principles for medical research involving human subjects. *JAMA*, *310*(20), 2191–2194. https://doi.org/10.1001/jama.2013.281053

Yeh, Z.-T., Lin, Y.-C., Liu, S.-I., & Fang, C.-K. (2017). Social Awareness and its Relationship with Emotion Recognition and Theory of Mind in Patients with Borderline Personality Disorder. *Journal of Social and Clinical Psychology*, *36*(1), 22–40. https://doi.org/10.1521/jscp.2017.36.1.22

Yeung, E. K. L., Apperly, I. A., & Devine, R. T. (2024). Measures of individual differences in adult theory of mind: A systematic review. *Neuroscience & Biobehavioral Reviews*, *157*, 105481. https://doi.org/10.1016/j.neubiorev.2023.105481

Zabihzadeh, A., Maleki, G., Hatami, A., Richman, M., Alimardani, Z., & Heidari, M. (2017). Affective and cognitive theory of mind in borderline personality disorder: The role of comorbid depression. *Psychiatry Research*, *257*. https://doi.org/10.1016/j.psychres.2017.07.034

Zerwas, S., Balaraman, G., & Brownell, C. (2004). Constructing an understanding of mind with peers. *Behavioral and Brain Sciences*, *27*(01). https://doi.org/10.1017/S0140525X04590039

Zoladz, P. R., & Diamond, D. M. (2013). Current status on behavioral and biological

markers of PTSD: A search for clarity in a conflicting literature. *Neuroscience &*

Biobehavioral *Reviews*, *37*(5), 860–895.

https://doi.org/10.1016/j.neubiorev.2013.03.024

List of Illustrations

Appendix

Faux-pas Recognition Test (Version für Erwachsene)

von Valerie Stone & Simon Baron-Cohen (übersetzt von Anika Ströbele)

überarbeitet von Juliane Burghardt, Magdalena Knopp, Silvia Gradl, Claudia Oppenauer, Manuel Sprung

Die Geschichten bleiben bei allen Fragen eingeblendet.

Lesen Sie sich die einzelnen Geschichten bitte sorgfältig durch und beantworten Sie anschließend die Fragen dazu. Es gibt immer 4 Antwortalternativen. Es kann vorkommen, dass Sie keine Antwort vollkommen passend finden, wählen Sie dann bitte die Antwort, die Ihrer Meinung am nächsten kommt.

1. Sandra war auf einer Party bei ihrem Freund Oliver zuhause. Sie redete mit Oliver, als eine andere Frau auf die beiden zukam. Sie war eine Nachbarin von Oliver. Die Frau sagte, „Hallo", dann wandte sie sich zu Sandra und sagte: „Ich glaube, wir kennen uns noch nicht. Ich heiße Maria und Du?". „Ich bin Sandra." „Möchte von euch jemand etwas trinken?", fragte Oliver.

1.1. Sagte irgendjemand etwas, das er oder sie nicht hätte sagen sollen oder irgendetwas Merkwürdiges?

1. Ja (1)
2. Nein (2)

[Wenn die Antwort auf Frage 1.1 „Nein" ist, dann wird weiter mit Frage 1.2.]

1.1.1. Wenn ja:

Wer sagte etwas, das er oder sie nicht hätte sagen sollen oder etwas Merkwürdiges?

1. Sarah (2)
2. Maria (4)
3. Oliver (1)
4. Sandra (3)

1.1.2. Warum hätte er oder sie dies nicht sagen sollen oder warum war es merkwürdig?

Hy_n: Oliver hat gleich gefragt ob die Frauen etwas trinken wollen. (2)

Ko: Oliver stellte Sandra nicht als seine Freundin vor. (1)

NT: Niemand stellte sich mit Nachnamen vor. (4)

Hy_p: Oliver sprach nur sehr kurz mit den beiden Frauen. (3)

1.1.3. Warum, denken Sie, hat er oder sie das gesagt?

NT: Aus Unachtsamkeit. (4)

Hy_p: Weil Oliver die Nachbarin hübscher findet als seine Freundin. (3)

Ko: Er dachte ein Gastgeber sollte seine Gäste bewirten. (1)

Hy_n: Weil er die Frauen betrunken machen will. (2)

1.1.4. Wusste Oliver, dass Sandra und Maria sich nicht kannten?

1. Ja (1)
2. Nein (2)

1.1.5. Wie, denken Sie, hat sich Sandra gefühlt?

Hy_n: besorgt (2)

hy_p: eifersüchtig (3)

NT: wach (4)

Ko: interessiert (1)

1.2. Kontrollfragen:

1.2.1. Wo war Sandra in dieser Geschichte?

1. Unterwegs (2)
2. Bei sich zu Hause (3)
3. Auf einer Party (1)
4. In einer Bar (4)

1.2.2. Kannten sich Sandra und Maria?

3. Ja (1)
4. Nein (2)

2. Der Ehemann von Claudia organisierte anlässlich ihres Geburtstages eine Überraschungsparty für sie. Er lud Sarah, eine von Claudias Freundinnen ein und sagte zu ihr: „erzähle niemandem davon, insbesondere nicht Claudia." Einen Tag vor der Party besuchte Claudia Sarah zuhause und verschüttete etwas Kaffee auf einem neuen Kleid, das über ihrem Stuhl hing. „Oh", sagte Sarah, „Ich hatte vor, das auf deiner Party zu tragen!" „Was für eine Party?", fragte Claudia. „Auf geht's" sagte Sarah, „mal sehen, ob wir den Fleck wieder herauskriegen."

2.1. Sagte irgendjemand etwas, das er oder sie nicht hätte sagen sollen oder irgendetwas Merkwürdiges?

1. Ja (1)
2. Nein (2)

[Wenn die Antwort auf Frage 2.1 „Nein" ist, dann wird weiter mit Frage 2.2.]

2.1.1. Wenn ja:

Wer sagte etwas, das er oder sie nicht hätte sagen sollen oder etwas Merkwürdiges?

1. Claudias Ehemann (2)
2. Lisa (3)
3. Sarah (1)
4. Claudia (4)

	Inwiefern hat sich Sarah in der Geschichte unpassend verhalten?					
		Trifft gar nicht zu	trifft nicht zu	trifft zu	trifft voll und ganz zu	Weiß ich nicht
1	Sarah hätte die Überraschungsparty nicht verraten sollen.	0	1	2	3	
2	Sarah hätte ihre Freundin nicht anlügen dürfen.					
3	Kleider sollten nicht im Wohnzimmer hängen.					
4	Durch das Verraten kann Claudia die Party nicht mehr genießen.					
5	Geburtstagsfeiern sind anstrengend.					

	Bitte bewerten Sie, warum Sarah so gehandelt hat.				
		Trifft gar nicht zu	trifft nicht zu	trifft zu	trifft voll und ganz zu
1	Sie hat aus Versehen die Party verraten. Sie war abgelenkt.	0	1	2	3
2	Sarah wollte aus Eifersucht die Überraschung verderben.				
3	Sarah weiß, dass Claudia keine Überraschungspartys mag und wollte es deshalb ihrer Freundin sagen.				
4	Sarah wollte, dass kein Fleck in ihrem Kleid bleibt.				

