# Does My Sadness Blind Me to What is Going on Inside You?
# Studies on Mental State Decoding and Reasoning in Unipolar Depression and Bipolar Disorder

**Dissertation**

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

M. Sc. Julia Feyerabend

aus Heilbronn

Tübingen

2022

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

| | |
|---|---|
| Tag der mündlichen Qualifikation: | 19.10.2022 |
| Dekan: | Prof. Dr. Thilo Stehle |
| 1. Berichterstatter/-in: | Prof. Dr. Martin Hautzinger |
| 2. Berichterstatter/-in: | Prof. Dr. Jennifer Svaldi |

# Acknowledgements

I would like to thank everyone who supported me in carrying out this work in the Department of Clinical Psychology and Psychotherapy at the Eberhard Karls University of Tübingen. The studies of this dissertation were funded by a research grant of the Deutsche Forschungsgemeinschaft awarded to Dr. Larissa Wolkenstein [WO- 1798/2-1].

My special thanks go to Professor Dr. Martin Hautzinger and Dr. Larissa Wolkenstein for setting the topic for this work, for the time they devoted to me and my work and especially for their valuable suggestions and corrections. I also owe a big thank you to my colleagues at the institute: In particular, Stefan Lüttke, Sibylla Wolter, and Fabienne Grosse-Wentrup for their help in preparing the studies and collecting data. I would also like to thank Anja Sommer, Yvonne Trautmann, Isa Scheufele, and Katrin Ziser for their support during data collection. Many thanks also to Professor Dr. Martin Hautzinger and Professor Dr. Jennifer Svaldi for reviewing this work.

# Abstract

The essential components of social functioning include both the ability to decode complex emotional facial expressions (decoding) and the ability - based on the integration of this and other information - to draw conclusions about the mental state of other people (reasoning). These two components are also referred to as Theory of Mind (ToM). Although the social functioning level of patients with major depression and patients with bipolar disorder is significantly impaired, only a few studies have examined the extent to which the two components of ToM are impaired in these patient groups. The studies that do exist predominantly use stimulus material with questionable ecological validity. Therefore, for the first time, the work presented here investigates ToM decoding and ToM reasoning abilities of patients with unipolar depression and bipolar disorder, using highly ecologically valid material. Furthermore, we investigated to what extent such impairments are related to the current state of the patients (acutely depressed vs. remitted) and whether they are possibly first induced or intensified by a negative mood induction. In addition, the impact of possible ToM deficits on the course of illness in patients with affective disorders was investigated.

# Zusammenfassung

Zu den wesentlichen Komponenten des sozialen Funktionierens gehören sowohl die Fähigkeit, komplexe emotionale Gesichtsausdrücke zu erkennen (Dekodierung) als auch die Fähigkeit - basierend auf der Integration dieser und anderer Informationen - Rückschlüsse auf den mentalen Zustand anderer Menschen zu ziehen (Reasoning). Diese beiden Komponenten werden auch als Theory of Mind (ToM) bezeichnet. Obwohl das soziale Funktionsniveau von unipolar depressiven Patienten und bipolaren Patienten deutlich beeinträchtigt ist, gibt es bisher nur wenige Studien darüber, inwieweit die beiden genannten Komponenten der ToM bei diesen Patientengruppen beeinträchtigt sind. Die wenigen Studien, die es gibt, verwenden überwiegend Erhebungsmaterial, dessen ökologische Validität fraglich ist. Daher wurden in der vorliegenden Arbeit erstmals die Decoding- und Reasoning-Fähigkeiten von affektiven Patienten mit ökologisch validem Material untersucht. Weiterhin wurde untersucht, inwieweit mögliche Beeinträchtigungen mit dem aktuellen Zustand der Patienten (akut depressiv vs. remittiert) zusammenhängen und ob sie möglicherweise erst durch eine negative Stimmungsinduktion ausgelöst oder verstärkt werden. Darüber hinaus wurde untersucht, welchen Einfluss mögliche ToM-Defizite auf den Krankheitsverlauf bei Patienten mit affektiven Störungen haben.

# Table of content

# List of abbreviations

## *A*

aBD · Acute depressive or manic mood states of bipolar disorder
ADFES · Amsterdam Dynamic Facial Expression Set
AIM · Affect Infusion Model
ANCOVA · Analysis of Covariance
ANOVA · Analyses of Variance
aUD · Acute depressive mood state of unipolar depression
AVLT · Auditory Verbal Learning Test

## *B*

BD · Bipolar disorder
BDI · Beck Depression Inventory
BDs · Persons with bipolar disorder
Bel · Social Distress (one of the F-SozU main scales)

## *C*

CAM · Cambridge Mindreading Face-Voice Battery
$CAM_{MI}$ · CAM proportion score for the part presented with mood induction
$CAM_{negativity}$ · Percentage of the negativity bias
$CAM\text{-}Negativity_{MI}$ · Percentage of the negativity bias for the part presented with mood induction
$CAM\text{-}Negativity_{NMI}$ · Percentage of the negativity bias for the part presented without mood induction
$CAM\text{-}Neg_{prop}$ · CAM proportion scores for negative mental states
$CAM\text{-}Neut_{prop}$ · CAM proportion scores for neutral mental states
$CAM_{NMI}$ · CAM proportion score for the part presented without mood induction
$CAM_{positivity}$ · Percentage of the positivity bias
$CAM\text{-}Positivity_{MI}$ · Percentage of the positivity bias for the part presented with mood indcution
$CAM\text{-}Positivity_{NMI}$ · Percentage of the positivity bias for the part presented without mood induction
$CAM\text{-}Pos_{prop}$ · CAM proportion scores for positive mental states
$CAM_{prop}$ · CAM proportion scores
CBASP · Cognitive Behavioral Analysis System of Psychotherapy
Com · Number of comorbidity
Con · Confidence scores

## *D*

d · Cohen's d
DSM IV · Diagnostic and Statistical Manual of Mental Disorders

## *E*

Edu · Education
EEMT · Emotional Expression Multimorph Task
EET · Emotion Evaluation Test
EU · Emotional Support (one of the F-SozU main scales)

## *F*

FAST · Functional Assessment Short Test
FB · False Belief task
FEEST · Facial Expressions of Emotions Stimuli and Tests series set
FER · Facial emotion recognition
FER score · Sum of all correctly answered items in the CAM task
FHA · Frith-Happé animations
F-SozU · Social support questionnaire

## *G*

GAF · General Assessment of Functioning scale

## *H*

HC · Healthy control group
HDRS17 · Hamilton Depression Rating Scale

## *I*

IDS-C30 · Inventory of Depressive Symptomatology

## *J*

JACFEE · Japanese and Caucasian Facial Expressions of Emotion

## *L*

LIFE · Longitudinal Interval Follow-up Evaluation

## *M*

MANOVA · Multivariate ANOVA
MASC · Movie for the Assessment of Social Cognition, Movie for the Assessment of Social Cognition
$MASC_{cog}$ · $MASC_{prop}$ for the cognitive mental state modality
$MASC_{emo}$ · $MASC_{prop}$ for the emotional mental state modality
$MASC_{exceeded}$ · MASC error score for the error type *exceeded* ToM
$MASC_{less}$ · MASC error score for the error type *less* ToM
$MASC_{MI}$ · MASC proportion score in the condition with mood induction
$MASC_{NMI}$ · MASC proportion score in the condition without mood induction
$MASC_{no}$ · MASC error score for the error type *no* ToM
$MASC_{prop}$ · MASC proportion scores
MBT · Mentalization-Based Treatment
MEPS · Means Ends Problem Solving Procedure
MI · Mood induction
MNs · Mirror neurons
MSDEF · Montreal Set of Facial Displays of Emotion
MSS · Self-Report Manic Inventory
MWT-B · Multiple Choice Word Fluency Test

## *N*

NumDep · Number of previous depressive episodes

## *P*

PANAS · Positive and Negative Affect Schedule
PEP · Positive Emotion Persistence
POET · Perception of Emotion Test
pSoSup · Perceived social support
PSR · Psychiatric Status Rating
PU · Practical Support (one of the F-SozU main scales)

## *R*

RAVLT · Rey Auditory Verbal Learning Test
rBD · Remitted bipolar disorder
ResSym · Presence of residual symptoms during remission
Rez · Reciprocity (one of the F-SozU additional scales)
RMET · Reading the Mind in the Eyes Task
rUD · Remitted unipolar major depression

## *S*

SCID · Structured Clinical Interview for DSM-IV
SCIT · Social Cognition and Interaction Training
SCWT · Stroop Color Word Test
SI · Social Integration (one of the F-SozU main scales)
SIGHD-IDS · Structured Interview Guide for the Hamilton Depression Scale and Inventory of Depressive Symptomatology
Simp · Simplicity scores
SZ · Schizophrenia

## *T*

t0 · Diagnostic session
t1 · Test session 1
t2 · Test session 2
t3 · Follow-up session
TASIT · The Awareness of Social Inference Test
TMT · Trail Making Test
TMT-A · TMT part A
TMT-B · TMT part B
ToM · Theory of Mind

## *U*

UD · Unipolar major depression
UDs · Persons with unipolar major depression

## *V*

VAS · Visual analogue scale

Vert · Availability of confidants (one of the F-SozU additional scales)
VIF · Variance inflation factor
VLMT · Verbal Learning and Memory Test
V-SIR · Versailles-Situational Intentionl Reading

## *W*

WASI · Wechsler Abbreviated Scale of Intelligence
WasU · F-SozU overall measure of perceived social support
WCST · Wisconsin card sorting test

## *Y*

YMRS · Young Mania Rating Scale

## *Z*

ZCAM · Standardized residuals for change in CAM proportion score in the condition with versus without mood induction
ZCAM-Negativity · Standardized residuals for change in CAM negativity bias in the condition with versus without mood induction
ZCAM-Positivity · Standard residuals for change in CAM positivity bias in the condition with versus without mood induction
ZMASC · Standardized residuals for change in MASC proportion score in the condition with versus without mood induction
Zuf · Satisfaction with social support (one of the F-SozU additional scales)

## *H*

$\eta_G^2$ · Generalized eta squared

# 1 Introduction

## 1.1 Social Cognition

Adaptive social behavior, including successful interaction with conspecifics (Adolphs, 2009) or taking advantage of being part of a certain social group (Frith, 2008) is important, in the broadest sense for ensuring one's own survival and, in a narrow sense, for well-being, life satisfaction and mental health. This requires a complex set of high order neuropsychological processes that have been investigated in the wide research field of *social cognition* (Brothers, 2002). The composite term social cognition relates the components of cognition, such as perception, attention, memory or action planning, to the social context in order to explain the processes by which people understand themselves and others (Beer & Ochsner, 2006; Frith, 2008). According to Beer and Ochsner (2006), these processes include three components: a) *the perception of others*, b) *the perception of self* and c) *social knowledge*. The first component – the perception of others - involves multiple stages of processing. Besides information from any number of sensory channels (e.g., verbal as well as non-verbal social cues), further information may be obtained from the context, or from stored information derived from previous experiences with the context and/or the persons involved. Beer and Ochsner (2006) stated that the second component – the self – serves as a cognitive filter through which others are perceived. Among other things, introspection and personal experience can be used to draw conclusions about the intentions and emotions of others (Meltzoff & Brooks, 2001; Nickerson, 1999) and one's own convictions can also be projected onto others (L. S. Newman, Duff, & Baumeister, 1997). The third component is social knowledge, which is composed of declarative processes (factual knowledge or abstract concepts about social scripts, relationships and phenomena) and procedural processes (rules, skills and strategies that enable us to select reactions or actions in social environments).

Due to the complexity of the field of social cognition, different lines of research have emerged that each refer to a specific part of the overall construct and therefore use different terms (Samamé, 2013). The National Institute of Mental Health distinguishes five dimensions of social cognition (Green et al., 2008): 1) *Theory of Mind* (ToM), the ability to attribute mental states to others (described in more detail later), 2) *social perception*, which refers to the ability to recognize social roles, social rules, and social context, 3) *social knowledge* that refers to awareness of what is socially expected in various situations, e.g., the typical behavior in a restaurant, 4) *attributional bias* or *style*, i.e., the way people typically infer the causes of certain positive and negative events, usually distinguishing between external personal attributions (i.e., causes attributed to other people), external situational attributions (i.e., causes attributed to situational factors), and internal

attributions (i.e., causes attributed to oneself) and 5) *emotional processing*, which refers to the perception and use of emotions and can be divided into the following 4 components: recognizing emotions, enabling emotions, understanding emotions, and managing emotions. There is some interconnect and overlap among these components of social cognition (Beer & Ochsner, 2006), resulting in a lack of conceptual clarity.

This dissertation addresses mainly ToM. The problem of conceptual overlap mentioned above also applies to this specific construct. In particular, under certain conditions, *emotion recognition* (part of emotional processing) is considered part of ToM ability. The overlap or delimitation of emotion recognition and ToM is shown in the following chapters in order to define the conceptual background of the terms to which this thesis mainly refers.

## 1.2 Theory of Mind

ToM refers by definition to a complex ability which results from a composite of different skills. It is defined as the competence to interpret and predict one's own and others' behavior by attributing *mental states* such as feelings, desires, knowledge, intents, expectations and opinions. The ability includes the understanding that mental states of others can be different from one's own (Fonagy, Gergely, Jurist, & Target, 2005; Premack & Woodruff, 1978). It is supposed to be particularly important for the successful maintenance of social contacts, and its impairment can severely hinder interpersonal communication (Van Neerven, Bos, & Van Haren, 2021). ToM includes affective and cognitive aspects (Shamay-Tsoory, Aharon-Peretz, & Perry, 2009). *Affective ToM* is the ability to infer other's emotional states. *Cognitive ToM* relates to the capability to make inferences about other people's beliefs and intentions. Furthermore, according to Sabbagh (2004), ToM encompasses two processes: *ToM decoding* (social-perceptual) and *ToM reasoning* (social-cognitive), a distinction which is sometimes used interchangeably with the affective-cognitive dichotomy (Van Neerven et al., 2021). The first process is the ToM decoding of mental states from observable social information, e.g., facial expression and posture. ToM decoding of facial expression is very similar to the definition of facial emotion recognition (described in more detail below in Chapter 1.2.1) – the ability to recognize and appraise others' emotions by processing relevant facial cues (Green et al., 2008; Lawlor-Savage, Sponheim, & Goghari, 2014; Samamé, 2013). Second, ToM contains reasoning about mental states by integrating observable, historical and further contextual information about a person to understand behavior. There is growing evidence that decoding and reasoning rely on different social brain networks (Martino, Strejilevich, Fassi, Marengo, & Igoa, 2011; Sabbagh, 2004). Activation of the orbitofrontal cortex and temporal cortex is more associated with the ToM decoding component, whereas activation of the frontal medial cortex is related to the ToM reasoning

component. Therefore, both components of the ToM need not necessarily be involved and can probably be dissociable (Martino et al., 2011).

This dissertation mainly refers to, first*, ToM decoding* or *complex facial emotion recognition* respectively and second, *ToM reasoning*. These terms are explained in more detail in the following two subsections.