	Bitte bewerten Sie, wie sich Claudia vermutlich fühlte.				
		Trifft gar nicht zu	trifft nicht zu	trifft zu	trifft voll und ganz zu
1	verletzt	0	1	2	3
2	entspannt				
3	erleichtert				
4	überrascht				
5	alt				

2.2. Kontrollfragen:

2.2.1. Für wen war die Überraschungsparty in dieser Geschichte?

1. Sarah (2)
2. Claudia (1)
3. Den Ehemann (3)
4. Susanne (4)

2.2.2. Was wurde auf dem Kleid verschüttet?

1. Milch (2)
2. Tee (3)
3. Wasser (4)
4. Kaffee (1)

3. Tom war unterwegs, um ein Hemd zu kaufen, das zu seinem Anzug passte. Der Verkäufer zeigte ihm mehrere Hemden. Tom schaute sie sich an und fand schließlich eines in der passenden Farbe. Aber als er das Hemd in der Umkleidekabine anprobierte, passte es nicht. „Ich fürchte, es ist zu klein" sagte er zu dem Verkäufer. „Keine Sorge" sagte der Verkäufer, „nächste Woche bekommen wir größere Größen" „Klasse, dann werde ich da wiederkommen" sagte Tom.

3.1. Sagte irgendjemand etwas, das er oder sie nicht hätte sagen sollen oder irgendetwas Merkwürdiges?

1. Ja (1)
2. Nein (2)

[Wenn die Antwort auf Frage 3.1 „Nein" ist, dann wird weiter mit Frage 3.2.]

3.1.1. Wenn ja:

Wer sagte etwas, das er oder sie nicht hätte sagen sollen oder etwas Merkwürdiges?

1. Tom (1)
2. Susi (2)
3. Der Verkäufer (3)
4. Tim (4)

3.1.2. Warum hätte er oder sie dies nicht sagen sollen oder warum war es merkwürdig?

Hy_n: Es ist nicht nett, gleich zu sagen, dass die Person größere Größen braucht. (2)

NT: Größenangaben in Geschäften sind heute sehr unzuverlässig. (4)

Hy_p: Der Verkäufer war enttäuscht, weil Tom nach der Beratung nichts kauft. (3)

Ko: Es ist nicht klar, ob Tom wirklich wiederkommt. (1)

3.1.3. Warum, denken Sie, hat er oder sie das gesagt?

Hy_n: Er wollte Tom schnell loswerden. (2)

Ko: Der Verkäufer möchte das der Kunde wiederkommt. (1)

NT: Weil das Hemd zu klein war. (4)

Hy_p: Er wollte, dass Tom gut aussieht. (3)

3.1.4. Wusste Tom, als er das Hemd anprobierte, dass sie es nicht mehr in seiner Größe haben?

1. Ja (1)
2. Nein (2)

3.1.5. Wie, denken Sie fühlte sich Tom?

Hy_n: gekränkt (2)

Ko: enttäuscht (1)

NT: gleichgültig (4)

Hy_p: geschmeichelt (3)

3.2. **Kontrollfragen:**

3.2.1. Was wollte Tom in dieser Geschichte kaufen?

1. Pullover (2)
2. T-Shirt (3)
3. Hemd (1)
4. Jacke (4)

3.2.2. Warum hat er vor, nächste Woche wiederzukommen?

1. Um eine Bestellung abzuholen (2)
2. Um einen anderen Pullover anzuprobieren (3)
3. Um einen Anzug zu kaufen (4)
4. Um das Hemd in einer größeren Nummer zu probieren (1)

4. Julia war eben erst in eine neue Wohnung gezogen. Sie ging einkaufen und kaufte neue Vorhänge für ihr Schlafzimmer. Als sie gerade mit der Dekoration der Wohnung fertig geworden war, kam ihre beste Freundin Lisa zu Besuch. Julia führte sie durch die Wohnung und fragte, „Wie gefällt dir mein Schlafzimmer?" „Diese Vorhänge sind schrecklich" sagte Lisa. „Ich hoffe, du wirst dir Neue besorgen!"

4.1. Sagte irgendjemand etwas, das er oder sie nicht hätte sagen sollen oder irgendetwas Merkwürdiges?

1. Ja (1)
2. Nein (2)

[Wenn die Antwort auf Frage 4.1 „Nein" ist, dann wird weiter mit Frage 4.2.]

4.1.1. Wenn ja:

Wer sagte etwas, das er oder sie nicht hätte sagen sollen oder etwas Merkwürdiges?

1. Luisa (2)
2. Julia (3)
3. Sandra (4)
4. Lisa (1)

4.1.2. Warum hätte er oder sie dies nicht sagen sollen oder warum war es merkwürdig?

Ko: Lisa hat Julia gekränkt, weil die Vorhänge neu waren und Julia gefallen. (1)

Hy_p: Julia findet es toll, wie ehrlich Lisa ist. (3)

NT: Es ist besser für die Umwelt nicht ständig neue Dinge zu kaufen. (4)

Hy_n: Lisa wollte Julia beleidigen. (2)

4.1.3. Warum, denken Sie, hat er oder sie das gesagt?

NT: In einer Freundschaft ist Ehrlichkeit wichtig. (4)

Hy_p: Weil sich Julia und Lisa so gut kennen, dass sie ihre Meinung nicht zurückhalten muss. (3)

Hy_n: Weil sie neidisch auf Julias neue Wohnung ist. (2)

Ko: Weil ihr die Vorhänge nicht gefallen haben. (1)

4.1.4. Wusste Lisa, wer die Vorhänge gekauft hatte?

1. Ja (1)
2. Nein (2)

4.1.5. Wie, denken Sie fühlte sich Julia?

NT: gleichgültig (4)

Ko: irritiert (1)

Hy_p: zufrieden (3)

Hy_n: persönlich angegriffen (2)

4.2. Kontrollfragen:

4.2.1. Was hatte Julia eben erst gekauft in der Geschichte?

1. Möbel fürs Schlafzimmer (2)
2. Vorhänge (1)
3. Pflanzen (3)
4. Bettwäsche (4)

4.2.2. Wie lange hatte Julia schon in dieser Wohnung gewohnt?

1. Das steht nicht im Text. (2)
2. Über mehrere Jahre (3)
3. Sie war gestern eingezogen. (4)
4. Erst seit Kurzem (1)

5. Ralf ging zum Friseur, um sich die Haare schneiden zu lassen. „Welche Frisur möchten sie denn haben?" fragte der Friseur. „Ich möchte gerne denselben Haarschnitt, den ich jetzt habe, nur zwei Zentimeter kürzer" antwortete Ralf. Der Friseur schnitt die Haare vorne etwas ungerade ab und um das wieder auszubessern, musste er die Haare kürzer schneiden. „Ich fürchte, es ist ein bisschen kürzer geworden, als sie wollten" sagte der Friseur. „Naja" sagte Ralf, „das wächst wieder nach."

5.1. Sagte irgendjemand etwas, das er oder sie nicht hätte sagen sollen oder irgendetwas Merkwürdiges?

1. Ja (1)
2. Nein (2)

[Wenn die Antwort auf Frage 5.1 „Nein" ist, dann wird weiter mit Frage 5.2.]

5.1.1. Wenn ja:

Wer sagte etwas, das er oder sie nicht hätte sagen sollen oder etwas Merkwürdiges?