### 1.2.1 Conceptual clarification of emotion recognition and ToM decoding

From an everyday perspective, at least two people are involved in the process of emotion recognition: The person who experiences and expresses the emotion, and the person who perceives the expressed emotion and tries to interpret it correctly. This implies that in order to understand the process of emotion recognition on the receiver side, the sender side must also be considered. For this purpose it is necessary to know what emotions are, how they are expressed and how they can be decoded and interpreted by the perceiver. As already defined several times in the literature (e.g., Adolphs, 2002; Damasio, 1995; Scherer, 2000), an *emotional experience* is the response to the value of a stimulus. Besides changes in subjective experience (e.g., have the subjective feeling of "fear") this response also contains changes in physiology, including endocrine, visceral and autonomic changes (Adolphs, 2002), as well as thoughts and behaviors (Gross, 2015; Mauss, Levenson, McCarter, Wilhelm, & Gross, 2005). The latter - emotion-related behavior – includes, for example, changes in facial expressions, body posture, intonation of the voice and situation-specific behaviors like avoidance or approach (Ekman et al., 1987; Frijda, 1986). Against the background of this definition, emotion recognition means the ability to decode other's emotional state correctly by processing the expressed emotion-related social cues. These cues can mainly be perceived visually (e.g., "reading" the feeling from the body posture or facial expression) or audibly (e.g., recognizing a certain tone of voice) (Lawlor-Savage et al., 2014; Samamé, 2013).

The most commonly investigated facet of emotion recognition is *facial emotion recognition* (FER)*,* i. e., how people perceive emotions in other's facial expressions. Facial emotional expressions are usually recognized quickly (Prkachin, 2003). Bombari et al. (2013) found that FER is still possible even if the presented faces are inverted, scrambled or blurred. Furthermore, it seems to be more robust for alterations than is facial identity recognition. The authors concluded that the importance of reliable FER for everyday interactions requires a solid and reliable system for recognizing emotions in the face, which should also work when information about features or configurations is less available (e.g., due to distance). This assertion is supported by the *evolutionary theory view*, which makes the presence of reliable FER performance a fundamental skill for successful adaption and interpersonal relationships (Surcinelli, Codispoti, Montebarocci, Rossi, & Baldaro, 2006).

Evidence from neuroimaging studies indicates that perceiving others' emotions is a complex mechanism involving at least *three interacting substeps* (Adolphs, 2002; Garrido-Vásquez, Jessen, &

Kotz, 2011; Schirmer & Kotz, 2006): 1) early sensory processing, in which simple and highly salient features are processed, 2) integration of sensory cues to form a salient percept, and 3) evaluation of the perceived cues, which means that cognitive processes operate on such percepts. According to Garrido-Vásquez et al. (2011), in FER, these three steps can be assigned to either the *perception* or the *recognition part*. In the *perception part*, the early sensory processing takes place. In the *recognition part* integrating and evaluating the percept takes place through more detailed emotion-processing mechanisms, as well as through making associations with previously known information about the context.

Different psychological theories of emotions

There are several popular approaches to what *emotions* are, and only some of them include explicit assumptions of how emotional perception works. The most popular models can be grouped into four categories: 1) the *discrete emotion models*, 2) the *dimensional models*, 3) the *meaning-oriented models* and 4) the *componential models* (Scherer, 2000).

Paul Ekman supports the idea of *separate, discrete emotional states* (Ekman, 1992). In this context Ekman also speaks of *basic emotions* that evolved to prepare us for the accomplishment of basic life-tasks. He uses the term basic emotion in order to designate separate, physiologically distinct emotions that are different from one another and in which evolution has played an important role in shaping the characteristics that these emotions display. The distinct basic emotions should differ in their expression, appraisal, behavioral responses and physiology. Based on the finding that basic emotions are recognized across cultures, Ekman (1993) concludes that the ability to have, express and perceive emotions is universal and innate. The theories of discrete emotions are historically linked to the study of facial expressions (Scherer, 2000). Ekmans' research on facial expressions revealed *six basic emotions*: happiness, sadness, fear, anger, disgust and surprise (Ekman, 1999; Ekman & Friesen, 2003). Other non-basic or *complex emotions* are seen as combinations of the basic emotions (Ekman, 1992). For example, hate can be seen as a combination of anger, fear, and disgust. The main focus of the discrete emotion approach – that is, motor expression or adaptive behavior patterns, including facial expression (Scherer, 2000) – can easily be linked to assumptions about how emotion recognition works. For example, FER is done by assigning labels to a certain facial expression.

In contrast, Russell (1980) takes a *dimensional view* in which he postulates that affective states are related to each other systematically and are best represented as a circle in a two-dimensional bipolar space. In his *circumplex model,* the horizontal dimension is *valence* (pleasant versus unpleasant) and the vertical dimension is *arousal* (activation vs. deactivation).

The meaning-oriented models can be further divided into *lexical models* and *social constructivist models*. According to Scherer (2000) the lexical models try to deduce emotion models

from the semantic domains of emotion terms. For example, Ortony, Clore, and Collins (1988) conducted a structural analysis of the emotion lexicon to uncover the underlying semantic implicational structure. Supporters of the social constructivist models do not deny the psychobiological response components of emotions, but regard them as secondary. They emphasize that the meaning of emotions is generally constituted or constructed by socioculturally determined patterns of behavior and values. Here, the emotion lexicon also plays a major role, as it is assumed that the emotion labels available in a language reflect the emotional meaning structures in that culture.

Special attention in emotion research has also been given to representatives of *componential models*. The basic assumptions are that the cognitive evaluation of antecedent events evokes emotions (Scherer, 2000). Furthermore, the outcome of this evaluation process determines the reaction pattern in the different response domains (physiology, expression, action tendencies and feelings). Whereas Lazarus postulates that a limited number of basic topics in appraisal produce a limited number of major emotions (Lazarus, 1968, 1991), Scherer supposes that there are as many different emotional states as there are different patterns of appraisal outcomes (Scherer, 1982, 1984a, 1984b, 1993).

Basic emotion recognition versus ToM decoding

With regard to the above-mentioned lack of conceptual clarity, it should be stressed that, although emotion recognition and ToM are conceptually separated to a certain extent, the decoding component of ToM can be considered to be closely related to what emotion recognition represents (Green et al., 2008). However, ToM decoding goes beyond pure *basic FER* and contains a ToM component that is not necessary for basic FER (Martino et al., 2011). As described in more detail below in this chapter, Damasio (2014) provided further specification by equating *ToM decoding* with FER of only *complex emotions* and other *mental states*.

The main difference between emotion recognition and ToM decoding is that ToM decoding can refer to *any* mental state, not just basic emotions. Mental states also include the complex emotions already mentioned, that are composed of two or more basic emotions. Furthermore, Adolphs (2002) separates basic emotions from *social emotions*. The latter are also referred to as moral and self-conscious emotions and serve explicitly to regulate social behavior. These include emotions such as embarrassment, pride, and feelings of guilt. He further assumes that the perception of social emotions requires a more comprehensive self-representation than the perception of basic emotions, since this includes the representation of oneself in a network of social relationships and requires the representation of the other individual's mental states (e.g., how someone else feels about oneself in a certain situation). Thus, this goes beyond "pure" emotion recognition and concerns the domain of ToM. Other mental states can be further specified and

differentiated from emotions. For example, emotions differ from *motivational states* like pain or hunger in that they involve social communication, and from *mood* in that they have a phasic character: an onset, a finite duration and an offset (Adolphs, 2002).

Similarly, more ToM than simple emotion recognition is the following content: On the perceiver side, correct emotion recognition also requires the knowledge that emotions depend on internal cognitive processes of the person expressing the emotion. In other words, the perceiver must have an understanding of *cognitive, belief-based emotions*. Belief-based emotions have been described by Howlin, Baron-Cohen, and Hadwin (1999) as "emotions caused by what something thinks is the case, even if what they think conflicts with reality" (p. 130). These processes clearly belong more to the domain of ToM than to simple emotion recognition.

Therefore, ToM decoding it is not so much FER of basic emotions as FER of complex emotions (Damasio, 2014), including social emotions (Baron-Cohen, Jolliffe, Mortimore, & Robertson, 1997; Kasari, Chamberlain, & Bauminger, 2001), and cognitive, belief-based emotions (Harris, 1989). Furthermore, ToM decoding is also FER of other mental states, including motivational states or moods (such as thoughtful, playful or bored).

### 1.2.2 Conceptual clarification of ToM reasoning

The assumption that others have minds is referred to as a *theory* – i. e. Theory of Mind – because no one has direct access to another's mind and its contents. So a mind's existence and functioning can only be inferred from observations of others behavior e.g., their verbal or facial expressions and motor actions (Premack & Woodruff, 1978). According to Sabbagh (2004), whereas ToM decoding involves some of the relatively rudimentary skills, ToM reasoning may include some of the more complex aspects of ToM. Although ToM decoding and ToM reasoning are two aspects within ToM, distinguishing these two processes is important, as they each rely mainly on different types of social information processing skills. ToM decoding relies primarily on social information from the immediate and observable environment (for example the person's actions, tone of voice etc. in addition to the facial expression). In contrast, ToM reasoning about the mental state of others to explain or predict actions requires access to knowledge and facts about either the person or their contextual circumstances.

The ability that enables ToM is called "mentalizing" or "mind reading" and refers to how we gain knowledge about the mental state of other people (Frith, 2008; Frith & Frith, 2005). The mechanisms underlying this ability are only rudimentarily known. It seems that it crucially depends on the ability to form metarepresentations, i.e., representations that are disconnected from reality. There are mainly two explanatory approaches of the underlying mechanisms of mentalizing (Frith & Frith, 2005): (1) *The theory theory* and (2) *The simulation theory*.

Different theories of how we acquire a ToM

*The theory theory:* The full scope of theory theory is a view of how all types of concepts (not just in the context of ToM) are structured, acquired, and used. The basic assumptions of the approach were also used as an explanation for the mindreading ability. Here it is assumed that people have a basic or "naive" theory of psychology ("folk psychology") – that is a partially tacit, systematically organized body of knowledge – to infer the mental states of others (Ratcliffe, 2006). In recent years, a *Bayesian framework* has been introduced to explain how people learn and understand concepts and assumptions about the world in a mathematical way (e.g., Ullman & Tenenbaum, 2020). *Bayesian inference* is an approach that describes how the probability of a domain-specific hypothesis or theory is updated as more empirical evidence, information, or insight becomes available (e.g., Box & Tiao, 2011). Bonawitz, van Schijndel, Friel, and Schulz (2012) report that a growing number of studies have shown that people act in ways consistent with optimal Bayesian inference. Therefore, people form theories – on the basis of which they mentalize – through a process of theory revision that is very similar to the way scientists propose and revise theories.

*The simulation theory of empathy:* This approach, on the other hand, assumes that conceptual thinking and explicit reflective mediation are not necessary to understand other people. Rather, this approach assumes that there are neural mechanisms (mirror mechanisms) in our brain that enable us to directly understand the meaning of others' actions and emotions by internally recreating ("simulating") them (Gallese, Keysers, & Rizzolatti, 2004). Therefore, according to this approach, the fundamental mechanism of mentalizing is the activation of the *mirror neuron system*. *Mirror neurons* (MNs) are a class of neurons first discovered in the premotor cortex of macaque monkeys, which respond selectively when the monkey observes one type of action (di Pellegrino, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992; Rizzolatti & Arbib, 1998). From this, the concept of MNs was derived, which states that there is a common neural representation for both one's own motor behavior and the observation of the same motor behavior in others. There is strong evidence that the mirror system exists in humans as well (e.g., Gallese & Goldman, 1998; Keysers & Gazzola, 2010). Watching a person performing an action, we activate a network of parietal and premotor areas that is also active when we perform similar actions ourselves. In addition, the concept of the MNs system has been extended to the field of emotions (Carr, Iacoboni, Dubeau, Mazziotta, & Lenzi, 2003; Gallese et al., 2004). Therefore, when we observe another person’s emotions, we activate neural structures that are normally involved in our own personally experienced emotions. For example, when we perceive a sad facial expression of our counterpart, we activate the part of our insula that is also active when we ourselves feel sadness.

Development of ToM

Much of the knowledge about ToM and its development has been gained in studies of how ToM unfolds in children (Sabbagh, 2004). ToM initially appears to be an innate potential ability of humans and needs social and other experiences over many years for its full development (Demetriou, Spanoudis, & Mouyi, 2010). Different people may develop a more or less effective ToM. Carlson, Koenig, and Harms (2013) summarize important milestones in children's development of ToM: Much of ToM develops in a stable, predictable sequence between the ages of 2 and 5 years. At age 2, children have a basic understanding of emotions, intentions, desires, and perceptions, but show very little understanding of knowledge and beliefs. At this age, it is still difficult for them to understand that other people may have different beliefs and levels of knowledge or that someone may think or believe something that is not true. They also have problems recognizing that appearances may be not the same as reality and that people can have different visual perspectives on the same event. However, by the age of 4 or 5, children already have a more adult-like understanding of these things.

Moreover, there is evidence that ToM operates in two ways - *implicitly* and *explicitly* – and that these different ways of functioning also manifest themselves at different times in child development (e.g., Frith, 2008; Schuwerk, Vuori, & Sodian, 2015; Sodian, 2011). The implicit form develops early. With it, one is able to consider an agent's mental state in order to make correct action anticipation without consciously thinking about it. This proceeds quickly, largely automatically, unconsciously, and inflexibly. The processes run automatically, whether we want them to or not, because we have no top-down control over them. Examples for these processes are gaze following (i.e., people follow the gaze of other people) and imitating actions (as already described in the paragraph on mirror neurons). Explicit ToM reasoning, on the other hand, develops later, around age 4. This enables conscious reasoning (e.g., judging) about the mental states of others in a conscious and flexible way and is cognitively demanding (Wellman, Cross, & Watson, 2001). At this higher level, the processes are slower and require mental effort (Frith & Frith, 2008). These high-level control processes can be used to override automatic (pro-)social behavior. Moreover, we use our cognitive abilities explicitly and ''rationally'' to reason why, in our case, morally questionable behavior or motives (such as greed or selfishness) are good.

## 1.3 Methods to investigate FER and ToM and their ecological validity

### 1.3.1 Measures of basic FER ability

Especially at the beginning of the research of FER, research studies used *static drawings of faces*, before more authentic material became available. For example, Cüceloglu (1970) generated sixty schematic drawings of facial expressions by combining four eyebrow types, three eye types and

five mouth types. In other studies, slightly modified versions of the line drawings of Cüceloglu (1970) were used (e.g., Bouhuys, Bloem, & Groothuis, 1995; Bouhuys, Geerts, & Gordijn, 1999).

Over time, researchers have been concerned with increasing the *ecological validity* of the stimuli, i.e., how test performance predicts behaviors in real-world settings (Barker, Musso, & Gouvier, n.d.). Increasingly, studies investigating FER have used *standardized static photographs* of facial expressions of basic emotions (fear, sadness, disgust, happiness, anger and surprise). The most commonly used stimulus set to study FER is a set of 60 photographs created by Ekman and Friesen (1976) showing the six basic emotions. Other examples of standardized static facial stimuli that express basic emotions are given in the article by Darke, Cropper, and Carter (2019) and include the Japanese and Caucasian Facial Expressions of Emotion (JACFEE; Biehl et al., 1997), the Montreal Set of Facial Displays of Emotion (MSDEF; Beaupré & Hess, 2005), and the NimStim Set of Facial Expressions (Tottenham et al., 2009). However, the ecological validity of emotion expressions that a) are presented in *static* images, b) represented only *basal* emotions, and c) have *high* expressed *intensity*, can be questioned.

In everyday life we are rarely confronted with fully developed facial emotional expressions, but rather with subtle expressions of a low emotional intensity level. To investigate the dependence of FER performance on the intensity level of emotional facial expression, *morphed pictures from facial emotional expressions* were created. The stimuli of a very common morphing task (A. W. Young et al., 1997) used the six basic emotion expressions of the standard set of expression photographs by Ekman and Friesen (1976). Morphing had been used to manipulate the intensity of the emotions expressed in the faces by adding increasing proportions of neutral expression to the prototypical emotional expression. A further development of this task is the Facial Expressions of Emotions Stimuli and Tests series set (FEEST; A. W. Young, Perrett, Calder, Sprengelmeyer, & Ekman, 2002). In research, the morphed pictures have been used primarily to identify possible differences in *sensitivity* to emotional expressions. Although these stimulus sets contain pictures of more subtle emotion expressions, they are still static, and therefore of low ecological validity. Darke et al. (2019) argue that expressions are inherently dynamic and that static images may be too impoverished to properly exploit emotion processing mechanisms (Fiorentini & Viviani, 2011).