1. Lorenz (2)
2. Ralf (1)
3. Der Friseur (3)
4. Bianca (4)

5.1.2. Warum hätte er oder sie dies nicht sagen sollen oder warum war es merkwürdig?

Hy_p: Der Friseur fand, dass Ralf mit kürzeren Haaren besser aussieht. (2)

NT: Der Friseur hätte einen Preisnachlass anbieten sollen. (4)

Hy_n: Der Friseur stört sich gar nicht daran, dass er die Haare schlecht geschnitten hat. (3)

Ko: Der Friseur hätte dem Kunden sofort von seinem Fehler erzählen sollen. (1)

5.1.3. Warum, denken Sie, hat er oder sie das gesagt?

Hy_n: Um sich nicht offen zu ärgern und verletzt zu zeigen. (2)

Ko: Ralf hätte den Fehler früher oder später bemerkt. (1)

NT: Friseure reden immer viel. (4)

Hy_p: Weil er Ralf attraktiver aussehen lassen wollte. (3)

5.1.4. Wusste Ralf, während der Friseur ihm die Haare schnitt, dass dieser die Haare gerade zu kurz schneidet?

1. Ja (1)
2. Nein (2)

5.1.5. Wie, denken Sie fühlte sich Ralf?

Hy_n: verärgert (2)

Hy_p: erfreut (3)

Ko: gleichgültig (1)

NT: hellwach (4)

5.2. **Kontrollfragen:**

5.2.1. Wie wollte Ralf seinen Haarschnitt in der Geschichte haben?

1. An den Seiten kurz und hinten lang (4)
2. Denselben Haarschnitt, nur zwei Zentimeter kürzer (1)
3. Ganz kurz (2)
4. Zur Seite geföhnt (3)

5.2.2. Wie schnitt ihm der Friseur die Haare?

1. Kürzer als er wollte. (1)
2. Zu langsam (2)
3. Genauso wie er es wollte. (3)
4. Länger als er wollte. (4)

6. Heiko hielt an der Tankstelle auf dem Weg nach Hause an, um sein Auto zu tanken. Er gab der Kassiererin seine Kreditkarte. Die Kassiererin zog die Karte durch das Gerät an der Kasse. „Es tut mir leid" sagte sie, „das Gerät akzeptiert Ihre Karte nicht." „Hmmm, das ist komisch!" sagte Heiko. „Gut, dann werde ich bar bezahlen." Er gab ihr zwanzig Euro und sagte, „und, ich hätte gerne noch einen Schokoriegel."

6.1. Sagte irgendjemand etwas, das er oder sie nicht hätte sagen sollen oder irgendetwas Merkwürdiges?

1. Ja (1)
2. Nein (2)

[Wenn die Antwort auf Frage 6.1 „Nein" ist, dann wird weiter mit Frage 6.2.]

6.1.1. Wenn ja, frage:

Wer sagte etwas, das er oder sie nicht hätte sagen sollen oder etwas Merkwürdiges?

1. Heiko (1)
2. Die Kassiererin (2)
3. Tanja (3)
4. Sebastian (4)

6.1.2. Warum hätte er oder sie dies nicht sagen sollen oder warum war es merkwürdig?

Hy_p: Er fand die Kassiererin attraktiv. (3)

Hy_n: Seine Reaktion war zu entspannt und unbesorgt. (2)

NT: Schokoriegel sind nicht gesund. (4)

Ko: Es ist unpraktisch etwas nachträglich zu einem Einkauf hinzuzufügen. (1)

6.1.3. Warum, denken Sie, hat er oder sie das gesagt?

Hy_p: Er wollte mit der Kassiererin flirten. (3)

Ko: Weil ihm erst später einfiel, dass er einen Schokoriegel wollte. (1)

Hy_n: Er vermutete schon, dass die Kreditkarte nicht akzeptiert werden würde und war nervös. (2)

NT: Schokoriegel schmecken gut. (4)

6.1.4. Wusste Heiko, dass das Gerät seine Kreditkarte nicht akzeptieren würde, als er der Kassiererin seine Karte gab?

1. Ja (1)
2. Nein (2)

6.1.5. Wie, denken Sie fühlte sich Heiko?

Ko: gelassen (1)

Hy_p: verliebt (3)

NT: hungrig (4)

Hy_n: verlegen (2)

6.2. **Kontrollfragen:**

6.2.1. Für was hielt Heiko in der Geschichte an?

1. Um ein Getränk zu kaufen (2)
2. Um mit der Kassiererin zu plaudern (3)
3. Um sein Auto zu waschen (4)
4. Um zu tanken (1)

6.2.2. Warum bezahlte er bar?

1. Weil er seine Karte vergessen hatte. (2)
2. Weil es ihm unangenehm war mit Karte zu bezahlen. (3)
3. Weil das Gerät seine Karte nicht akzeptierte. (1)
4. Weil er das Bargeld passend hatte. (4)

7. Lena ist ein dreijähriges Mädchen mit einem runden Gesicht und kurzen blonden Haaren. Sie war zu Besuch bei ihrer Tante Sofie. Es klingelte an der Tür und ihre Tante Sofie öffnete. Marianne, eine Nachbarin schaute vorbei. „Hallo" sagte Tante Sofie, „Schön, dass Sie vorbeikommen." Marianne sagte, „Hallo", dann schaute sie zu Lena und sagte, „Oh, ich glaube, ich kenne diesen kleinen Jungen noch nicht. Wie heißt du denn?"

7.1. Sagte irgendjemand etwas, das er oder sie nicht hätte sagen sollen oder irgendetwas Merkwürdiges?

1. Ja (1)
2. Nein (2)

[Wenn die Antwort auf Frage 7.1 „Nein" ist, dann wird weiter mit Frage 7.2.]

7.1.1. Wenn ja:

Wer sagte etwas, das er oder sie nicht hätte sagen sollen oder etwas Merkwürdiges?

1. Luisa (2)
2. Lena (3)
3. Tante Sofie (4)
4. Marianne, die Nachbarin (1)

7.1.2. Warum hätte er oder sie dies nicht sagen sollen oder warum war es merkwürdig?

Hy_n: Marianne wollte Lena ärgern. (2)

NT: Die Nachbarin hätte Lena nicht „klein" nennen sollen. Kinder wollen sich erwachsen fühlen. (4)

Ko: Marianne hat Lena fälschlicherweise für einen Jungen gehalten, obwohl sie ein Mädchen ist. (1)

Hy_p: Marianne schenkt dem Kind sehr viel Aufmerksamkeit. (3)

7.1.3. Warum, denken Sie, hat er oder sie das gesagt?

Hy_n: Weil sie Lena kränken wollte. (2)

NT: Weil das Kind noch sehr jung ist. (4)

Hy_p: Weil sie sich einen Spaß mit Lena erlauben wollte. (3)

Ko: Marianne glaubte wegen Lenas kurzen Haare, dass sie ein Junge sei. (1)

7.1.4. Wusste Marianne, dass Lena ein Mädchen ist?

1. Ja (1)
2. Nein (2)

7.1.5. Wie, denken Sie fühlte sich Lena?

Ko: verwirrt (1)

Hy_n: sie schämt sich (2)

NT: hungrig (4)

Hy_p. geschmeichelt (3)

7.2. **Kontrollfragen:**

7.2.1. Wo war Lena in der Geschichte?

1. Zu Besuch bei ihrer Tante Marianne (2)
2. Zu Besuch bei ihrer Tante Sofie (1)
3. Zu Besuch bei dem kleinen Mädchen (3)
4. Zu Besuch bei dem kleinen Jungen (4)

7.2.2. Wer kam zu Besuch?

1. Marianne, die Nachbarin (1)
2. Lena (2)
3. Marleen (3)
4. Sofie (4)

8. Steffi ging mit ihrem Hund, Charly, hinaus in den Park. Sie warf ihm einen Stock zu, dem er hinterherjagte. Als sie dort einige Zeit waren, begegnete ihnen Bianca, Steffis Nachbarin. Steffi und Bianca redeten einige Minuten. Dann fragte Bianca „Gehst du auch in Richtung Wohnung? Möchtest du mit mir zusammen gehen?" „Sicher" sagte Steffi. Sie rief Charly, aber dieser war damit beschäftigt, Tauben zu jagen und kam nicht. „Es sieht so aus, als wäre er noch nicht bereit zu gehen" sagte sie. „Ich denke, wir bleiben noch." „Ok", sagte Bianca. „Ich sehe dich dann später."