In order to address this insufficiency, *dynamic materials* also have been developed to examine basic FER. There are two different types of dynamic stimuli: a) *dynamic morphs* and b) *short videos* in which actors express a certain emotion on their face. Regarding the former, the dynamic morphs, some researchers have created dynamic stimuli that slowly morph through several iterations. For example, continuous morphs of facial emotions are presented from neutral intensity (0%) up to 100% or even 150% emotional intensity. The participants are asked to react (e.g., by pressing a button) as soon as they are sure they recognize the expression. One example is the

Emotional Expression Multimorph Task (EEMT; Rich et al., 2008) that is a variation of a task by Blair, Colledge, Murray, and Mitchell (2001). In the EEMT too, the pictures used to develop the stimuli were also taken from Ekman and Friesen (1976) and further processed by blending a picture of a prototypical emotional expression with a picture of neutral facial expression. The faces slowly morph through 39 iterations from neutral intensity (0%) to full emotional intensity (100%). Even more realistic displays than dynamic morphs are *video-based* instruments. One video-based instrument with high ecological validity is *The Awareness of Social Inference Test* (TASIT; McDonald, 2002). The TASIT comprises videotaped vignettes of everyday social situations enacted by professional actors. TASIT has three parts. Basic emotion recognition is investigated by part 1, called the *Emotion Evaluation Test* (EET). In EET, the actor's behavior (voice, facial expression, and gestures) together with the social situation indicate the emotional meaning. Actors express one of six basic emotions or a neutral state, whereas the scripts are ambiguous in content.

Darke et al. (2019) listed some of the dynamic stimuli sets of facial expressions of basic emotions: the *Perception of Emotion Test* (POET; Kilts, Egan, Gideon, Ely, & Hoffman, 2003), the Cohn-Kanade Facial Expression Database (Kanade, Cohn, & Tian, 2000), the CMU-Pittsburgh AU-Coded Face Expression Image Database (Kanade et al., 2000), and the Amsterdam Dynamic Facial Expression Set (ADFES; van der Schalk, Hawk, Fischer, & Doosje, 2011).

It must be borne in mind, that – as mentioned above – FER as a component of ToM refers to the recognition of complex emotions and mental states rather than of pure basic emotions (Damasio, 2014). Thus, the instruments listed in this chapter assess basal FER ability, which is distinct from ToM decoding ability.

### 1.3.2 Measures of ToM decoding ability

Interestingly, there are only a few studies and also a few stimuli sets that address complex FER ability or ToM decoding respectively. Studies that do exist used the *Reading the Mind in the Eyes Task* (RMET; Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001). In RMET, a series of static photographs displaying only the eye region of the face is shown. Participants are instructed to choose which of four words best describe what the person in the photograph is thinking or feeling. This instrument refers to the recognition of *complex* and *social emotions* as well as other *mental states,* like motivational or mood states. Examples are emotions or states like skeptical, worried, panicked, anticipating, playful, and flirting. Therefore, the RMET is considered suitable for capturing ToM decoding (Bora & Zorlu, 2017). Again, the ecological validity of these stimuli is limited, since unlike in everyday experience, the RMET only presents the eye region and uses static instead of dynamic stimuli.

Dynamic stimuli to capture ToM decoding capability are the *Frith-Happé animations* (FHA; Abell, Happé, & Frith, 2000). These are animations of a big red triangle and a small blue triangle

moving about the computer screen. The movement of the two triangles mimic human behavior and implied complex mental states. Although this material contains dynamic stimuli, these are merely abstract shapes instead of showing real people with naturalistic human movements, severely limiting ecological validity.

So far, the measures used to investigate ToM decoding are of questionable ecological validity. In everyday life, we can use information from the entire facial expression to decode the emotional state of a certain person, instead of just the eye area as in the RMET. In addition, in real situations we are confronted with *dynamic* facial expressions of humans which are often ambiguous, complex and subtle, and not with movements of abstract forms. We define ToM decoding tasks as being of high ecological validity when the following demands are met: the stimuli a) are dynamic and b) display real persons who in turn display c) complex emotional expressions or mental states in d) different developmental stages (ranging from subtle to fully developed). As far as we know, there is only one instrument that meets these criteria: the *Cambridge Mindreading Face-Voice Battery* (CAM) by Golan, Baron-Cohen, and Hill (2006), described in more detail in Chapter 2.1. It goes back to a lexical model, that has been introduced by Baron-Cohen, Golan, Wheelwright, and Hill (2004) and contains short videos of actors expressing complex emotions and mental states on their faces.

### 1.3.3 Measures of ToM reasoning ability

To date, ToM reasoning has mostly been investigated by means of *stories* (e.g., Fletcher et al., 1995) and *cartoon comprehension* (e.g., Kerr, Dunbar, & Bentall, 2003). The participants are presented with stories or cartoons and asked comprehension questions whose correct answer requires reasoning about the mental state of the protagonists. The ToM questions are sometimes combined with *picture sequence tasks* in which participants are asked to correctly sequence pictures (usually drawn or in form of a cartoon) of a social event before answering ToM questions about the story (e.g., Inoue, Tonooka, Yamada, & Kanba, 2004). One of the most frequently used tasks from the category story comprehension is the *Faux pas recognition task* (Stone, Baron-Cohen, & Knight, 1998). Participants are to read a story that involves a faux pas. In addition, the story is placed in written form in front of the participant while it is read aloud and during the subsequent interview. An example for the cartoon tasks is the *Brüne's cartoon picture story test* (Brüne, 2003). Brüne's test includes a series of cartoon picture stories with three types: (1) cooperation of two people; (2) a deception scenario in which one person cheated a second person; and (3) two people cooperating at the cost of a third person. Each picture story consists of four cards, presented face down and in reverse order. Participants were asked to turn the cards over and put them in a logical order. If they failed to sequence the picture story correctly, the experimenter corrected them so that the correct order was available for the subsequent ToM test. *False Belief tasks* (FB) are also classically used in ToM reasoning research. Frith and Corcoran (1996) introduced a FB task that is brief and made

minimal demands on general cognitive resources. Stories are read to the participants, examining their ability to appreciate f*irst order false beliefs* (A person has a false belief about the state of the world) and *second order false beliefs* (A person has a false belief about the beliefs of another character), and deception. While the story is read to the participant, the participant is simultaneously presented with cartoon drawings depicting the action sequences of the story read aloud. Another often used task is the *Hinting task* (Corcoran, Mercer, & Frith, 1995). It tests the participants' ability to infer the true intentions behind indirect speech utterances. Short passages describe interactions between two people which end with one person making hints to the other. The participant is asked what the character really meant.

These reasoning tasks are primarily *verbal* material based on comprehension of stories read aloud, or self-read, and/or on processing a series of *static* images. Therefore, they do not meet the demands of highly ecologically valid ToM reasoning tasks, which we defined as: a) using *dynamic* stimuli, b) showing *real* persons, and c) displaying everyday life situations to ensure the possibility of using contextual information to draw conclusions. A common, ecologically highly valid measurement method of ToM reasoning that meets these criteria is the *Movie for the Assessment of Social Cognition* (*MASC*; Dziobek et al., 2006). This task is a naturalistic measure of ToM reasoning, since it includes real peoples' actions, voices and gestures, as well as contextual information to attribute mental states (Samamé, 2013). In this task, participants watch a movie about four people meeting for a dinner. The movie shows everyday interactions and is paused at particular points for asking questions referring to the mental states of the characters. Besides its ecological validity, a further advantage of MASC is its lower requirement on general neurocognitive abilities, in comparison to e.g., story comprehension or sequencing tasks which require a higher level of abstraction. Therefore, it may be expected that the MASC measure reflects a purer measure of ToM reasoning which is less contaminated by general neurocognitive abilities. Another video-based ToM reasoning source is TASIT. It consists of three parts. Part 1 (The Emotion Evaluation Test) shows short videos of basic emotion expression (described in more detail above, in Chapter 1.3.1). Part 2 (Social Inference – Minimal) contains short video vignettes of sincere, sarcastic or paradoxical sarcastic communication between actors. Dialogues are ambiguous, so participants have to pay attention to the general behavior, tone of voice, facial expressions and/or gestures to interpret the situation. At the end of each clip, participants were asked four questions designed to elicit interpretations of what the speaker was thinking, doing (e.g., criticizing), meaning to say, and feeling. The TASIT, Part 3 (Social Inference – Enriched), also consists of videotaped scenes. The scenes include two types of conversations, presented as either a lie (a white lie or a sympathetic one) or sarcasm. After each scene, participants were asked questions (yes/no) about the actors' communicative intentions, such as whether the characters wanted the literal or non-literal meaning of their statements to be

believed. Another instrument, the *Versailles-Situational Intentional Reading* (V-SIR) has been proposed by Bazin et al. (2009). It's a task that requires interpretation of the mental states of one of the characters who interact socially in a complex way (for example, hints, lies, and indirect speech) in a videotaped scene, each lasting 10 to 70 seconds. After each scene, a question is asked about one of the characters.

One main difference between MASC and the other two tasks (TASIT and V-SIR) is that MASC consists of one continuous story (about 15 minutes). The other two tasks include short video clips that last only a few seconds and are not related to each other. Moreover, the videos of TASIT and V-SIR target specific ToM aspects such as implicit intentions or sarcasm. Thus, the MASC is superior to the other two instruments in terms of ecological validity, and it also considers a wider range of different ToM reasoning facets (described in more detail in Chapter 2.2).

## 1.4 Variations in FER and ToM performance

### 1.4.1 Variations in FER and ToM performance in healthy people

The capacity for ToM is considered a stable personality trait and is subject to large individual differences (Leiberg & Anders, 2006). Kinderman, Dunbar, and Bentall (1998) already reported that ToM is not an ability that is either present or absent but seems to be more or less pronounced in different individuals. Therefore, some people are generally more successful at mentalizing than others (Leiberg & Anders, 2006). In this context, there is a considerable body of research on individual differences in basic FER and ToM ability and the causes of these differences. Factors and variables have been identified that have an impact on basic FER and ToM ability or are at least related to it. Some of these research findings are outlined in the following.

Social cognition is partly influenced by "pure" *general cognitive abilities* like intelligence, executive functions or working memory. In line with this, it was found that basic FER performance is related to years of education (Kessels, Montagne, Hendriks, Perrett, & de Haan, 2014) and cognitive abilities like fluid intelligence, processing speed and memory (e.g., Horning, Cornwell, & Davis, 2012; Schlegel et al., 2020). By using the RMET (Baron-Cohen et al., 2001), which is a complex FER task (ToM decoding) it was found that individual differences in ToM decoding are partially attributable to fluid intelligence variability (Ibanez et al., 2013) as well as general intelligence variability (Coyle, Elpers, Gonzalez, Freeman, & Baggio, 2018). Coyle et al. (2018) further provided evidence that general intelligence also influences ToM reasoning ability.

In addition, there is evidence that there is an influence of *gender*, such that females outperform males in tasks of basic FER (Campbell et al., 2002; Kessels et al., 2014; Montagne, Kessels, Frigerio, de Haan, & Perrett, 2005; Thayer & Johnsen, 2000; Vaskinn, Sundet, Friis, Simonsen, Birkenæs, et al., 2007) and ToM decoding (RMET) (DeSoto, Bumgardner, Close, & Geary, 2007; Ibanez

et al., 2013). Regarding ToM reasoning, Turkstra et al. (2020) report in their review that only a few studies have addressed gender differences in this. Most studies found no significant difference between men and women in ToM reasoning tasks, although some studies found a trend toward better outcomes for women.

Furthermore, there appears to be an *age-related* decline in basic FER in adults (e.g., Horning et al., 2012; Kessels et al., 2014; Ruffman, Henry, Livingstone, & Phillips, 2008), especially in the recognition of negative emotions like anger, fear and sadness (West et al., 2012). L. M. Williams et al. (2009) found that, across a lifespan, the basic FER performance curve follows an inverted U-shaped trajectory. Accordingly, young and middle-aged adults are the most accurate compared to children and older adults. This age effect has been associated with age-related changes in general cognitive abilities, but has not been fully explained by these changes (Horning et al., 2012). In their meta-analytic review, Henry, Phillips, Ruffman, and Bailey (2013) investigated the age-related difference in ToM (decoding and reasoning) performance. They further distinguished according to domain (affective, cognitive, mixed) and modality (verbal, visual-static, visual-dynamic, verbal and visual-static, or verbal and visual-dynamic). Overall, older adults were found to perform worse than younger adults across all task types, domains and modalities.

The *current mood* also has an influence on basic FER with regard to mood-congruent biases, i. e. perceiving more sadness in ambiguous line drawings of faces after a negative mood induction (e.g., Bouhuys et al., 1995).

The experience of *stress* also appears to have a significant impact on social cognition. Deckers et al. (2015) found that basic FER performance was higher after than before stress in healthy people. Smeets, Dziobek, and Wolf (2009) found that a higher stress-induced cortisol reactivity impaired ToM reasoning performance (measured by MASC) in woman, but enhanced MASC performance in men, compared to non-stressed controls. However, regarding ToM decoding (measured by RMET), the authors found no difference among non-stressed controls, stressed participants with high cortisol reactivity, and stressed participants with low cortisol reactivity.

Furthermore, there are studies suggesting an influence of certain *hormones*. For example, healthy females during their follicular phase (increasing estrogen level, low progesterone level) show higher accuracy in basic FER performance in comparison to females during their luteal phase (declining estrogen level, higher progesterone level) (Derntl, Kryspin-Exner, Fernbach, Moser, & Habel, 2008). Domes, Heinrichs, Michel, Berger, and Herpertz (2007) found that oxytocin improves ToM decoding (measured by RMET) in healthy people. Also testosterone seems to play a role in ToM: According to DeSoto et al. (2007) higher levels of testosterone are associated with more errors in ToM decoding (RMET). However, a different pattern was observed among men with the highest testosterone levels: Specifically, males who were nearly 2 SD's above the mean testosterone levels

made fewer errors. Furthermore, genetic variations in the receptors associated with oxytocin, vasopressin, dopamine and serotonin could account for individual differences in social cognition and social behavior (e.g., Skuse & Gallagher, 2011).

Ludyga et al. (2020) examined whether basic FER and ToM decoding (RMET) performance is related to *cardiorespiratory fitness*. Results indicate cardiorespiratory fitness to explain a unique proportion of variance in models predicting overall performance (basic FER and ToM decoding) and basic FER performance, but not with respect to the ToM decoding performance, although the trend is in the same direction. The authors discuss that, statistically, this could be because other covariates (such as gender, age, and education) already explained a greater proportion of the variance in models predicting ToM decoding performance.

Also *characteristics of the stimulus material* with which social cognition is measured seem to be crucial for the performance of the participants in the respective tests. In terms of basic FER, Darke et al. (2019) concluded that dynamic faces appear to have an advantage over static faces, as there are a variety of studies demonstrating better performance in recognizing emotions from dynamic faces compared to static faces. Darke et al. (2019) further underpinned this claim with findings that dynamic compared to static emotional face displays elicit greater self-reported "emotional experience" (Yoshikawa & Sato, 2006), and that dynamic faces are perceived as more intense than static faces (Biele & Grabowska, 2006). They also reported results from a PET study showing significantly different patterns of brain activation for dynamic compared to static emotional face displays (Kilts et al., 2003).