8.1. Sagte irgendjemand etwas, das er oder sie nicht hätte sagen sollen oder irgendetwas Merkwürdiges?

1. Ja (1)
2. Nein (2)

[Wenn die Antwort auf Frage 8.1 „Nein" ist, dann wird weiter mit Frage 8.2.]

8.1.1. Wenn ja:

Wer sagte etwas, das er oder sie nicht hätte sagen sollen oder etwas Merkwürdiges?

1. Marc (3)
2. Steffi (2)
3. Bianca (1)
4. Charly (4)

8.1.2. Warum hätte er oder sie dies nicht sagen sollen oder warum war es merkwürdig?

NT: Sie führt den Hund nicht an der Leine. (4)

Hy_p: Sie findet es wichtiger, dass der Hund glücklich ist als die Nachbarin. (3)

Ko: Sie widerspricht sich. (1)

Hy_n: Sie sucht eine Ausrede, um nicht mehr mit der Nachbarin reden zu müssen. (2)

8.1.3. Warum, denken Sie, hat er oder sie das gesagt?

Hy_n: Sie mag die Nachbarin nicht. (2)

Hy_p: Sie mag ihren Hund. (3)

NT: Weil sie ihren Hund nicht unter Kontrolle hat. (4)

Ko: Es ist ihr egal was der Hund macht. (1)

8.1.4. Als Bianca Steffi fragte, ob sie mit ihr zusammen nach Hause gehen möchte, wusste

Bianca, dass Steffi nicht mitkommen konnte?

1. Ja (1)
2. Nein (2)

8.1.5. Wie, glauben Sie fühlte sich Bianca?

Ko: genervt (1)

Hy_n: es ist ihr peinlich (2)

Hy_p: gerührt (3)

NT: gleichgültig (4)

8.2. **Kontrollfragen:**

8.2.1. Wohin ging Steffi mit Charly in der Geschichte?

Zum Einkaufen (2)

In den Park (1)

In die Garage (3)

Zu Bianca (4)

8.2.2. Warum ging sie nicht zusammen mit ihrer Nachbarin Bianca nach Hause?

Weil Steffi noch Einkaufen gehen wollte. (3)

Weil der Hund Bianca nicht mochte. (4)

Weil der Hund noch spielen wollte. (1)

Weil Steffi noch nicht nach Hause wollte. (2)

9. Heidi hatte letztes Jahr die Hauptrolle in einer Theateraufführung ihrer Schule gespielt und wollte auch dieses Jahr wieder eine Hauptrolle haben. Sie nahm Schauspielunterricht und im Frühjahr hatte sie ein Vorsprechen für das Theaterschauspiel. Am Tag, als die Entscheidung bekannt gegeben wurde, ging sie vor dem Unterricht zu der Liste, um zu sehen, wer das Rennen gemacht hatte. Sie hatte die Hauptrolle nicht bekommen, sondern war stattdessen für eine Nebenrolle vorgesehen. Auf dem Gang traf sie zufällig ihren Freund und erzählte ihm, was passiert war. „Tut mir leid", sagte er „du bist bestimmt enttäuscht". „Ja", sagte Heidi, „Ich muss mich entscheiden, ob ich diese Rolle annehme."

9.1. Sagte irgendjemand etwas, das er oder sie nicht hätte sagen sollen oder irgendetwas Merkwürdiges?

1. Ja (1)
2. Nein (2)

[Wenn die Antwort auf Frage 9.1 „Nein" ist, dann wird weiter mit Frage 9.2.]

9.1.1. Wenn ja:

Wer sagte etwas, das er oder sie nicht hätte sagen sollen oder etwas Merkwürdiges?

1. Der Freund (2)
2. Der Schauspieler (3)
3. Heidi (1)
4. Hannah (4)

9.1.2. Warum hätte er oder sie dies nicht sagen sollen oder warum war es merkwürdig?

NT: Es gibt wichtigeres als Theater. (4)

Hy_p: Ihr Freund hat sich wirklich angestrengt sie wieder aufzuheitern. (3)

Ko: Heidi hatte extra für die Rolle geübt und sie trotzdem nicht bekommen. (1)

Hy_n: Heidi war sich sicher die Hauptrolle zu bekommen. Für die Nebenrolle wollte sie nicht auftreten. (2)

9.1.3. Warum, denken Sie, hat er oder sie das gesagt?

NT: Theater ist leichter als Mathematik. (4)

Hy_n: Weil die Nebenrolle unter ihrem Niveau war. (2)

Ko: Um sie zu trösten (1)

Hy_p: Weil er will, dass sie glücklich ist. (3)

9.1.4. Wusste Heidis Freund, als er sie zufällig auf dem Gang traf, dass sie die Rolle nicht bekommen hatte?

1. Ja (1)
2. Nein (2)

9.1.5. Wie, denken Sie fühlte sich Heidi?

NT: gelangweilt (4)

Hy_n: empört (2)

Ko: enttäuscht (1)

Hy_p: aufgeheitert (3)

9.2. **Kontrollfragen:**

9.2.1. Welche Rolle bekam Heidi in dieser Geschichte?

1. Die Hauptrolle in einer Theateraufführung (1)
2. Eine Nebenrolle in einer Theateraufführung (2)
3. Sie was zuständig für den Ton der Theateraufführung. (3)
4. Sie bekam keine Rolle in der Theateraufführung. (4)

9.2.2. Welche Rolle hatte sie letztes Jahr gehabt?

1. Sie hatte die Hauptrolle in der Theateraufführung. (2)
2. Sie hatte eine Nebenrolle in der Theateraufführung. (1)
3. Sie was zuständig für den Ton der Theateraufführung. (3)
4. Sie hatte keine Rolle in der Theateraufführung. (4)

10. Thomas war in der Bücherei. Er fand ein Buch über Wandern in den Alpen, das er haben wollte und ging zur Ausleihtheke, um das Buch zu entleihen. Als er seinen Geldbeutel aufmachte, entdeckte er, dass er seine Büchereikarte zu Hause gelassen hatte. „Es tut mir leid", sagte er zu der Frau an der Ausleihtheke. „Ich glaube, ich habe meine Karte zu Hause gelassen." „Das ist ok", antwortete sie. „Nennen Sie mir einfach Ihren Namen. Wenn wir Sie im Computer haben, können Sie das Buch auch ausleihen, in dem Sie mir Ihren Führerschein zeigen."

10.1. Sagte irgendjemand etwas, das er oder sie nicht hätte sagen sollen oder irgendetwas Merkwürdiges?

1. Ja (1)
2. Nein (2)

[Wenn die Antwort auf Frage 10.1 „Nein" ist, dann wird weiter mit Frage 10.2.]

10.1.1. Wenn ja:

Wer sagte etwas, das er oder sie nicht hätte sagen sollen oder etwas Merkwürdiges?

1. Die Mitarbeiterin der Bücherei (1)
2. Thomas (2)
3. Sarah (3)
4. Martin (4)

10.1.2. Warum hätte er oder sie dies nicht sagen sollen oder warum war es merkwürdig?

Ko: Man sollte prüfen, ob man alles dabeihat, bevor man seine Wohnung verlässt. (1)

Hy_n: Er nahm sofort das Schlimmste an. (2)

Hy_p: Sie hat für ihn eine Ausnahme gemacht. (3)

NT: Wer geht heute noch in eine Bibliothek? (4)

10.1.3. Warum, denken Sie, hat er oder sie das gesagt?

Hy_p: Um sich mit ihm unterhalten zu können. (3)

Ko: Sie versucht die Kunden zufriedenzustellen. (1)

NT: Thomas ist sehr altmodisch. (4)

Hy_n: Sie wollte ihn schikanieren. (2)

10.1.4. Als Thomas in die Bücherei ging, war ihm da bewusst, dass er seine Büchereikarte nicht dabeihatte?

1. Ja (1)
2. Nein (2)

10.1.5. Wie denken Sie hat sich Thomas gefühlt?

Hy_n: verärgert (2)

Ko: erleichtert (1)

Hy_p: geschmeichelt (3)

NT: entspannt (4)

10.2. **Kontrollfragen:**

10.2.1. Welches Buch hat Thomas in der Geschichte in der Bücherei ausgeliehen?

1. Wandern in den Alpen (1)
2. Skifahren in den Bergen (2)
3. Tauchen im Mittelmeer (3)
4. Gemüseanbau im Garten (4)

10.2.2. Konnte er das Buch ausleihen?

1. Ja (2)
2. Nein (3)
3. Weiß man noch nicht (1)

11. Simone Groß, eine Managerin von Abco Software Design hatte alle Mitarbeiter zu einer Versammlung einberufen. „Ich habe Ihnen etwas mitzuteilen", sagte sie. „Tim Becker, einer unserer Buchhalter, ist schwer an Krebs erkrankt und liegt im Krankenhaus." Alle waren still, um die Nachricht zu verarbeiten, als Robert, ein Softwareentwickler, verspätet dazukam. „Hey, ich habe gestern Abend einen guten Witz gehört!" sagte Robert. „Was sagt ein todkranker Patient zu seinem Arzt?" Daraufhin sagte Simone „Okay, kommen wir wieder zurück zum Thema unserer Versammlung!"

11.1. Sagte irgendjemand etwas, das er oder sie nicht hätte sagen sollen oder irgendetwas Merkwürdiges?

1. Ja (1)
2. Nein (2)

[Wenn die Antwort auf Frage 11.1 „Nein" ist, dann wird weiter mit Frage 11.2.]

11.1.1. Wenn ja:

Wer sagte etwas, das er oder sie nicht hätte sagen sollen oder etwas Merkwürdiges?

1. Robert (1)
2. Simone, die Managerin (2)
3. Der Arzt (3)
4. Tim Becker (4)

11.1.2. Warum hätte er oder sie dies nicht sagen sollen oder warum war es

merkwürdig?

Hy_p: Robert heitert gerne Leute auf und macht deshalb andauernd Witze. (3)

Ko: Es war nicht der richtige Moment für diese Art von Witz. (1)

NT: Robert kam zu spät. (4)

Hy_n: Robert interessiert sich nicht für das Schicksal seines Kollegen. (2)

11.1.3. Warum, denken Sie, hat er oder sie das gesagt?

NT: Humor ist sympathisch. (4)

Hy_p: Weil er die Anspannung in den Gesichtern sah und die Stimmung auflockern wollte. (3)

Ko: Weil er später dazukam und die Nachricht nicht mitbekommen hat. (1)

Hy_n: Weil er seinen Kollegen nie gemocht hat. (2)

11.1.4. Als Robert zur Versammlung dazu kam, wusste er, dass der Buchhalter an Krebs erkrankt war?

1. Ja (1)
2. Nein (2)

11.1.5. Wie, denken Sie fühlte sich Simone, die Managerin?

Hy_n: verärgert (2)

Ko: irritiert (1)

NT: konzentriert (4)

Hy_p: amüsiert (3)

11.2. Kontrollfragen:

11.2.1. Was erzählte Simone, die Managerin in der Geschichte, den Leuten bei der Versammlung?

1. Dass eine Mitarbeiterin Mutter geworden ist. (2)
2. Dass ein Mitarbeiter gekündigt hat. (3)
3. Dass eine Mitarbeiterin immer zu spät kommt. (4)
4. Dass ein Mitarbeiter an Krebs erkrankt ist. (1)

11.2.2. Wer kam zu spät zur Versammlung?

1. Robert (1)
2. Simone (2)
3. Tim (3)

4. Der Arzt (4)

12. Michael, ein neunjähriger Junge, ging seit kurzem auf eine neue Schule. Er war in einer der Toilettenkabinen der Schule. Jonas und Peter, zwei andere Jungs, kamen zur Toilette herein, stellten sich an die Waschbecken und unterhielten sich. Jonas sagte, „Kennst du diesen neuen Typen aus unserer Klasse? Sein Name ist Michael. Sieht der nicht komisch aus? Und er ist so klein!" Michael kam aus der Kabine und Jonas und Peter schauten zu ihm. Peter sagte, „Oh, hallo Michael! Gehst du jetzt raus zum Fußball spielen?"

12.1. Sagte irgendjemand etwas, das er oder sie nicht hätte sagen sollen oder irgendetwas Merkwürdiges?

1. Ja (1)
2. Nein (2)

[Wenn die Antwort auf Frage 12.1 „Nein" ist, dann wird weiter mit Frage 12.2.]

12.1.1. Wenn ja:

Wer sagte etwas, das er oder sie nicht hätte sagen sollen oder etwas Merkwürdiges?

1. Markus (2)
2. Peter (3)
3. Michael (4)
4. Jonas (1)

12.1.2. Warum hätte er oder sie dies nicht sagen sollen oder warum war es

merkwürdig?

Ko: Weil er Michael dadurch gekränkt hat. (1)

Hy_n: Weil sie dadurch verhindern, dass Michael in der Klasse anerkannt wird. (2)

NT: Man redet nicht schlecht über andere. (4)

Hy_p: Sie wollten so mit Michael in Kontakt treten. (3)

12.1.3. Warum, denken Sie, hat er oder sie das gesagt?

NT: Er wollte die Toilette benutzen. (4)

Hy_p: Weil er wollte, dass Michael sein Freund wird. (3)

Hy_n: Weil er mit sich selbst unzufrieden ist und deshalb andere beleidigen will. (2)

Ko: Er war ehrlich und sagt deshalb, dass Michael klein ist und komisch aussieht. (1)

12.1.4. Wusste Jonas, als er sich mit Peter unterhielt, dass Michael in einer der Kabinen war?

1. Ja (1)
2. Nein (2)

12.1.5. Wie denken Sie fühlte sich Michael?

Ko: verunsichert (1)

Hy_n: ausgegrenzt (2)

NT: gestört, weil der Raum so voll ist (4)

Hy_p: interessiert (3)

12.2. **Kontrollfragen:**

12.2.1. Wo war Michael, als Jonas und Peter sich unterhielten?

1. In der Toilettenkabine (1)
2. Auf dem Pausenhof (2)
3. Auf dem Flur (3)
4. Im Klassenzimmer (4)

12.2.2. Was sagte Jonas über Michael?

1. Dass Michael viel zu groß ist. (2)
2. Dass Michael dreckige Klamotten trägt. (3)
3. Dass Michael neu ist. (4)
4. Dass Michael komisch aussieht und klein ist. (1)

13. Andreas Cousin, Frank kam zu Besuch und Andrea hatte extra einen Apfelkuchen für ihn gemacht. Nach dem Abendessen sagte sie, „Ich habe einen Kuchen nur für dich gemacht. Er ist in der Küche." „Mmmm" antwortete Frank, „Er riecht köstlich! Ich liebe Kuchen – außer Apfelkuchen natürlich."

13.1. Sagte irgendjemand etwas, das er oder sie nicht hätte sagen sollen oder irgendetwas Merkwürdiges?

1. Ja (1)
2. Nein (2)

[Wenn die Antwort auf Frage 13.1 „Nein" ist, dann wird weiter mit Frage 13.2.]