The answer to the question of what is associated with variations in FER and ToM seems to need to include an even more multifaceted perspective, as there are studies that point to complex interrelationships of FER and ToM with other variables (interactions, moderating and mediating effects of some variables) (e.g., Kanske, Böckler, & Singer, 2015; Smeets et al., 2009). And characteristics of the stimulus material (e.g., static versus dynamic) also seem to play a crucial role.

### 1.4.2 Variations in FER and ToM performance in people suffering from psychiatric disorders

Impaired social and interpersonal functioning is often associated with mental health disorders, either as a defining criterion (e.g., autism spectrum disorders, American Psychiatric Association, 1994) or as a side effect (e.g., schizophrenia, Baez et al., 2013). The presence of deficits or abnormalities in social cognition is one explanation used to understand these observed problems (e.g., Berecz, Tényi, & Herold, 2016; Emre Bora, C. Bartholomeusz, & Christos Pantelis, 2016; Samamé, 2013; Weightman, Air, & Baune, 2014). Indeed, social cognition has been consistently found to be impaired in various psychopathological groups, and this impairment is associated with poorer social and functional outcome (e.g., Fett, Viechtbauer, Penn, van Os, & Krabbendam, 2011;

Fulford, Peckham, Johnson, & Johnson, 2014). Previous research on social cognitive dysfunction in psychiatric disorders has mostly focused on FER and ToM (Van Neerven et al., 2021). Deficits or abnormalities in FER and ToM have been reported for patients suffering from autism (e.g., Baron-Cohen, 2000; Fernandes, Cajão, Lopes, Jerónimo, & Barahona-Corrêa, 2018), schizophrenia (e.g., Bora, Yücel, & Pantelis, 2009; Garrido-Vásquez et al., 2011), bipolar disorder (e.g., Emre Bora, C. Bartholomeusz, et al., 2016), major depression (e.g., Bora & Berk, 2016; Garrido-Vásquez et al., 2011), anxiety disorders (e.g., Hezel & McNally, 2014), eating disorders (e.g., Tapajóz Pereira de Sampaio, Soneira, Aulicino, & Allegri, 2013), obsessive-compulsive disorder (e.g., Liu et al., 2017), and also in people with personality disorders like borderline personality disorder (e.g., Hillmann et al., 2021; Németh et al., 2018).

Even in healthy individuals, a variety of factors have been identified (see Chapter 1.4.1) that influence FER and ToM, and it is likely that the degree of complexity increases further when psychopathological processes become influential. Some of the variables that have been identified as relevant to FER and ToM performance in healthy individuals are altered in individuals with certain mental illnesses and have also been linked to abnormalities in social cognition in these patient groups. Here are some examples: Schizophrenia is associated with disturbances in the dopaminergic system, which is also linked to abnormalities in FER (e.g., Garrido-Vásquez et al., 2011). Canli and Lesch (2007) found that the gene encoding the serotonin transporter (5-HTT), which contains a regulatory variation that has been linked to anxiety-related traits and susceptibility for depression, has a high impact on behavior and may play a role in social cognition. Oxytocin might facilitate social information processing (comprehension of affective speech) not only in a healthy control group but also in those with autism (Hollander et al., 2007). Furthermore, mental illness and stress are interrelated (e.g., Pearlin, 1999), and stress-related changes in FER and ToM performance have already been identified in healthy individuals (see Chapter 1.4.1).

There are also changes in FER, ToM and other aspects of social cognition that seem to be specifically linked to particular mental disorders, such as *negative cognitive biases* in individuals with depression (e.g., Weightman et al., 2014) or *attributional bias* (i.e., the tendency to make systematic errors in evaluating or explaining one's own behavior or the behavior of others) in individuals with anxiety disorder, e.g., the tendency to misinterpret ambiguous social situations as threatening (Plana, Lavoie, Battaglia, & Achim, 2014). These cognitive biases take place at a later stage of processing (e.g., Garrido-Vásquez et al., 2011; Gotlib & Joormann, 2010). Yet, in other mental illnesses that are associated with impairments in general cognitive abilities at early stages of processing, such as schizophrenia, the abnormalities in basic FER are also found at early sensory stages of processing (e.g., Johnston, Stojanov, Devir, & Schall, 2005).

Although a large number of studies point to deficits in FER and ToM in individuals with mental illnesses, there is also a substantial number of inconsistent findings. In these, patients do not differ from the healthy control sample, or specific conditions - such as a specific symptom or other conditions - must be present for differences in FER and ToM to emerge (e.g., Garrido-Vásquez et al., 2011). Furthermore, there is sometimes even evidence that a certain group of patients is even better in a specific area of social cognition than the healthy control group. For example, Harmer, Grayson, and Goodwin (2002) found that euthymic patients with bipolar disorder performed better on disgust recognition than healthy controls. This reaffirms – as stated above – that social cognition is a multifactorial and highly complex composition of different skills, each of which is partly independent of the other and can thus interact differently with other factors.

## 1.5 FER and ToM in the context of affective disorders

The doctoral dissertation presented here focuses on the two ToM components according to Sabbagh (2004), namely *ToM decoding* and *ToM reasoning*, in individuals with *affective disorders,* specifically with *unipolar major depression* (UD) and *bipolar disorder* (BD). The presence of deficits or abnormalities in social cognition has been frequently reported in *persons with bipolar disorder* (BDs) as well as in *persons with unipolar major depression* (UDs). Basic FER and ToM (decoding and reasoning) are the most frequently investigated domains of social cognition in UDs and BDs (Bora & Berk, 2016; Weightman et al., 2014). But the findings are inconsistent. Therefore, in the following section, the empirical findings regarding basic FER, ToM decoding and ToM reasoning within BD and UD are reported. Although basic FER is not the research subject of this dissertation, an overview of it is also given. The reason for this is the overlapping of basic FER and complex FER (as part of ToM decoding) described above.

### 1.5.1 Inconsistent empirical findings

In the FER and ToM studies that examined BD patients, most of the samples examined were *remitted* BDs, whereas in most of the studies that investigated UD patients, the patients were in an *acute* phase of the disease. It is also noticeable that most of the studies used static materials and that there are hardly any studies that used dynamic materials, especially in the domain of ToM decoding. In the following, the results are reported separately by construct of interest, patient group and method. When reporting results, the following affective disorder states are distinguished: 1) *acute depressive or manic mood states of BD* (aBD), 2) *acute depressive mood state of UD* (aUD), 3) *remitted BD* (rBD), 4) *remitted UD* (rUD), 5) *healthy control group* (HC).

**Basic FER**

*rBD:* Most studies that have examined basic FER in BD used *static photographs of basic emotion expressions*. Using these kinds of prompts, remitted or subsyndromal BDs perform worse than HC in FER of *all basic emotions* (e.g., Bio, Soeiro de-Souza, Otaduy, Machado-Vieira, & Moreno, 2013; Lahera et al., 2015; Lahera et al., 2012; Neves et al., 2015; Soeiro-de-Souza, Otaduy, et al., 2012). However, there are also studies that found no difference in overall performance between rBDs and HC (e.g., Foland-Ross et al., 2012; J. Lee et al., 2013; J. L. Robinson et al., 2008; Rowland et al., 2013; Shamay-Tsoory, Harari, Szepsenwol, & Levkovitz, 2009; Vaskinn, Sundet, Friis, Simonsen, Birkenaes, et al., 2007). Yet some studies did not find any overall deficit in basic FER, but only in the recognition of individually *specific* basic emotions, e.g., a specific deficit only in the recognition of fear (Martino et al., 2011; Martino et al., 2008; Yurgelun-Todd et al., 2000), disgust (Martino et al., 2008) or happiness (Almeida, Versace, Hassel, Kupfer, & Phillips, 2010). Interestingly, in some studies, rBDs compared to HCs showed even *better performance* in recognizing specific basic emotions, e.g., a facilitated recognition of disgust (Harmer et al., 2002). There are only a few studies using *dynamic stimuli* (continuous morphs or short videos) to investigate basic FER in rBDs. Venn et al. (2004) found preserved basic FER performance of rBDs and no differences in sensitivity between rBDs and HC when using videos of continuous morphs. Rowland et al. (2013) also found no performance differences between rBDs and HCs when using short videos of actors expressing basic facial emotions (TASIT, Part 1). In contrast, Baez et al. (2013), who also used the TASIT, Part 1, found that a sample of BDs, most of them remitted, had a lower total score than HCs.

*aBD:* The studies that examined BDs in an acute state (manic or depressed) also found inconsistent results. When using *static photographs* of basic emotions, most studies found impaired basic FER of aBDs (e.g., Derntl, Seidel, Kryspin-Exner, Hasmann, & Dobmeier, 2009; Getz, Shear, & Strakowski, 2003; Lembke & Ketter, 2002; Soeiro-de-Souza, Bio, et al., 2012; Vederman et al., 2012) but some did not (e.g., Foland et al., 2008; Hulvershorn et al., 2012). Similarly to the findings in rBD, Gray et al. (2006) found preserved performance of aBDs (depressed and manic) compared to HCs when using *videos of continuous morphs*. Using this type of stimuli, Schaefer, Baumann, Rich, Luckenbaugh, and Zarate (2010) also found that in the case of fully developed emotion expression there was no difference between aBDs and HCs. However, aBDs required a more intense facial expression to make a first response and also to correctly identify the facial emotion compared to HCs. Furthermore, Summers, Papadopoulou, Bruno, Cipolotti, and Ron (2006) found a *specific* impairment in the depressed aBD group compared to HC in processing videos of continuous morphemes of surprise.

*aUD:* Within aUD, when using *static photographs* of basic emotions, some studies found that aBDs had impaired FER of all basic emotions (e.g., Csukly, Czobor, Szily, Takács, & Simon, 2009;

Langenecker et al., 2005; Naranjo et al., 2011; Rubinow & Post, 1992), whereas the results of other studies indicated no overall deficit of aUDs compared to HCs (Bediou et al., 2005; Gaebel & Wölwer, 1992; Gollan, Pane, McCloskey, & Coccaro, 2008; Seidel et al., 2010; Sprengelmeyer et al., 2011). Therefore, in aUD patients, the evidence for a general deficit in FER of all basic emotions is not particularly consistent and rather questionable. Somewhat clearer, however, is the evidence that there are abnormalities in recognition accuracy or *sensitivity* to quite *specific* basic emotions, and also that there is a *negativity bias* in the processing of emotional facial expressions. Deficits in the recognition of specific facial expressions have been reported, for example, for disgust (e.g., Douglas & Porter, 2010; Sprengelmeyer et al., 2011). Some studies have also found aUD to be worse than HC in FER either exclusively (e.g., Gollan, McCloskey, Hoxha, & Coccaro, 2010) or more clearly (e.g., Csukly et al., 2009) for *subtle* expressions of emotion. Furthermore, the results of the study by Leppänen, Milders, Bell, Terriere, and Hietanen (2004) revealed that aUDs and HCs were equally accurate at recognizing happy and sad faces. However, HCs also recognized neutral faces accurately as happy and sad faces, whereas aUDs recognized neutral faces less accurately than either happy or sad faces. The impairment in processing neutral faces was still present even after symptoms had remitted. It has also been observed in other studies that neutral faces were more likely to be interpreted as sad and less likely to be interpreted as happy in aUDs compared to HC (e.g., Douglas & Porter, 2010; Gollan et al., 2008; Naranjo et al., 2011). As in FER studies of BD patients, there are also findings for the UD patient group that – contrary to the findings indicating deficits – point to an advantage of aUD participants over HCs. For example, according to Milders, Bell, Platt, Serrano, and Runcie (2010) aUDs showed higher accuracy and a higher response bias than HCs for sad expressions only, which remained stable over a 6-month interval. *Dynamic* FER stimuli have also been used with UDs. Joormann and Gotlib (2006) used a dynamic face morph task and observed that aUDs and HCs were equally accurate. However, there is evidence of biases: aUDs required a significantly greater intensity of emotion than did HCs to correctly identify happy expressions and less intensity to identify sad than angry expressions. The results of Schaefer et al. (2010), who also used dynamic morphed stimuli, likewise indicate no significant differences between aUDs and HCs. In line with this, when using videotaped facial expressions of the six basic emotions displayed by professional actors (Kan, Kawamura, Hasegawa, Mochizuki, & Nakamura, 2002), Kan, Mimura, Kamijima, and Kawamura (2004) found no significant difference between aUDs and HCs. In contrast to this, Zwick and Wolkenstein (2017), who also used short video sequences displaying basic facial expressions, found aUDs to be worse in decoding happy faces compared to HC.

*rUD:* Data in rUD are particularly scarce and inconsistent. By using *static photographs* of basic emotions, female rUDs showed a selectively *greater* recognition of fear (Bhagwagar, Cowen, Goodwin, & Harmer, 2004) and sad faces (Biyik, Keskin, Oguz, Akdeniz, & Gonul, 2015). Another

study, which also tends to indicate *better* basic FER of rUDs compared to HCs is the one by Anderson et al. (2011). They investigated aUDs and rUDs by using static morphs for each of the 6 basic emotions. Results indicate that rUDs correctly identified a greater number of basic facial emotion expressions than did aUDs and HCs, owing to increased bias in identifying the presence of emotion, whereas aUDs had impaired accuracy. Investigating basic FER in rUD by using a *dynamic* morphing task, LeMoult, Joormann, Sherdell, Wright, and Gotlib (2009) found that HC made more errors than did rUDs. However, rUDs required a greater emotional intensity in the faces to correctly identify happy expressions. In summary, it seems that rUDs tend to have a better basic FER ability than aUDs and HCs (Bora & Berk, 2016).

**ToM decoding**

The most commonly used task to study ToM decoding in BD and UD patients is the RMET, wherein the stimuli consist of static photographs showing only the eye region of a face expressing a mental state.

*rBD:* By using the RMET, some found impaired overall ToM decoding performance in *remitted* (e.g., Bora, Bartholomeusz, & Pantelis, 2016; Bora et al., 2005; Budak, 2011) *or subsyndromal* BD (e.g., Cusi, MacQueen, & McKinnon, 2012; Donohoe et al., 2012), whereas a relatively large number of studies did not (e.g., Barrera, Vázquez, Tannenhaus, Lolich, & Herbst, 2013; Caletti et al., 2013; Duman et al., 2019; Ibanez et al., 2012; Martino et al., 2011; Purcell, Phillips, & Gruber, 2013; L. J. Robinson, 2010; Shamay-Tsoory, Harari, et al., 2009; Thaler, Allen, Sutton, Vertinski, & Ringdahl, 2013). Lahera et al. (2012) used the Emotion Recognition Test (Baron-Cohen, Wheelwright, Jolliffe, & Therese, 1997) in which patients were presented with static photos of 10 basic emotions (e.g., happiness or sadness) but also of 10 complex emotions (e.g., guilt, reflection, boredom). One third of the images shows the whole face, another third only the eyes, and the final third only the mouth. They found that globally, rBDs showed significant impairment in FER compared with HCs. By using a *dynamic* measure, namely the Frith-Happé animations, Malhi et al. (2008) observed impaired task performance in rBDs compared to HCs.

*aBD:* Again, for aBDs, studies are rather scarce. By using the RMET, most studies identified impaired overall ToM decoding performance in aBDs (e.g., Bora et al., 2016; Wiener, Andrzejewska, Bodnar, & Rybakowski, 2011).

*aUD:* For aUDs, there are studies that observed impaired RMET performance (e.g., Harkness, Washburn, Theriault, Lee, & Sabbagh, 2011; L. Lee, Harkness, Sabbagh, & Jacobson, 2005; Nejati, Zabihzadeh, Maleki, & Tehranchi, 2012; Wang, Wang, Chen, Zhu, & Wang, 2008) and there are also some studies that found no differences between aUDs and HCs (e.g., Kettle, O'Brien-Simpson, & Allen, 2008; Wolkenstein, Schönenberg, Schirm, & Hautzinger, 2011). There is also evidence for the presence of deficits for only *specific* emotions or the presence of abnormalities in only *certain rUD*

*subgroups*. Szily and Kéri (2009) divided their depressed sample into those with and those without psychosis risk. In addition to the overall RMET score, they also looked at the type of expression: social negative/positive (e.g., looking friendly or hostile) and cognitive (e.g., looking pensive). aUDs without psychosis risk only displayed impaired ToM decoding for negative social emotions, whereas patients with psychosis risk were also impaired in ToM decoding of cognitive expressions. Szanto et al. (2012) investigated a sample of aUDs divided into suicide attempers and non-suicidal depressed. They found only RMET impairment in aUDs who had attempted suicide. There are also studies which used the *dynamic* Frith-Happé animations, for example, Ladegaard, Larsen, Videbech, and Lysaker (2014), who found aUDs to be impaired in comparison to HCs.

*rUD:* Only a few studies have examined ToM decoding abilities in rUDs (Bora & Berk, 2016). By using RMET, there seemed to be no difference between rUDs and HCs in RMET total score (e.g., Purcell et al., 2013). Harkness, Jacobson, Duong, and Sabbagh (2010) investigated RMET performance of rUDs following a sad versus happy mood induction. They even found that rUDs where significantly *more accurate* in their ToM decoding than HCs. In addition, rUDs whose positive mood increased in response to the happy mood induction showed poorer task performance levels, similar to the never-depressed group. These findings are consistent with the results mentioned above reporting better basic FER performance in rUD patients.

**ToM reasoning**

*rBD:* Studies using *story comprehension, cartoon picture or picture sequence tasks* have found reasoning deficits in rBDs (e.g., Andrews, 2013; Bora et al., 2005; Ibanez et al., 2012; Inoue et al., 2004; Lahera et al., 2008; Martino et al., 2011; Olley et al., 2005; Sakarya & Ozgüven, 2012; Shamay-Tsoory, Harari, et al., 2009). And again, there are also some studies of rBDs that did not observe any differences compared to HCs (e.g., Caletti et al., 2013; Duman et al., 2019; Ioannidi, Konstantakopoulos, Sakkas, & Oulis, 2015; Kerr et al., 2003; Ozel-Kizil et al., 2012). Lahera et al. (2012) also discovered that globally, rBDs did not show any significant impairments compared with HCs. However, the subgroup of low-functioning BDs showed a significantly poorer performance compared with the subgroup of high-functioning BDs. There are also some studies that have examined ToM reasoning performance in *subsyndromal* BD samples. Simon et al. (2013) examined a BD sample, which was then further divided into remitted vs. subsyndromal BD patients. By using the Faux Pas Task, they found only the subsyndromal sample to be impaired in ToM reasoning. The other studies which investigated subsyndromal BD mostly found impaired ToM reasoning performance in rBD compared to HCs (e.g., Donohoe et al., 2012; Lahera et al., 2015; McKinnon, Cusi, & MacQueen, 2010; Van Rheenen & Rossell, 2013; Wolf, Brüne, & Assion, 2010) but some did not (e.g., Thaler et al., 2013). Ioannidi et al. (2015) applied a multi-level battery of ToM tasks to the subsyndromal BDs and HCs. They found BDs only to be impaired in the Faux Pas Task but not in the Hinting Task or in

False Belief (1st order). Results are also inconsistent when looking at studies using *dynamic, video-based material*. By using the MASC, Montag et al. (2010) and Santos et al. (2017) identified worse ToM reasoning performance in rBDs compared to HCs, whereas Aidelbaum and Goghari (2022) found no differences. By using the TASIT in investigating samples of subsyndromal BD, Rowland et al. (2013) found no difference in Part 2 but a significant impairment in Part 3, whereas J. Lee et al. (2013) found no difference in TASIT, Part 3.

*aBD:* By using *story comprehension, cartoon picture or picture sequence tasks*, some studies have found reasoning deficits in aBDs (e.g., Bora, Bartholomeusz, et al., 2016; Ioannidi et al., 2015; Kerr et al., 2003; Rossell & Van Rheenen, 2013; Wolf et al., 2010) and some have not (e.g., Bazin et al., 2009; Doody, Götz, Johnstone, Frith, & Cunningham Owens, 1998; Sarfati, Hardy-Baylé, Brunet, & Widlöcher, 1999). To the best of our knowledge, there is no study that has examined aBD patients with *dynamic* ToM material.

*aUD:* Results of most of the studies using *story comprehension, cartoon picture or picture sequence* tasks indicate reasoning deficits in aUDs (e.g., Mattern et al., 2015; Uekermann et al., 2008; Wang et al., 2008; Zobel et al., 2010), even in mildly depressd aUDs (Cusi, Nazarov, MacQueen, & McKinnon, 2013). In some studies, only a specific ToM deficit was found. For example, Thoma, Schmidt, Juckel, Norra, and Suchan (2015) found that aUD patients show an imparment in the interpretation of other peoples' sarcastic remarks but not of the mental states underlying other peoples' actions. On the other hand, some studies indicate equal ToM reasoning performances in aUDs compared with HCs (e.g., Bazin et al., 2009; Bertoux et al., 2012; Doody et al., 1998; Sarfati et al., 1999). There are two studies using *dynamic* measures, the MASC (Wolkenstein et al., 2011) and the TASIT, Part 2 (Ladegaard et al., 2014). Both studies found impaired ToM reasoning performance of aUDs compared to HCs. However, there are also two studies, using the MASC (Wilbertz, Brakemeier, Zobel, Härter, & Schramm, 2010) and the V-SIR (Bazin et al., 2009), which observerd no significant difference between aUDs and HCs in ToM reasoning.

*rUD:* There is a dearth of studies examining ToM in rUDs. There is a study by Inoue et al. (2004) that used *a cartoon picture story task* and indeed identified ToM reasoning deficits in rUDs.

**Summary**

For BD, deficits in basic FER have been reported in all three disease phases (depression, mania, remission) (Samamé, 2013). When it comes to ToM, Bora, Bartholomeusz, et al. (2016) concluded in their meta-analysis that ToM decoding and ToM reasoning performance (across different ToM tasks) is impaired in aBDs compared to HCs and is also evident in rBDs. Furthermore, ToM decoding and reasoning deficits are more severe during acute episodes compared to remission.

For UD, the longitudinal study by Milders et al. (2010) indicates that there are stable deficits in basic FER in aUDs, even when symptom severity decreases over a period of 6 month. Regarding

ToM, in another meta-analysis by Bora and Berk (2016), they found out that there are significant medium ToM decoding and ToM reasoning deficits in aUDs. Furthermore, they concluded from their analyses that there are deficits in affective as well as cognitive ToM-stimuli (see Chapter 1.2). Besides this, in UDs, deficits in ToM reasoning seem to be larger than deficits in basic FER. The latter appear to be only marginally impaired. In accordance with this, Dalili, Penton-Voak, Harmer, and Munafò (2015) reported very small effect size for basic FER deficits in UDs. Weightman et al. (2014) stated that deficits in basic FER and ToM in UD are more subtle than in other disorders and may not be identified by broad measures. Data for rUDs are sparse. Especially in basic FER and ToM decoding, rUDs seem to have a higher sensitivity and better performance in the respective tests.

Moreover, in general, in the studies using dynamic material, there is a larger proportion of studies finding no difference between patients and healthy participants than the proportion found in studies using static material. In their review, Garrido-Vásquez et al. (2011) state that not only healthy samples but also patient groups benefit from dynamic information, as dynamic stimuli have a *higher ecological validity* and therefore provide more information than static photographs. Differences between patients and HCs also seem to tend to decrease in the use of fully expressed, explicit and clear emotion expressions. Regarding these findings, the question rises as to whether deficits in basic FER and ToM performance of BDs and UDs are perhaps limited to material of low ecological validity (e.g., static photographs or stories) and relatively low emotional intensity (e.g., subtle or ambiguous emotional expressions).

### 1.5.2 Possible reasons for the inconsistent findings

Overall, the reasons for the inconsistent findings can be divided into two categories: First, the presence of *methodological shortcomings* in the previous studies, and second, the *heterogeneity of the disease* and the failure to account for *moderating* or *mediating variables*.

**Methodological shortcomings**

Authors of reviews and meta-analyses point to important methodological shortcomings in the studies that investigate basic FER and ToM in BDs (Bora, Bartholomeusz, et al., 2016; Samamé, 2013) as well as in UDs (Bora & Berk, 2016), such as poor between group matching on clinical/demographic variables, heterogeneous samples, inconsistent definition of euthymia and a low statistical power due to small sample sizes. Furthermore, Samamé (2013) criticize that most studies have assessed depressive symptoms by using the Hamilton Depression Rating Scale (Hamilton, 1960) which is claimed to be psychometrically and conceptually deficient (Bagby, Ryder, Schuller, & Marshall, 2004).

Another weakness concerns the *measurement instruments* used. A variety of different ToM tasks with different characteristics were used in the previous studies. As a result, differences in the

operationalization of social cognitive constructs, the difficulty, complexity, and psychometric properties of the tasks, as well as the type of stimuli used and the number of items included, may play a role in the differences among groups. For example, the *mode* (i.e., visual vs. verbal) and *complexity* (i.e., simple vs. advanced) of the ToM stimuli varied considerably. As a consequence, comparison of results between studies is also complicated (Bora & Berk, 2016; Samamé, 2013). This is problematic, as there is evidence that these features address separate neural networks (Schurz, Radua, Aichhorn, Richlan, & Perner, 2014) that theoretically could be more or less impaired with respect to the presence of mood symptoms and their severity (Aidelbaum & Goghari, 2022). Another very important point in this context, which is described in more detail in Chapter 1.3, is that most of the methods used – with the exception of the MASC – do not have sufficient *ecological validity*.

**Heterogeneity of the disease and moderating or mediating variables**

First, *demographic variables* related to FER or ToM performance, such as age, gender, and education, (see Chapter 1.4) should also be considered when examining patients. This is particularly important because there is evidence that the association between FER or ToM performance and another variable (e.g., gender) may not exist at all in a certain patient sample or may exist differently than in the healthy sample. For example, Vaskinn, Sundet, Friis, Simonsen, Birkenæs, et al. (2007) found that healthy male participants performed worse than their female counterparts in basic FER, whereas no gender difference was observed in BDs.

Bora and Berk (2016) also pointed out that *clinical variables* such as age of onset, duration of illness, comorbidity, number of past episodes, pharmacological variables and the presence of psychotic symptoms may be relevant to ToM deficits, but that studies addressing this issue are lacking. Some other variables are discussed in more detail in the following.

*The role of general neurocognitive functioning*

It is not yet clear whether and to what extent non-social cognitive deficits contribute to ToM deficits. The definition of social cognition has a large overlap with the definition of general neurocognitive functioning. Several authors suggest that BD and UD are often associated with dysfunctions in a number of general neurocognitive domains (e.g., Godard, Baruch, Grondin, & Lafleur, 2012; Haldane & Frangou, 2006). Impairments of attention, working memory, long-term memory and executive functions have been consistently reported for BDs, both during acute episodes (e.g., Kurtz & Gerraty, 2009) and during remission (e.g., Mann-Wrobel, Carreno, & Dickinson, 2011), although in milder form (Volkert et al., 2016). Based on these persistent deficits and the finding that not affected, first-degree relatives of BDs have similar impairments (Bora, Yücel, et al., 2009), Samamé (2013) argued that general cognitive dysfunction (including emotion recognition and ToM) may be a core feature of BD that reflects a strong genetic component. In contrast, for UD, McClintock, Husain, Greer, and Cullum (2010) reported that UD is inconsistently

associated with deficits in neurocognitive functions. Although cognitive deficits are also reported in UDs, they appear to be weaker in contrast to BDs (Mansell, Colom, & Scott, 2005).

There is some evidence that neurocognitive dysfunction partly explains the differences in ToM performance between patients and HCs (e.g., Lahera et al., 2008; Martino et al., 2011). Bora et al. (2005) even observed that the impairment in ToM decoding and reasoning in BDs was no longer significant after correction for working memory deficits. In line with this, Ioannidi et al. (2015) concluded that ToM reasoning deficits in BD (all three phases of remission, mania and depression) might reflect underlying cognitive deficits rather than represent a specific trait marker. On the other hand, Wolf et al. (2010) reported that the poorer performance of subsyndromal BDs' on executive tasks did not fully explain the differences in ToM reasoning between BDs and HCs, suggesting at least a partial selective deficit in ToM reasoning in BD. Therefore, Bora, Bartholomeusz, et al. (2016) suggest that further studies are needed to investigate the separability of neurocognition and social cognition.

In addition, it must also be noted that these conclusions are drawn from studies that used material with poor ecological validity. These conclusions may no longer be transferable if ecologically valid material is used.

*Influence of the current mood on basic FER and ToM*

One additional explanation for the inconsistent findings concerning social cognition in affective disorders is the presence (or absence) of an acutely present negative mood. The presumed effects would perhaps only become apparent if the test participants showed a negative mood in the examination situation. In recent years, there has been great progress in the experimental study of how affective states or moods affect the way people process social information (Forgas, 2017). According to Forgas (2017), psychological explanations for this phenomenon can be divided into three main theories that explain mood congruence: (1) *associative network theories* that emphasize memory processes (Bower, 1981; Bower & Forgas, 2000), (2) *affect-as-information theory* that relies on inferential processes (Clore, Gasper, & Garvin, 2001; Clore & Storbeck, 2006), and (3) an integrative *Affect Infusion Model* (*AIM*) (Forgas, 2002). These three models will now be explained in more detail. *The associative network theory:* The associative network theory by Bower (1981) proposes that moods are associated with an associative network of memory representations. Thus, a mood state can automatically activate representations that were associated with that mood in the past. Such affective *priming* has been demonstrated in several experiments with happy or sad people, who were more likely to remember mood-congruent details or events (Bower, 1981). In addition, mood congruence has also been observed in the interpretation of ongoing social behaviors (Forgas, Bower, & Krantz, 1984) and in the formation of impressions of others (Forgas & Bower, 1987). *Affect-as-information theory:* This model is closely aligned with research on misattribution and

judgment heuristics. It states that people misattribute a preexisting mood state as an indication of their response to an unrelated target. People rely on their mood as simple and convenient heuristic information mainly under the following condition: when "the task is of little personal relevance, when little other information is available, when the problems are too complex to be solved systematically, and when time or attentional resources are limited" (Fiedler, 2001, p. 175). The model is supported by studies showing, for example, that mood influenced by good or bad weather can significantly affect judgment on a variety of opinion questions in a telephone interview (Schwarz & Clore, 1983). Forgas (2017, p. 94) concluded that "respondents presumably had little time, interest, motivation, or capacity to engage in elaborate constructive processing in such a survey situation requiring rapid responses, and so relied on their mood as a simple and convenient heuristic shortcut to infer their evaluative reactions". *Affect Infusion Model:* The AIM (Forgas, 1995, 2002) defines affect infusion as a process where affectively charged information has an impact on attention, learning, decision making, and judgment processes, in that it can influence the cognitive processes of the judge, as well as the judgment itself. For affect infusion to occur, a relatively open and constructive information search is required. The AIM distinguishes four processing strategies depending on their openness, constructiveness and the degree of effort in search for a solution: 1) The *direct access* strategy is used when a topic is of low importance and has high familiarity, so that one simply and directly retrieves an already stored preexisting response. As this requires little effort and is not constructive, affect infusion should not occur. 2) The *motivated processing* strategy is dominated by a specific motivational objective. It is used when the topic is highly relevant and requires effortful, highly selective and targeted thinking. This open, constructive processing should be insensitive to affect infusions and may even produce mood incongruent effects. 3) *Heuristic processing* refers to the affect as information model described above. 4) Finally, *substantive* processing refers to the associative network theory (Affect Priming Model), described above.