13.1.1. Wenn ja:

Wer sagte etwas, das er oder sie nicht hätte sagen sollen oder etwas Merkwürdiges?

1. Nora (3)
2. Andrea (4)
3. Frank (1)
4. Manuel (2)

13.1.2. Warum hätte er oder sie dies nicht sagen sollen oder warum war es merkwürdig?

Ko: Weil Andrea extra für ihn einen Apfelkuchen gemacht hat. (1)

Hy_n: Weil er Andrea zutiefst beleidigt. (2)

Hy_p: Weil er damit die Überraschung verdirbt. (3)

NT: Man isst, was auf den Tisch kommt. (4)

13.1.3. Warum, denken Sie, hat er oder sie das gesagt?

NT: Er will eine Unterhaltung führen. (4)

Hy_n: Er wollte, sich an ihr rächen. (2)

Ko: Er weiß nicht, welchen Kuchen sie gebacken hat und sagt einfach die Wahrheit. (1)

Hy_p: Weil er schon so viel gegessen hat und nicht noch mehr essen will. (3)

13.1.4. Wusste Frank, als er den Kuchen roch, dass es ein Apfelkuchen war?

1. Ja (1)
2. Nein (2)

13.1.5. Wie denken Sie fühlte sich Andrea?

Hy_p: mitfühlend. (3)

Ko: enttäuscht. (1)

Hy_n: wertlos (2)

NT: satt (4)

13.2. **Kontrollfragen:**

13.2.1. Was für einen Kuchen hatte Andrea in der Geschichte gebacken?

1. Mohnkuchen (2)
2. Apfelkuchen (1)
3. Birnenkuchen (3)
4. Kirschkuchen (4)

13.2.2. Woher kannten sich Andrea und Frank?

1. Frank war der Cousin von Andrea. (1)
2. Andrea war die Freundin von Frank. (2)
3. Frank war der Onkel von Andrea. (3)
4. Andrea war die Tante von Frank. (4)

14. Isabel kaufte ihrer Freundin, Anne eine Kristallschale als Hochzeitsgeschenk. Die Hochzeitsfeier von Anne war sehr groß und so galt es, den Überblick über die vielen Geschenke zu behalten. Etwa ein Jahr später war Isabel bei Anne zum Abendessen eingeladen. Isabel ließ aus Versehen eine Weinflasche auf die Kristallschale fallen und die Schale zersprang. „Es tut mir sehr leid. Ich habe die Schale zerstört" sagte Isabel. „Mach dir keine Sorgen" sagte Anne. „Ich habe die Schale sowieso nie gemocht. Irgendjemand hat sie mir zur Hochzeit geschenkt."

14.1. Sagte irgendjemand etwas, das er oder sie nicht hätte sagen sollen oder irgendetwas Merkwürdiges?

1. Ja (1)
2. Nein (2)

[Wenn die Antwort auf Frage 14.1 „Nein" ist, dann wird weiter mit Frage 14.2.]

14.1.1. Wenn ja

Wer sagte etwas, das er oder sie nicht hätte sagen sollen oder etwas Merkwürdiges?

1. Julia (2)
2. Isabel (3)
3. Anne (1)
4. Sarah (4)

14.1.2. Warum hätte er oder sie dies nicht sagen sollen oder warum war es

merkwürdig?

Hy_n: Weil Anne durch diese Äußerung ihre jahrelange Freundschaft zerstört hat. (2)

Ko: Weil die Schale ein Geschenk von Isabel gewesen ist. (1)

NT: Weil man sich über Geschenke nicht beschweren sollte. (4)

Hy_p: Isabel hatte bemerkt, dass Anne die Kristallschale nicht mochte. Sie wollte deshalb nicht, dass sie sich weiterhin gezwungen fühlt, sie zu behalten. (3)

14.1.3. Warum, denken Sie, hat er oder sie das gesagt?

Ko: Sie hatte vergessen, dass sie die Schale von Isabel geschenkt bekommen hatte. (1)

NT: Sie war schon hungrig. (4)

Hy_p: Anne ist eigentlich traurig über das Missgeschick, sie wollte es sich aber nicht anmerken lassen und hat deshalb gesagt, dass sie ihr nicht gefiel. (3)

Hy_n: Weil sie von der Schale als Hochzeitsgeschenk enttäuscht war und sie es Isabel nun spüren lassen wollte. (2)

14.1.4. Erinnerte sich Anne daran, dass Isabel ihr die Schale geschenkt hatte?

1. Ja (1)
2. Nein (2)

14.1.5. Wie denken Sie fühlte sich Isabel?

Hy n: persönlich angegriffen (2)

NT: hungrig (4)

Ko: verletzt (1)

Hy p: erleichtert (3)

14.2. **Kontrollfragen:**

14.2.1. Was schenkte Isabel Anne zur Hochzeit?

1. Eine Kristallschale (1)
2. Eine Weinflasche (2)
3. Weingläser (3)
4. Einen Urlaubsgutschein (4)

14.2.2. Wie ging die Schale kaputt?

1. Isabel ließ eine Flasche auf die Schale fallen. (1)
2. Anne ließ die Schale fallen. (2)
3. Isabel warf die Schale an die Wand. (3)
4. Anne stieß gegen die Schale. (4)

15. In der Adam-Riese-Gesamtschule gab es einen Geschichten-Wettbewerb. Jeder war eingeladen, daran teilzunehmen. Etliche Fünftklässler taten dies auch. Svenja, eine Fünftklässlerin, liebte die Geschichte, mit der sie am Wettbewerb teilgenommen hatte. Einige Tage später wurden die Ergebnisse des Wettbewerbs verkündet: Svenjas Geschichte hatte überhaupt nichts gewonnen, während ein Klassenkamerad, Lukas, den ersten Preis abräumte. Am nächsten Tag saß Svenja mit Lukas auf einer Bank. Sie schauten sich seinen „Erster Platz" - Pokal an. Lukas sagte „es war so einfach, diesen Wettbewerb zu gewinnen. Alle anderen Geschichten in diesem Wettbewerb waren schrecklich." „Wo wirst du deinen Pokal hinstellen?" fragte Svenja.

15.1. Sagte irgendjemand etwas, das er oder sie nicht hätte sagen sollen oder irgendetwas Merkwürdiges?

1. Ja (1)
2. Nein (2)

[Wenn die Antwort auf Frage 15.1 „Nein" ist, dann wird weiter mit Frage 15.2.]

15.1.1. Wenn ja:

Wer sagte etwas, das er oder sie nicht hätte sagen sollen oder etwas Merkwürdiges?