In the domain of social cognition, this assumption that affective states may influence the content and valence of thinking and processing is supported by studies that show that even in HC, experimentally induced negative mood has a negative influence on emotion recognition (Chepenik, Cornew, & Farah, 2007), ToM reasoning (Feyerabend et al., 2018) as well as on the reaction of facial mimicry to emotional facial expressions (Likowski et al., 2011). Furthermore, it has already been shown that there is an interaction between the presence of a negative mood and the presence of an affective disorder: LeMoult et al. (2009) found that remitted patients with recurrent depression show a significantly lower sensitivity to happy faces under negative mood induction than people without a previous depressive illness (LeMoult et al., 2009). It has also been pointed out in other areas of social cognition (e.g, negative communication patterns; Rehman, Ginting, Karimiha, & Goodnight, 2010),

that certain effects in depressed persons can only be detected when a negative mood state is also acutely induced.

### 1.5.3 Performance in ToM as a predictor of relapse in affective disorders

Even though ToM appears to be impaired in affective disorders, there are very few studies investigating ToM performance as a potential predictor of the course of illness. Regarding basic FER, there is only one study by Bouhuys et al. (1999), who investigated emotion recognition in patients suffering from UD or BD. The authors used static, schematic line drawings of positive and negative emotional facial expressions in different developmental stages and found that the attribution of negative emotions to ambiguous faces was associated with relapse. The authors conclude that patients who show a negativity bias are more likely to relapse than patients who do not show a negativity bias. As to ToM reasoning, the study by Inoue, Yamada, and Kanba (2006) investigated the association between ToM reasoning and relapse in a sample of patients with rUD and rBD, by using a cartoon picture story task. Participants had to answer a second order false belief question. Patients who failed to answer this question correctly showed a significantly higher number of relapses in a 1-year follow-up than patients who answered correctly. In line with this, Yamada, Inoue, and Kanba (2015) evaluated ToM reasoning (Brüne cartoon picture story) in rUDs. After 1 year, UDs who had a ToM deficit according to the second-order false belief question relapsed more frequently. In addition, Purcell et al. (2013) examined ToM decoding (RMET) in rBDs. Although they found no group difference between rBDs and HCs in ToM decoding accuracy, they identified that rBDs responded significantly faster than HCs. Interestingly, faster response times in rBDs predicted greater impairment in life functioning at a one-year-follow-up. However, the tasks used in the studies described above are of questionable ecological validity. Therefore, it cannot be ruled out that these results are due to the use of tasks of low ecological validity.

## 1.6 Studies of the present work

The inconsistent results of studies concerning ToM in affective disorders may be due to methodological shortcomings, the failure to consider relevant third variables or the failure to use ecologically valid stimuli. The studies of the present work aim to counter some of these criticisms. In three studies, we investigated ToM decoding and ToM reasoning in patients with affective disorders, using ecologically valid stimulus material. We used the CAM to assess ToM decoding and MASC to examine ToM reasoning. Additionally, the effects of negative/sad *mood induction* (MI) are considered. Therefore, both tasks were divided into two parts, to conduct one part with and one without MI.

**Study 1: Theory of mind in acute and remitted unipolar depression: Patients struggle with reasoning task of high ecological validity**

This study aimed to investigate ToM decoding and ToM reasoning in aUD (*N* = 42) and rUD (*N* = 43) patients compared to HCs (*N* = 40), using ecologically valid stimulus material. By also investigating rUDs, the persistence of the deficits during remission was examined. In addition to a sad MI, we also assessed the effects of certain third variables that might have an impact on ToM performance, such as age and gender. Furthermore, besides investigating the accuracy in the CAM task, this study looked at how *difficult* the patients rated the CAM items and how *confident* they were in their judgments.

**Study 2: Theory of Mind in remitted bipolar disorder: Younger patients struggle in tasks of higher ecological validity**

By using the measures of high ecological validity mentioned above, the second study focuses on ToM decoding and ToM reasoning in rBD patients. A relatively large sample size of *N* = 44 rBD patients and *N* = 40 HCs was investigated. Besides the effects of sad MI, the impact of *age* and *gender* on ToM performance was also considered.

**Study 3: Theory of Mind in remitted bipolar disorder: Decoding predicts relapse**

This study investigated whether the performance on ToM decoding and ToM reasoning tasks as well as mood-linked changes in ToM performance predicted the clinical course of BD, that is, whether or not relapse occurred within a certain period of time after study participation. *N* = 40 rBD patients completed the CAM and the MASC. To assess mood-linked changes both tasks were divided into two parts, to conduct one part with, and one without, negative MI. The course of illness (9-month-follow-up-period) was assessed using the *Longitudinal Interval Follow-up Evaluation* (Keller et al., 1987).

# 2 Operationalization of the constructs of interest in this work

This chapter describes the methods used in our studies to measure the main constructs described above, i. e., a) ToM decoding ability, b) ToM reasoning ability and c) the course of illness and further d) how we induced negative mood.

## 2.1 Evaluation of ToM decoding using the Cambridge Mindreading Face-Voice Battery

The CAM tests the recognition of 20 complex emotions and mental states from faces and voices. In constructing the battery, the authors had the following goals: 1) to test different complex emotions and mental states and to investigate the recognition of specific emotions, and 2) to use dynamic stimuli with the purpose of generating a more naturalistic test. The test is based on the taxonomy of emotions introduced by Baron-Cohen et al. (2004) in which 412 emotions are classified into 24 mutually exclusive emotion groups. The CAM battery included a selection of 20 emotion concepts taken from this taxonomy, representing 18 of the 24 emotion groups. Overall, the CAM items include 5 positive concepts (empathic, exonerated, intimate, reassured, vibrant), 12 negative concepts (appalled, confronted, distaste, grave, guarded, insincere, mortified, resentful, stern, subdued, subservient, uneasy) and 3 neutral concepts (appealing, lured, nostalgic). Moreover, both subtle and intensive forms of emotion expressions are included. The battery includes two tasks: Emotion recognition in 1) the face (facial task) and 2) the voice (vocal task). In the present study, only the facial task was used and will be described hereafter. To create dynamic stimuli, 50 short film clips (3-5 seconds) were recorded in which actors of both sexes, different age groups and ethnicities enacted the emotions. After watching each clip, four adjectives are presented to the participant, who is asked to "choose the word that best describes how the person feels". In the present studies, the main measure of ToM decoding ability used was the *facial emotion recognition score* (FER score). This FER score is defined as the sum of all correctly answered items and ranges from 0 to 50. Depending on the study, additional scores from the CAM were used and described in the respective chapters.

## 2.2 Evaluation of ToM reasoning using the Movie for the Assessment of Social Cognition

The MASC was introduced by Dziobek et al. (2006) as a sensitive video-based test to evaluate subtle difficulties in mind reading. Participants are asked to watch a short movie (about 15 minutes)

and to answer multiple-choice questions related to the actors' mental states. Mental states involve emotions, thoughts or intentions of the characters portrayed, each of which can have a positive, negative or neutral valance. Items vary in difficulty and show a broad range of mental states. The range and quality of language, gestures and facial expressions used in the items were deliberately varied by the developers of the MASC and include several classic concepts of mental states: false beliefs (first and second order), deception, faux pas, manipulation, metaphor and sarcasm, among others. Furthermore, the items vary according to their conversational content. While some items represent verbal content (some to be understood literally and some symbolically), other items refer to nonverbal contents. Nonverbal items require consideration of facial expression, gestures, and body language. Therefore, one advantage of the MASC is its high ecological validity due to its approximation to the depiction of everyday social interaction situations. Compared to other ToM-paradigms, the MASC has shown the greatest discriminative ability to distinguish between individuals with and without Asperger's syndrome or schizophrenia respectively (Dziobek et al., 2006, Fleck, 2007).

The movie is about four characters meeting for a dinner party and consists of 45 items. Each item consists of a movie segment followed by a question about the mental state of a particular character in that segment along with 4 response options. Only one response option represents the correct answer, while the other three options appear as distractors. There is always one distractor representing an exaggerated mindreading because it goes beyond what is made accessible by the scene. Another distractor stands for a superficial mindreading, i.e., it is close to the correct answer, but too vague or taken word-for-word from the scene. The last distractor reflects no/missing mindreading, as it is based only on objective occurrences of the scene.

The *MASC total score* is calculated by finding the sum of the number of correctly selected response options and therefore ranges from 0 to 45. In addition, *error scores* can be formed separately for exaggerated, superficial, and absent mindreading by finding the sum of the number of selections of the respective distractor. Depending on the study, additional scores from the MASC have been used and described in the respective chapters.

## 2.3 Evaluation of the course of illness using the Longitudinal Interval Follow-up Evaluation

In Study 3, we conducted the *Longitudinal Interval Follow-up Evaluation* (LIFE; Keller et al., 1987) 9 months after the assessment of ToM decoding and ToM reasoning performance of rBDs. LIFE is a semi-structured interview and rating system. It is used to record retrospectively the course of psychiatric illnesses (e.g., recovery from previous episodes or occurrence of new episodes) and its

concomitant circumstances since a defined point in time, in this case the time of the first examination session to determine ToM performance.

Using this diagnostic tool, the interviewer assigns *Psychiatric Status Ratings* (PSRs) for each mental illness of interest. In the present study, weekly ratings for depressive and (hypo-)manic symptoms were used, resulting in 2 x 39 PSRs for a rBD participant. The PSRs comprise 6 levels that indicate the extent to which the symptomatology meets DMS-IV criteria and thus reflect the severity of the illness. PSRs of 5 and 6 represent the presence of full DMS-IV criteria for either a depressive or manic episode, 3 and 4 indicate states of partial impairment, and 1 or 2 reflect no or very little symptomatology. Several studies suggest that the LIFE is a valid and reliable method (Warshaw, Dyck, Allsworth, Stout, & Keller, 2001; Warshaw, Keller, & Stout, 1994).

In the present study, data obtained with the LIFE interview formed the basis for an index of the course of the illness, namely whether or not a patient had *relapsed* (two-factorial relapse variable; yes versus no).

## 2.4 Induction of a sad mood

To induce a sad mood we applied a 10-minute and a 5 minute version of Peer Gynt Suite No. 1 Op. 46 “The Death of Ase” from Grieg, which is known to be suitable for inducing a sad mood (Rojas, Geissner, & Hautzinger, 2014). While listening to the piece of music, participants were asked to think of a sad event in their life that had happened at least two years ago (to avoid destabilization). This combination of sad music and reflection on a personal experience has proven to be particularly effective in inducing a sad mood (Kuehner, Huffziger, & Liebsch, 2009). It is assumed that the mood induction (through music) causes a change in experienced affective processes and that this induced mood can have an influence on cognitive processes (Västfjäll, 2001).

# 3 Empirical Papers and Manuscripts

All studies were improved by the local ethics committee and were carried out in accordance with the provisions of the World Medical Association Declaration of Helsinki.

## 3.1 Study 1: Theory of mind in acute and remitted unipolar depression: Patients struggle with reasoning task of high ecological validity

Research addressing ToM in UD (aUD and rUD) has yielded inconclusive results. As described in detail in previous chapters, this can be attributed to methodological shortcomings, failure to account for relevant third variables, and the scarcity of studies using ecologically valid stimuli. This study investigated ToM decoding and ToM reasoning in aUDs and rUDs using ecologically valid stimuli and considering the effects of a sad MI. We also examined the influence of demographic variables like gender and age on ToM in patients and HCs. In addition, this study looked at how difficult patients found the ToM decoding task and how confident they were in their judgment.

### 3.1.1 Introduction

Numerous studies record losses in the social functional level of patients with UD (e.g., Berecz et al., 2016; Hirschfeld et al., 2000; Judd et al., 2000) that persists even during remission (Coryell et al., 1993; Ladegaard, Videbech, Lysaker, & Larsen, 2016). One explanation for these losses is the presence of abnormalities in social cognition (e.g., Berecz et al., 2016; Weightman et al., 2014). Among the most frequently investigated domains of social cognition is ToM (Bora & Berk, 2016; Weightman et al., 2014), subdivided into a) *decoding* and b) *reasoning* about others' mental states (Adolphs, 2009; Ladegaard et al., 2016; Sabbagh, 2004) (see Chapter 1.2 for detailed information).

One operationalization of ToM decoding is *complex FER*. However, FER has already been investigated in several studies in people with UD but mostly by using standardized, static images of facial expressions that reflect the *basic* emotions (basic FER), such as the stimuli created by Ekman and Friesen (1976). For aUD as well as rUD, findings are inconsistent and are summarized in Chapter 1.5.1. In brief, in many cases, an impaired FER of aUD compared to HC was observed – partly as a fundamental impairment in the recognition of all basic emotions, partly as a specific impairment regarding individual basic emotions. To check whether the abnormalities in FER could be a vulnerability factor, studies were also conducted with rUD patients. Most of them also indicated impairments in rUD. Furthermore, a negativity bias has been repeatedly demonstrated, indicating that UDs misinterpreted neutral faces as sad, and happy faces as less happy/neutral. In contrast,

however, most studies using dynamic stimuli of basic emotions found no deficit in basic FER in UDs compared to HCs, except the study by Zwick and Wolkenstein (2017) (see Chapter 1.5.1).

It is noticeable that deficits are more likely to be found in static than in dynamic image material and more likely to be found in morphed than in fully developed material. Accordingly, the question arises whether the deficit may be limited to stimuli with comparatively low information content. In everyday life, we are rarely confronted with prototypical facial expressions of full intensity. Therefore, UD patients could be impaired in everyday life when it comes to the recognition of more subtle and complex facial expressions. Furthermore, it should be noted that ToM decoding in the narrower sense does not refer to the recognition of basic emotions but rather to the recognition of complex emotions (Damasio, 2014). There has been little research in this area so far. The few studies generally used the RMET and yielded inconsistent results (see Chapter 1.5.1). However, in everyday life, we usually have not only the eye area of our counterpart (like in RMET), but the entire, moving facial expression (including mouth, cheeks, nose) at our disposal to decode emotional states. Consequently, one of the main objectives of this study is to examine the extent to which patients with aUD and rUD show impairments in ToM decoding when ecologically valid stimuli are used. Ecologically valid stimuli are defined as stimuli that a) are dynamic and b) display the whole face of real persons who in turn display c) complex emotional expressions in d) different developmental stages (ranging from subtle to fully developed).

As Zwick and Wolkenstein (2017) mentioned, ToM studies have focused on accuracy and biases and ignored other aspects that might also matter for the quality of social interaction. One example of these neglected aspects is how *confident* people are about their judgments and how *difficult* they find it to recognize certain facial expressions. Therefore, in this study, for the first time, we also collected *confidence* and *perceived simplicity* ratings in a ToM decoding task of high ecological validity.

Up to now, ToM reasoning abilities of UD patients have often been investigated by means of story comprehension tasks or cartoon comprehension tasks - such as the *Advanced Theory of Mind Test* by Happé (1994) as well as the *picture sequence tasks* (see Chapter 1.3). Although reasoning deficits were found in UD patients using the paradigms described above, results are inconsistent (see Chapter 1.5.1) and the ecological validity and the suitability of such paradigms to detect potentially weaker ToM deficits is questionable (see Chapter 1.3). Therefore, the question remains to what extent reasoning deficits in patients with UD can also be found with ecologically valid measuring instruments. In the meantime, there are some studies which address this issue by capturing the reasoning ability with the MASC or the TASIT. Wilbertz et al. (2010) found no reasoning impairment with MASC in depressed patients. This could be due to the rather restrictive sample of chronically depressed patients, but also to the fact that the healthy sample in this study was not examined for

potentially present psychopathologies. By using the same task, Wolkenstein et al. (2011) only found a trend indicating decreased reasoning ability in aUD, maybe due to the low sample size. Zwick and Wolkenstein (2017) who also used the MASC with a larger sample size, found impaired reasoning in aUD but not in rUD. Furthermore, Ladegaard et al. (2014) examined ToM reasoning in medication-naïve first-episode depressed patients by using the TASIT and found depressed patients performing below the controls in the subscale "paradoxical sarcasm". Hence, there are a wide variety of view points on existing reasoning deficits. A further aim of the present study is therefore to provide further evidence for ToM reasoning deficits in UD when using the MASC as ecologically valid material.

In addition, we aim to extend previous research by investigating whether the presence of a *sad mood* has a moderating influence on the presence of a ToM decoding and ToM reasoning deficit within aUD and rUD. The presence of a negative mood in the acute investigation situation has been repeatedly discussed as a moderating factor, and there are several studies (see Chapter 1.5.2: *Influence of the current mood on basic FER and ToM*) that collectively suggest that the presence of an acute sad mood may contribute significantly to whether or not certain effects can be proven.

This study additionally considers *third variables* that may also play a role in the inconsistent results of previous studies. Indeed, it has become evident that associations between ToM and a third variable (e.g., gender) present in non-clinical samples may be different, or even absent, in patient samples (Vaskinn, Sundet, Friis, Simonsen, Birkenaes, et al., 2007) (see Chapter 1.5.2: *Heterogeneity of the disease and moderating or mediating variables*). Besides gender, another variable that might be of interest in this context is age. Some studies have investigated the change in FER performance across age groups in non-clinical samples and found a significant decrease with age (Calder et al., 2003; Mill, Allik, Realo, & Valk, 2009; Phillips, MacLean, & Allen, 2002; Richter, Dietzel, & Kunzmann, 2011; Sullivan, Ruffman, & Hutton, 2007). However, it is not yet clear whether this relation also exists in UD. This is problematic because the probability of overlooking differences between UD and HCs would be relatively high – even when matching for age – if the relationship between gender or age and ToM was different in UDs compared to HCs. Therefore, besides matching for age and gender, the current study explored whether the relation between these variables and ToM is the same for aUDs, rUDs and HCs.

The primary aim of this study is to investigate the two ToM components – decoding (FER) and reasoning – in patients with aUD using ecologically valid stimulus material. Furthermore, the persistence of the deficits will be investigated using a rUD sample. In addition, the influence of a sad MI is investigated. In summary, in the present study we made the following hypotheses:

1) Compared to HC, aUD and rUD show deficits in complex FER (ToM decoding) as well as in ToM reasoning, if ecologically valid stimulus material is used to determine the ToM components.
2) These deficits are stronger if a negative mood is induced beforehand.
3) UD patients show deficits in decoding and reasoning even during remission. These deficits should be weaker than in the acute depressive state. Here, too, the deficits are expected to be intensified by negative mood induction.
4) It is assumed that a significant variance in reasoning is due to the ability to decode complex emotional states.

We also wanted to explore the confidence and perceived simplicity ratings of the ToM decoding task. Furthermore, we wanted to explore the influence of age and gender on ToM decoding and ToM reasoning comparing aUD, rUD and HCs.

## 3.1.2 Methods

### 3.1.2.1 Sample

*N* = 125 participants (aUD: *N* = 42; rUD: *N* = 43; HC: *N* = 40) were recruited by advertisements in inpatient and outpatient clinics as well as on the internet.

*Inclusion criteria* for the patient groups were: a) for aUD: acute episode of major depression, for rUD: lifetime major depressive disorder (single episode or recurrent) but currently remitted according to the *Diagnostic and Statistical Manual of Mental Disorders (DSM IV; American Psychiatric Association, 1994)*; b) for rUD: currently remitted as defined by a *Hamilton Depression Rating Scale* score below 8 (HDRS17; Hamilton, 1967) and an *Inventory of Depressive Symptomatology* score below 12 (IDS-C30; Rush, Gullion, Basco, Jarrett, & Trivedi, 1996), for aUD: a HDRS17 score above 13 and IDS-C30 score above 23. Healthy participants were included if there was no evidence of current or lifetime mental disorder.

*Exclusion criteria* for all participants were: a) insufficient knowledge of the German language, b) diseases affecting the central nervous system, c) age below 18 or above 69 years, and d) neurocognitive impairments as defined by an IQ-score below 85 according to the *Multiple Choice Word Fluency Test* (MWT-B; Lehrl, 2005) or extreme outlier downwards (< 3 * interquartile range) in either the *Verbal Learning and Memory Test* (VLMT; Helmstaedter, Lendt, & Lux, 2001) or the *Trail Making Test* (TMT; Reitan, 1992). Of the original 129 participants, four (aUD: *N* = 2; rUD: *N* = 2) were excluded due to abnormal scores below the standard on the neuropsychological tests, resulting in the sample size described above.

Further exclusion criteria for the patient groups were: a) current alcohol or substance abuse and/or dependency or lifetime dependency if abstinence period < 3 years, b) acute or lifetime

psychotic symptoms except mood congruent delusions during affective episodes, c) current anorexia nervosa (Body Mass Index ≤ 18 kg/m²), d) comorbid schizoid, schizotypical, paranoid, antisocial and/or borderline personality disorder according to DSM IV (Saß, Wittchen, & Zaudig, 1996).

The *Structured Clinical Interview for DSM-IV* (SCID-I and SCID-II; Wittchen, Zaudig, & Fydrich, 1997) was conducted by trained interviewers in all three groups (aUD, rUD and HC). The patient and HC groups were matched with regard to gender, level of education, and age (+/- 5 years).

Table 1A shows the sample characteristics. Of the clinical sample, 61.91% of the aUD sample and 55.82% of the rUD sample had one or more comorbid mental illness(es) [$\chi^2(1)=0.12$, p=.726], and 85.71% of the aUD sample and 48.84% of the rUD sample were taking medication at the time of testing [$\chi^2(1)=11.46$, p=.001].

#### 3.1.2.2 Materials & Procedure

*Assessment of symptoms*

To determine both the HDRS17 score and the IDS-C30 score within the clinical groups we used the *Structured Interview Guide for the Hamilton Depression Scale and Inventory of Depressive Symptomatology* (SIGHD-IDS; Kobak, Williams, & Rush, 2007). Additionally, to ensure the absence of symptoms in the rUD and HC groups or the presence of depressive symptoms in the aUD group, a self-report measure were also applied: The *Beck Depression Inventory* (BDI; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961).

*Assessment of neurocognitive functioning*

For controlling significant neurocognitive impairments, we applied the VLMT, the TMT-A and -B and the MWT-B. Table 1B & C presents the clinical and neurocognitive variables separately for aUDs, rUDs and HCs and shows significant group differences for some of them. The neuropsychological tests are described in more detail in the following:

VLMT: The VLMT is the translation and further development of the *Auditory Verbal Learning Test* (AVLT) (Lezak, 1976, 1983, 1995). It can be used to test individuals aged 6 to 79. The administration of the VLMT is divided into 5 phases. The first phase is a learning phase: In five passes, a *learning list* of 15 words is read to the participants by the test leader. After each pass (pass 1 - pass 5), the participant is asked to recite all the words of the learning list that he or she is able to remember. The order of reciting is irrelevant (free reproduction). The second phase includes the reading of an *inference list*, also consisting of 15 words, and the subsequently free recall of it by the participant. In the third phase, the participant is asked to recall the learning list without repeated reading (pass 6). In the fourth phase, after a time delay of 20-30 minute and without repeated reading, the participant is again requested to recall the learning list (pass 7). In the fifth phase, the test leader examines which words of the learning list are recognized by the participant when a list of

50 words, containing all of the learning list's words, all of the interference list's words, and 20 additional words, is read to the participant. The actual testing time lies between 20 and 25 minutes, and between 50 and 55 minutes when the time delay is included. According to several previous analyses (e.g., Vakil & Blachstein, 1993), all recorded parameters can be represented by 3 factors: First, learning or data acquisition, calculated by finding the sum of all correctly recalled words in pass 1 to 5. Second, consolidation in long-term memory, resulting from subtracting the number of correctly recalled words in pass 7 from the number of correctly recalled words in pass 5. Third, recognition performance, examined within the fifth phase. Therefore, we considered the following parameters in our study: 1) *short-term verbal retentiveness* (summation of all correctly recalled words in pass 1 – pass 5), 2) *memory loss after delayed recall* (after a 20-30 minute time lag) 3) and recognition performance after a time lag corrected by the number of false positive items in the recognition test (*corrected recognition*). Test-retest reliability after 8-12 months ranged from $r$ = .68 to $r$ = .86 (Helmstaedter & Durwen, 1990).

TMT-A/B: The TMT is used to test visual attention and task switching and provides information about visual search speed, scanning, processing speed, mental flexibility, and executive functioning (Arnett & Labovitz, 1995). There are two parts, A and B, which can be conducted sequentially. In both parts, participants are asked to connect a sequence of 25 consecutive targets on a sheet of paper as quickly as possible while still maintaining accuracy. In part A (*TMT-A*), the participants must connect a sequence of scattered numbers from 1 to 25 in the correct order (from 1 to 2 to 3 etc., up to 25). In part B (*TMT-B*), a working memory component is added: alternately, numbers and letters are to be connected in "corresponding" order, i.e., the numbers in ascending order and the letters according to the alphabet (from 1 to A, to 2, to B, to 3, to C etc., up to number 13). During the test performance, the experimenter immediately points out errors and requests the participant to correct them. Thus, the occurrence of errors increases the duration of test execution. Test scores are the *execution times in seconds* for each part. Part A primarily examines visual-perceptual abilities, particularly processing speed. Part B is designed to examine executive functions (mainly working memory) and task-switching ability (Sánchez-Cubillo et al., 2009). Performance decreases with increasing age and lower levels of education (Tombaugh, 2004). For intervals of 3 weeks to 1 year, test-retest reliability is moderate to high for part A ($r$ =.36 to .79) and part B ($r$ =.44 to .89) (Bornstein, Baker, & Douglass, 1987; Dikmen, Heaton, Grant, & Temkin, 1999; Matarazzo, Wiens, Matarazzo, & Goldstein, 1974).

MWT-B: MWT is the generic term for several tests which are constructed according to the same pattern. The MWT versions are performance tests for measuring the general level of crystalline intelligence. Since only low demands are made on the actual available power, scores are hardly influenced by mild to moderate mental disorders. Therefore, the MWT versions can be used to

estimate the premorbid intelligence levels. The most common version is the MWT-B, which we also used in the present work. It consists of 37 rows (itmes) with 5 words each. For each item, participants are asked to distinguish the one really existing word from the 4 fictitious, newly constructed words. The 37 items are ordered by difficulty and there is no time limit for completion. The test score is calculated by determining the number of correct selections of the really existing word, so the score can range from 0 to 37. Repeated measurements revealed high retest reliabilities of $r$ = .95 for 6 months and $r$ = .87 for 14 months. There are also conversion tables for converting raw values into percentile ranks, IQ scores, and standard values. These norms are based upon a representative sample of $N$ = 1952 German speaking adults aged 20 to 64 years. Both raw values as well as IQ scores are used in the present study.

*Assessment of ToM decoding*

The face task of the CAM was used to asses ToM decoding (see Chapter 2.1). We subdivided the CAM into two parts to present one part with and the other part without sad MI: one part consisted of 22 and the other of 23 film clips. Items were assigned to the both parts by matching them with regard to difficulty, valence and emotion concept (Dolde, 2012). The remaining 5 film clips showed a poor discriminatory power. We used them as practice trials at the beginning of the session. Film clips were presented in a randomized order.

As an index of FER performance, we used the *CAM proportion scores* ($CAM_{prop}$), defined as "number of correct answers in the respective CAM part divided by the total number of items in the respective CAM part". Furthermore, we calculated the *percentage of the positivity bias* ($CAM_{positivity}$) and *negativity bias* ($CAM_{negativity}$). $CAM_{positivity}$ (or $CAM_{negativity}$) was calculated based on the number of incorrectly chosen distractors that – according to a preliminary study (Choudhery, 2012) – were more positive (or more negative) than the correct answer, divided by the total number of items in the respective CAM part. This quotient was multiplied by 100. Moreover, in accordance with, e.g., Wolkenstein et al. (2011), we calculated separate proportion scores for items representing *negative* ($CAM\text{-}Neg_{prop}$), *positive* ($CAM\text{-}Pos_{prop}$) and *neutral* ($CAM\text{-}Neut_{prop}$) mental states. Concepts were coded for valence when creating the taxonomy by 3 independent judges (Baron-Cohen, Golan, Hill, & Wheelwright, n.d.).

Additionally, after each clip, participants were presented two *visual analogue scales* (VAS). The VAS is a line whose endpoints represent extreme states, such as "very unconfident" and "very confident". The respondent marks his or her subjective feeling or assessment by placing a mark on the line. The indicated value is then quantified as a percentage (%) from 0 to 100. However, this continuous scale was not visible to the respondent. The first VAS was present to participants to indicate how confident they were in their judgment (left endpoint = *very unconfident*, right endpoint = *very confident*) and the second VAS to indicate how easy/difficult the expression was for them to

judge (left endpoint = *very difficult*, right endpoint = *very simple*). *Confidence scores* (Con) and *simplicity scores* (Simp) were averaged across all trials.

**Table 1**

*Demographic, clinical and neurocognitive variables*

| **Variable** | **aUD** ***N* = 42** | **rUD** ***N* = 43** | **HC** ***N* = 40** | ***df*** | ***χ²*** | ***p*** |
|---|---|---|---|---|---|---|
| *A) Demographic variables* | | | | | | |
| Level of education (Abitur) in % | 61.90 | 69.77 | 77.5 | 2 | 1.73 | .42 |
| Gender (female) in % | 76.19 | 86.05 | 77.5 | 2 | 1.52 | .47 |
| Current living condition in % | | | | 6 | 11.46 | .07 |
| living alone | 45.24 | 41.86 | 25.00 | | | |
| living together with life partner | 30.95 | 51.17 | 62.5 | | | |
| living together with parents/relatives | 7.14 | 2.33 | 2.50 | | | |
| Other | 9.52 | 4.65 | 17.5 | | | |
| Current social contact in % | | | | 10 | 12.14 | .26 |
| more than 1 time per week | 50.00 | 55.82 | 77.5 | | | |
| not more than 1 time per week | 19.05 | 18.60 | 15.00 | | | |
| once in two weeks | 11.90 | 6.98 | 5.00 | | | |
| once per month | 9.52 | 13.95 | 2.50 | | | |
| only in hallway or at work | 7.14 | 4.65 | 0.00 | | | |
| | **M (SD)** | **M (SD)** | **M (SD)** | ***df*** | ***F*** | ***p*** |
| Age | 40.43 (15.01) | 45.60 (14.73) | 42.35 (14.21) | 2 | 1.354 | .26 |
| *B) Clinical variables* | | | | | | |
| HDRS17-Score | 18.05 (3.25) | 2.19 (2.82) | - | 80.81 | -23.99 | < .001 |
| IDS-C30-Score | 35.14 (7.77) | 3.91 (4.74) | - | 67.58 | -22.31 | < .001 |
| BDI-Score | 27.32 (7.84) | 5.48 (6.35) | 1.53 (2.46) | 2 | 217.4 | < .001[1] |
| *C) Neurocognitive variables* | | | | | | |
| MWT-B (IQ-score) | 115.19 (16.32) | 114.41 (15.94) | 118.90 (15.81) | 2 | 0.923 | .400 |
| TMT-A (seconds) | 28.81 (9.97) | 28.75 (9.65) | 26.70 (8.31) | 2 | 0.674 | .512 |
| TMT-B (seconds) | 65.33 (22.57) | 69.93 (24.32) | 56.83 (17.15) | 2 | 3.91 | .023[2] |
| *VLMT* | | | | | | |
| short-term verbal retentiveness | 59.02 (8.90) | 58.86 (8.38) | 60.43 (7.39) | 2 | 0.44 | .643 |
| memory loss after delayed recall | 8.49 (2.68) | 8.50 (2.19) | 8.88 (2.11) | 2 | 0.46 | .633 |
| corrected recognition | 14.45 (0.89) | 14.61 (0.87) | 14.53 (0.75) | 2 | 0.40 | .673 |

*Note*. aUD = patients with acute unipolar depressive disorder, rUD = patients with remitted unipolar depressive disorder, HC = healthy control group, HDRS17 = Hamilton Depression Rating Scale, IDS-C30 = Inventory of Depressive Symptomatology, BDI = Beck Depression Inventory, MWT-B = Multiple Choice Word Fluency Test, TMT = Trail Making Test, VLMT = Verbal Learning and Memory Test

[1] t.tests between all three groups are significant

[2] t.tests between aUD and HC and between rUD and HC are significant

*Assessment of ToM reasoning*

For the evaluation of subtle difficulties in mindreading, participants are presented to the MASC (see Chapter 2.2). For the present study we divided the MASC into two parts. The first part consisted of the first 23 items, the second part of the remaining 22 items. The first part was repeated as an introduction in order to avoid memory effects in the second test session - but this time without breaks and questions. For each part, we calculated *proportion scores* ($MASC_{prop}$). Moreover, in accordance with the study by Montag et al. (2010), we calculated proportion scores for two different *mental state modalities*: *cognitive* ($MASC_{cog}$): "What is X thinking or intending", and *emotional* ($MASC_{emo}$): "What is X feeling?". If the correct answer was not given, we also recorded the type of mistake: *no ToM* ($MASC_{no}$), *less ToM* ($MASC_{less}$) or *exceeded ToM* ($MASC_{exceeded}$).