1. Marcus (2)
2. Lukas (1)
3. Luisa (3)
4. Svenja (4)

15.1.2. Warum hätte er oder sie dies nicht sagen sollen oder warum war es merkwürdig?

Hy_p: Er versucht zu angestrengt Svenja zu beeindrucken. (3)

Ko: Weil er Svenja mit seiner Aussage gekränkt hat. (1)

Hy_n: Lukas denkt er sei der Beste. Durch seine überheblichen Äußerungen mindert er die Leistungen aller anderen. (2)

NT: Wettbewerbe an Schulen sind unnötig. (4)

15.1.3. Warum, denken Sie, hat er oder sie das gesagt?

NT: Er ist noch ein Kind. (4)

Hy_p: Er wollte bei Svenja Eindruck schinden und sie mit seinem Preis beeindrucken. (3)

Ko: Lukas hat in diesem Moment nicht nachgedacht. (1)

Hy_n: Er wollte Svenja verletzen. (2)

15.1.4. Wusste Lukas, dass Svenja mit einer Geschichte am Wettbewerb teilgenommen hatte?

1. Ja (1)
2. Nein (2)

15.1.5. Wie denken Sie fühlte sich Svenja?

Hy_p: erfreut (3)

Ko: verletzt (1)

Hy_n: wütend (2)

NT: gelangweilt (4)

15.2. **Kontrollfragen:**

15.2.1. Wer gewann den Wettbewerb in dieser Geschichte?

1. Lukas (1)
2. Svenja (2)
3. Marcus (3)
4. Luisa (4)

15.2.2. Gewann Svenja mit ihrer Geschichte etwas?

1. Ja (1)
2. Nein (2)

16. Markus war in einem Restaurant. Er verschüttete versehentlich etwas Kaffee auf dem Tisch. „Ich werde Ihnen eine neue Tasse Kaffee bringen", sagte die Bedienung und kam eine Weile lang nicht wieder. Thomas, ein anderer Besucher des Restaurants, stand an der Theke, um zu bezahlen. Markus ging auf Thomas zu und sagte, „Ich habe Kaffee auf meinem Tisch verschüttet. Könnten Sie das aufwischen?"

16.1. Sagte irgendjemand etwas, das er oder sie nicht hätte sagen sollen oder irgendetwas Merkwürdiges?

1. Ja (1)
2. Nein (2)

[Wenn die Antwort auf Frage 16.1 „Nein" ist, dann wird weiter mit Frage 16.2.]

16.1.1. Wenn ja:

Wer sagte etwas, das er oder sie nicht hätte sagen sollen oder etwas Merkwürdiges?

1. Die Bedienung (3)
2. Der Restaurantbesitzer (4)
3. Markus (1)
4. Thomas (2)

16.1.2. Warum hätte er oder sie dies nicht sagen sollen oder warum war es merkwürdig?

Hy p: Markus wollte mit Thomas ins Gespräch kommen. (3)

NT: Der Kellner kommt mit der Arbeit nicht hinterher. (4)

Ko: Markus hielt Thomas fälschlicherweise für einen Kellner. (1)

Hp n: Markus wollte Thomas bloßstellen. (2)

16.1.3. Warum, denken Sie, hat er oder sie das gesagt?

Hy p: Er fand Thomas sympathisch. (3)

NT: Markus trinkt gerne Kaffee. (4)

Hy n: Er wollte ihn beleidigen. (2)

Ko: Er hat ihn verwechselt. (1)

16.1.4. Wusste Markus, dass Thomas ein anderer Besucher war?

1. Ja (1)
2. Nein (2)

16.1.5. Wie denken Sie fühlte sich Thomas?

NT: gleichgültig (4)

Hy_n: beleidigt (2)

Ko: überrascht (1)

Hy p: interessiert (3)

16.2. Kontrollfragen:

16.2.1. Warum stand Thomas in der Geschichte an der Theke?

1. Um einen Kaffee zu bestellen (2)
2. Um zu bezahlen (1)
3. Um die Aussicht zu genießen (3)
4. Um jemanden zu treffen (4)

16.2.2. Was verschüttete Markus?

1. Kaffee (1)
2. Tee (2)

3. Saft (3)
4. Milch (4)

17. Gisela wartete an der Bushaltestelle. Der Bus hatte Verspätung und sie stand dort schon eine lange Zeit. Sie war 65 Jahre alt und es machte sie müde, so lange zu stehen. Als der Bus endlich kam, war er überfüllt und es war kein Sitzplatz mehr frei. Paul, der im Mittelgang des Busses stand, sagte „Hallo, Gisela", „Hast du lange auf den Bus gewartet?" „Etwa 20 Minuten" antwortete sie. Ein junger Mann, der saß stand auf und sagte „Möchten Sie meinen Platz haben?"

17.1. Sagte irgendjemand etwas, das er oder sie nicht hätte sagen sollen oder irgendetwas Merkwürdiges?

1. Ja (1)
2. Nein (2)

[Wenn die Antwort auf Frage 17.1 „Nein" ist, dann wird weiter mit Frage 17.2.]

17.1.1. Wenn ja:

Wer sagte etwas, das er oder sie nicht hätte sagen sollen oder etwas Merkwürdiges?

1. Gisela (1)
2. Der junge Mann (2)
3. Paul (3)
4. Kathi (4)

17.1.2. Warum hätte er oder sie dies nicht sagen sollen oder warum war es merkwürdig?

Hy_n: 65 Jahre sind nicht wirklich alt. (2)

Ko: Ich würde nicht in einen überfüllten Bus einsteigen. (1)

NT: Der Bus hatte Verspätung. (4)

Hy_p: Der Mann ist ein Schleimer. (3)

17.1.3. Warum, denken Sie, hat er oder sie das gesagt?

NT: Es gab einen Stau. (4)

Hy_n: Weil er sie für alt und schwach hielt. (2)

Ko: Weil er dachte, dass sie schon sehr lange stand. (1)

Hy_p: Weil er sich einschmeicheln wollte. (3)

17.1.4. Wusste Paul, als Gisela in den Bus stieg, wie lange sie schon gewartet hatte?

1. Ja (1)
2. Nein (2)

17.1.5. Wie denken Sie fühlte sich Gisela?

Ko: dankbar (1)

Hy_p: gerührt (3)

NT: müde (4)

Hy_n: beleidigt (2)

17.2. **Kontrollfragen**:

17.2.1. Warum wartete Gisela in dieser Geschichte 20 Minuten an der Bushaltestelle?

1. Weil sie zu früh zur Haltestelle kam (4)
2. Weil der Bus Verspätung hatte (1)
3. Weil im Bus kein Platz mehr frei war (2)
4. Weil sie den letzten Bus verpasst hatte (3)

17.2.2. Waren im Bus freie Sitzplätze, als sie einstieg?

1. Ja (1)
2. Nein (2)

18. Martin hatte gerade einen Job in einem neuen Büro angefangen. Eines Tages unterhielt er sich im Pausenraum mit seinem neuen Kollegen Sven. „Was macht denn deine Frau beruflich?" fragte Sven. „Sie ist Anwältin" antwortete Martin. Einige Minuten später kam Sabine in den Pausenraum und sah verärgert aus. „Ich hatte gerade das schlimmste Telefongespräch, das man sich vorstellen kann" erzählte sie. „Anwälte sind alle so arrogant und habgierig. Ich kann sie nicht ausstehen!" „Kommst du mit und siehst dir die Berichte durch?" fragte Sven. „Nicht jetzt" antwortete Sabine, „Ich brauche meinen Kaffee!"

18.1. Sagte irgendjemand etwas, das er oder sie nicht hätte sagen sollen oder irgendetwas Merkwürdiges?

1. Ja (1)
2. Nein (2)

[Wenn die Antwort auf Frage 18.1 „Nein" ist, dann wird weiter mit Frage 18.2.]

18.1.1. Wenn ja:

Wer sagte etwas, das er oder sie nicht hätte sagen sollen oder etwas Merkwürdiges?

1. Sabine (1)
2. Susanne (2)
3. Martin (3)
4. Sven (4)

18.1.2. Warum hätte er oder sie dies nicht sagen sollen oder warum war es merkwürdig?

Hy_n: Weil Martins Frau Anwältin ist, und Sabines Aussage Martin gekränkt hat. Die kommende Zusammenarbeit zwischen Martin und Sabine wird deshalb sehr schwierig werden. (2)

Ko: Weil Martins Frau selbst Anwältin ist, und er Sabines Aussage über Anwälte gehört hat. (1)

Hy_p: Sabine wollte Martin auf sich aufmerksam machen. (3)

NT: Solche Pauschalurteile sind falsch. (4)

18.1.3. Warum, denken Sie, hat er oder sie das gesagt?

NT: Sie hatte zu viel bezahlt. (4)

Hy_p: Sie wollte, dass Martin die schlechten Seiten seiner Frau erkennt. (3)

Ko: Sabine war so aufgebracht, dass sie sich durch diese Aussage Luft machen wollte. (1)

Hy_n: Sie wollte Martin dadurch verletzen. (2)

18.1.4. Wusste Sabine, dass Martins Frau Anwältin war?

1. Ja (1)
2. Nein (2)

18.1.5. Wie denken Sie fühlte sich Martin?

NT: gleichgültig (4)

Hy p: an Sabine interessiert (3)

Ko: verärgert (1)

Hy n: entsetzt (2)

18.2. **Kontrollfragen:**

18.2.1. Was macht Martins Frau in der Geschichte beruflich?

1. Sie ist Lehrerin. (3)
2. Sie ist Musikerin. (4)
3. Sie ist Anwältin. (1)
4. Sie arbeitet bei einer Telefonzentrale. (2)

18.2.2. Wo unterhielten sich Martin und Sven?

1. Im Pausenraum (1)
2. Im Café (2)
3. In der Empfangshalle (3)
4. In Martins Büro (4)

19. Joachim kaufte sich ein neues Auto - einen roten Peugeot. Einige Wochen nach dem Kauf fuhr er rückwärts in das Auto seines Nachbarn Eugen, einen alten kaputten Volvo. Sein neues Auto war überhaupt nicht kaputt und auch Eugens Wagen war nur geringfügig beschädigt worden – nur ein Kratzer im Lack über dem Kotflügel. Dennoch ging er zu Eugen und klopfte an die Tür. Als Eugen aufmachte, sagte Joachim, „es tut mir echt leid. Ich habe deinem Auto gerade einen kleinen Kratzer zugefügt." Eugen schaute sich sein Auto an und sagte, „mach dir mal keine Gedanken. Das war nur ein Versehen."

19.1. Sagte irgendjemand etwas, das er oder sie nicht hätte sagen sollen oder irgendetwas Merkwürdiges?

1. Ja (1)
2. Nein (2)

[Wenn die Antwort auf Frage 19.1 „Nein" ist, dann wird weiter mit Frage 19.2.]

19.1.1. Wenn ja:

Wer sagte etwas, das er oder sie nicht hätte sagen sollen oder etwas Merkwürdiges?

1. Gerhard (4)
2. Joachim (1)
3. Hannes (2)
4. Eugen (3)

19.1.2. Warum hätte er oder sie dies nicht sagen sollen oder warum war es merkwürdig?

Hy_n: Joachim wollte Eugen beweisen, wie toll sein neues Auto ist. (2)

Ko: Joachim hat einen Unfall gebaut. (1)

Hy_p: Eugen reagierte viel zu ruhig auf den Unfall. (3)

NT: Neue Autos sind ihren Preis nicht wert. (4)

19.1.3. Warum, denken Sie, hat er oder sie das gesagt?

NT: Die Nachbarn sind befreundet. (4)

Ko: Er war nicht konzentriert. (1)

Hy_n: Er wollte Eugen unter die Nase reiben, wie toll sein neues Auto war. (2)

Hy_p: Er hatte Angst vor seinem Nachbarn. (3)

19.1.4. Wusste Joachim, wie sein Nachbar Eugen reagieren würde?

1. Ja (1)
2. Nein (2)

19.1.5. Wie denken Sie fühlte sich Joachim?

Hy_p: besorgt (3)

Hy_n: stolz (2)

NT: gestresst (4)

Ko: schuldig (1)

19.2. **Kontrollfragen:**

19.2.1. Was hat Joachim in der Geschichte mit Eugens Auto gemacht?

1. Joachim machte einen Kratzer in Eugens.
   Auto (1)

2. Joachim verkaufte Eugens Auto. (2)
3. Joachim zerstörte Eugens Auto. (3)
4. Joachim reparierte Eugens Auto. (4)

19.2.2. Wie reagierte Eugen?

1. entspannt (1)
2. aufgeregt (2)
3. wütend (3)
4. erleichtert (4)

20. Michaela ging zum Metzger, um etwas Fleisch zu kaufen. In dem Geschäft war es überfüllt und laut. Sie fragte den Metzger „Haben sie Freiland–Hühner?" Er nickte und fing an, ein gebratenes Huhn für sie einzupacken. „Entschuldigen Sie", sagte sie, „Sie haben mich offensichtlich nicht richtig verstanden. Ich fragte Sie, ob Sie Freiland–Hühner haben." „Oh, entschuldigen Sie", sagte der Metzger, „die sind uns heute ausgegangen."

20.1. Sagte irgendjemand etwas, das er oder sie nicht hätte sagen sollen oder irgendetwas Merkwürdiges?

1. Ja (1)
2. Nein (2)

[Wenn die Antwort auf Frage 20.1 „Nein" ist, dann wird weiter mit Frage 20.2.]

20.1.1. Wenn ja:

Wer sagte etwas, das er oder sie nicht hätte sagen sollen oder etwas Merkwürdiges?

1. der Metzger (1)
2. Michaela (2)
3. Maria (3)
4. die Mitarbeiterin (4)

20.1.2. Warum hätte er oder sie dies nicht sagen sollen oder warum war es merkwürdig?

Ko: Der Metzger reagierte nicht wirklich auf die Frage der Kundin. (1)

NT: Die Kundin war zu forsch. (4)

Hy_p: Der Metzger war von Michaela abgelenkt. (3)

Hy_n: Der Metzger wollte das Falsche einpacken. (2)

20.1.3. Warum, denken Sie, hat er oder sie das gesagt?

Hy_p: Weil er Michaela attraktiv fand. (3)

Ko: Der Metzger dachte er wüsste was die Kundin wollte. (1)

Hy_n: Der Metzger wollte seine alten Hühner loswerden. (2)

NT: Michaela war genervt, weil der Metzger das Falsche einpackte. (4)

20.1.4. Wusste der Metzger, als er das gebratene Huhn für Michaela einpackte, dass sie ein Huhn aus Freilandhaltung wollte?

1. Ja (1)
2. Nein (2)

20.1.5. Wie denken Sie fühlte sich Michaela?

Hy_p: geschmeichelt (3)

Ko: enttäuscht (1)

Hy_n: verärgert (2)

NT: ruhig (4)

20.2. **Kontrollfragen:**

20.2.1. Wohin ging Michaela in dieser Geschichte?

1. In den Supermarkt (4)
2. Zum Metzger (1)
3. Zum Markt (2)
4. Zum Bauern (3)

20.2.2. Warum fing der Metzger an, ein gebratenes Huhn für sie einzupacken?

1. Weil er ein gebratenes Huhn verkaufen wollte (2)
2. Weil ihm die Freiland-Hühner ausgegangen waren (3)
3. Weil die Kundin es bestellt hatte (4)
4. Weil er die Kundin falsch verstanden hatte (1)

Gesamtauswertung:

Geschichten mit Fauxpas (Geschichten 2,4,7,11,12,13,14,15,16, 18)

erste Frage („Sagte irgendjemand etwas, das er oder sie nicht hätte sagen sollen oder irgendetwas Merkwürdiges?") = Ja 1 P./ Nein 0 P.

Geschichten ohne Fauxpas (Geschichten 1,3,5,6,8,9,10,17,19,20)

erste Frage („Sagte irgendjemand etwas, das er oder sie nicht hätte sagen sollen oder irgendetwas Merkwürdiges?") = Nein 1 P./ Ja 0 P.

Gesamtwert Summe