*Negative mood induction*

We used a 10-minute and a 5-minute version of The Peer Gynt suite no. 1 op. 46 "The Death of Ase" from Grieg to present it to the participants who, during this time, should think of a sad event in their lives that have happened at least two years ago (to avoid destabilization) (see Chapter 2.4).

*Positive and Negative Affect Schedule (PANAS)*

We repeatedly used 5 items of the *Positive and Negative Affect Schedule* (PANAS; Crawford & Henry, 2004) to verify successful MI: *worried*, *elated*, *bad*, *content*, *sad.* On a 5-point Likert scale ranging from *"very slightly"* to *"very much"*, participants should indicate the extent to which they experience these emotions at the present time.

*Procedure*

Participation in this study commenced with informed consent at the beginning of the diagnostic session (t0) in which the anamnestic interview, the clinical assessments and the neuropsychological tests were subsequently performed[3]. Within two weeks after t0, two test sessions (t1 & t2) were performed on separate days. If there was a period of more than two weeks between t0 and t1 and/or t2, we re-evaluated SIGHD-IDS and BDI to guarantee – depending on group membership – sustained acute depression, remission or health.

In both test sessions one part of the CAM was presented first, followed by one part of the MASC. The allocation of the respective CAM part and the MI condition (yes versus no) to t1 and t2 was completely balanced and randomly assigned to the participants. In the MI session the piece of music was presented directly before the CAM (10-minute version) and again before the MASC (5-

[3] Besides the material described in this section we used additional questionnaires in this study that are, however, not relevant to the question that is examined here. Participants were asked to complete the following questionnaires at home: Inventory of Interpersonal Problems (Horowitz, Strauß, & Kordy, 2000); Social support questionnaire (Fydrich, Sommer, & Brähler, 2007); Cognitive Emotion Regulation Questionnaire (Garnefski & Kraaij, 2007); Emotion regulation questionnaire (Gross & John, 2003); Empathy Quotient (Baron-Cohen & Wheelwright, 2004). The respective results are reported elsewhere.

minute version). The PANAS items were completed immediately before and after the MIs. In the session without MI, the PANAS was given both before the CAM and before and after the MASC. Participants were requested to sit on a fixed chair because of concomitant eye tracking recordings[4], and during CAM, electromyography recordings[5] made it necessary to attach electrodes to the participants' face. During the MASC, we then removed the electrodes because no electromyography or eye tracking recordings were made during that time.

#### 3.1.2.3 Statistical analyses

We accomplished the statistical analyses with R, version 4. 1. 2 (R Core Team, 2021). The level of significance was set to $\alpha < .05$.

*Pre-Analyses*: To decide whether age and gender should be included as additional factors in the main analyses, we first analyzed whether accuracy in ToM decoding ($CAM_{prop}$) and ToM reasoning ($MASC_{prop}$), as a function of group assignment (aUD vs. rUD vs. HC), was related to age or gender, respectively. To test that for age, we conducted *Analyses of Covariance* (ANCOVA) with $CAM_{prop}$ and $MASC_{prop}$ as dependent variables and group and age as independent variables. To test that for gender, we conducted separate *Analyses of Variance* (ANOVA) with gender and group as between-subject factors for each of the dependent variables. We planned to include age or gender in the main analysis, if there were a significant interaction between group and age/gender.

For the *main analyses*, we conducted *mixed-factor ANOVAs* and *mixed-factor ANCOVAs* with $CAM_{prop}$, $CAM_{negativity}$ and $CAM_{positivity}$, CAM-$Neg_{prop}$, CAM-$Pos_{prop}$, CAM-$Neut_{prop}$, $MASC_{prop}$, $MASC_{cog}$ and $MASC_{emo}$, $MASC_{no}$, $MASC_{less}$, $MASC_{exceeded}$ respectively as dependent variables. The factor *group* was defined as a between-subjects factor and sad MI (with versus without) as within-subject factor.

To explore the confidence ratings (Con) and the simplicity ratings (Simp), we performed two different kinds of analyses. First, we wanted to investigate if the three different groups differed by their confident and simplicity ratings, depending on whether a mood induction occurred or not. Therefore, we conducted ANOVAs with group as between subject factor, MI as within subject factor, and Con or Simp as dependent variable. Second, we wanted to find out to what extent performance in the tasks was related to the confident and simplicity ratings, as a function of group membership, and whether mood induction occurred. For this purpose, we performed ANCOVAS with $CAM_{prop}$ as dependent variable and group, MI and Con or Simp as predictors.

Post-hoc tests were performed when needed. For all analyses, we centered the *age* variable. We report *generalized eta-squared* ($\eta_G^2$) for the ANOVAs and ANCOVAs, as indicators of effect size (Olejnik & Algina, 2003). $\eta_G^2$ is calculated by the sum of squares of an effect for one variable divided by the total sum of squares in the ANOVA model. Values between 0.01-0.05 reflect small effect sizes,

[4] This is outside the scope of this work, and the respective results are reported elsewhere

[5] This is outside the scope of this work, and the respective results are reported elsewhere

values between 0.06-0.13 medium effect sizes, and values ≥ 0.14 large effect sizes. For the t-tests we report *Cohen's d (d)*. *d* is determined by calculating the mean difference between two groups, and then dividing the result by the pooled standard deviation. For *d*, values between 0.2-0.5 correspond to small effect sizes, values between 0.5-0.8 to medium effect sizes, and values ≥0.8 to large effect sizes.

### 3.1.3 Results

*Manipulation check of mood induction*

We performed a *multivariate ANOVA* (MANOVA). The four-level *time* factor (pre_1 = *before the first MI*, post_1 = *after the first MI*, pre_2 = *before the second MI*, post_2 = *after the second MI*) was the within-subject factor. The *group* (aUD vs. rUD vs. HC) factor represented the between-subject factor and the item scores (worried, elated, bad, content, and sad) were the dependent variables. Means and standard deviations are presented separately for aUD, rUD and HCs in Table 2.

We found the expected main effect of time [$V$=.375, $F$(3,492)=14.02, $p$<.001, $\eta_G^2$=.125]. Mean scores before the first MI were significantly different from mean scores after the first MI in the expected direction [worried: $t$(233.77)=-5.69, $p$<.001, $d$=0.717; elated: $t$(232.69)=6.45, $p$<.001, $d$=0.812; bad*:* $t$(232.30)=-6.42, $p$<.001, $d$=0.808; content: $t$(249.70)=5.84, $p$<.001, $d$=0.735; sad*:* $t$(220.32)=-9.46, $p$<.001, $d$=1.192]. Mean scores before the second MI were also significantly different from mean scores after the second mood induction [worried*:* $t$(253.23)=-4.65, $p$<.001, $d$=0.585; elated*:* $t$(213.42)=5.51, $p$<.001, $d$=0.694; bad: $t$(242.96)=-4.02 , $p$<.001, $d$=0.507; content: $t$(250)=4.31 , $p$<.001, $d$=0.543; sad*:* $t$(227.91)=-7.18, $p$<.001, $d$=0.904].

There was also a significant group effect [$V$=.364, $F$(2,492)=21.75, $p$<.001, $\eta_G^2$=.182] indicating that irrespective of time*,* rUD scored higher than HC in worried [$t$(67.89)=4.53*,* $p$<.001, $d$=0.964], bad [$t$(55.97)=3.99*,* $p$<.001, $d$=0.841] and sad mood [$t$(76.29)=2.66*,* $p$=.010, $d$=0.570]. Furthermore, aUD scored higher than rUD [worried: $t$(66.75)=3.68*,* $p$<.001, $d$=0.804; bad: $t$(60.11)=5.77*,* $p$<.001, $d$=1.262; sad: $t$(70.61)=4.35*,* $p$<.001, $d$=0.948] and also higher than HC [worried: $t$(50.23)=6.97*,* $p$<.001, $d$=1.509; bad: $t$(44.13)=8.40*,* $p$<.001, $d$=1.814; sad: $t$(57.64)=6.62*,* $p$<.001, $d$=1.438]. aUD scored lower in elated and content than rUD [elated: $t$(82.18)=-2.53*,* $p$<.013, $d$=0.544; content: $t$(83.52)=-6.30*,* $p$<.001, $d$=1.354] and lower than HC [elated: $t$(72.32)=-3.18*,* $p$=.002, $d$=0.706; content: $t$(72.78)=-6.58*,* $p$<.001, $d$=1.463] whereas HC and rUD scored equal [elated: $t$(79.31)=-0.79*,* $p$=.431; content: $t$(77.36)=-0.95*,* $p$=.344].

We found no significant interaction between time and group [$V$=.064, $F$(6,492)=1.07, $p$=.363].

**Table 2**

*Means and standard deviations separately for participants with acute depression, participants with remitted depression and healthy control participants*

| *PANAS-items* | Participants within acute depression | | | | Participants with remitted depression | | | | Healthy control participants | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | First mood induction | | Second mood indcuction | | First mood induction | | Second mood induction | | First mood induction | | Second mood indcuction | |
| | Pre *M* (*SD*) | Post *M* (*SD*) | Pre *M* (*SD*) | Post *M* (*SD*) | Pre *M* (*SD*) | Post *M* (*SD*) | Pre *M* (*SD*) | Post *M* (*SD*) | Pre *M* (*SD*) | Post *M* (*SD*) | Pre *M* (*SD*) | Post *M* (*SD*) |
| Worried | 2.17 (1.06) | 2.76 (1.21) | 1.93 (1.05) | 2.57 (1.13) | 1.34 (0.61) | 2.32 (1.09) | 1.39 (0.54) | 1.91 (0.91) | 1.05 (0.22) | 1.60 (0.67) | 1.05 (0.32) | 1.48 (0.64) |
| Elated | 1.57 (0.74) | 1.17 (0.38) | 1.50 (0.71) | 1.17 (0.38) | 1.98 (0.85) | 1.27 (0.76) | 1.89 (0.99) | 1.27 (0.59) | 2.13 (0.91) | 1.38 (0.74) | 1.90 (0.74) | 1.38 (0.63) |
| Bad | 2.19 (1.09) | 2.74 (1.23) | 2.17 (1.08) | 2.33 (1.12) | 1.09 (0.29) | 1.75 (0.97) | 1.27 (0.54) | 1.59 (0.90) | 1.03 (0.16) | 1.13 (0.33) | 1.10 (0.30) | 1.23 (0.48) |
| Content | 2.12 (0.99) | 1.69 (0.90) | 2.21 (0.75) | 1.69 (0.72) | 3.43 (0.90) | 2.41 (1.00) | 3.09 (1.05) | 2.32 (0.98) | 3.60 (0.78) | 2.65 (1.10) | 2.98 (0.97) | 2.65 (1.08) |
| Sad | 2.10 (1.05) | 3.29 (1.15) | 1.95 (1.10) | 2.79 (1.18) | 1.16 (0.37) | 2.52 (1.19) | 1.20 (0.51) | 2.34 (1.18) | 1.05 (0.22) | 2.10 (0.87) | 1.10 (0.38) | 1.80 (0.79) |

*Note.* PANAS = *Positive and Negative Affect Schedule* (PANAS; Crawford & Henry, 2004)

*Pre-Analyses: Identification of additional relevant factors*

In the following, we checked whether age or gender should be included as additional covariates in the main analyses. For $Cam_{prop}$, there was only a significant main effect of age [$F(1,119)=42.79$, $p<.001$, $\eta_G^2=.264$] but no significant interaction between age and group [$F(2,119)=0.97$, $p=.380$]. Post hoc correlation tests revealed that there was a significant negative correlation between $CAM_{prop}$ and age for HCs ($r=-0.51$, $t(38)=-3.66$, $p=.001$, $d=.116$), rUDs ($r=-0.38$, $t(41)=-2.62$, $p=.012$, $d=.202$) as well as for aUDs (r=-0.65, $t(40)=-5.36$, p<.001, $d=.291$). Furthermore, for $MASC_{prop}$, there was a significant main effect of age [$F(1,119)=34.00$, $p<.001$, $\eta_G^2=.222$] but no significant interaction between age and group [$F(2,119)=1.64$, $p=.197$]. Post hoc correlation tests revealed that there was a significant negative correlation between $MASC_{prop}$ and age for HCs ($r=-0.41$, $t(38)=-2.75$, $p=.009$, $d=.129$), rUD ($r=-0.38$, $t(41)=-2.60$, $p=.013$, $d=-.195$) as well as for aUD (r=-0.60, $t(40)=-4.73$, p<.001, $d=.297$).

On $CAM_{prop}$, ANOVAs revealed neither a significant effect of gender [$F(1,119)=0.71$, $p=.402$] nor of the interaction of gender and group [$F(2,119)=2.02$, $p=.137$]. Moreover, on $MASC_{prop}$, there was a significant effect of gender [$F(1,119)=7.78$, $p=.006$, $\eta_G^2=.061$] but no significant interaction between gender and group [$F(2,119)=0.96$, $p=.386$]. Across both groups, female participants ($M$ = 0.786 , $SD$ = 0.094) scored significantly higher than male participants ($M$ = 0.728 , $SD$ = 0.131) in $MASC_{prop}$.

Based on these analyses, we decided to refrain from including age or gender as additional independent variables in the main analyses.

*ToM decoding (FER)*

For $CAM_{prop}$ there was neither a significant effect of group [$F(2,122)=0.16$, $p=.856$] nor of MI [$F(1,122)=0.18$, $p=.671$] and no significant interaction between group and MI [$F(2,122)=0.02$, $p=.978$]. The same applies for $CAM_{negativity}$ [group: $F(2,122)=0.04$, $p=.964$; MI: $F(1,122)=0.02$, $p=.895$; group x MI: $F(2,122)=0.09$, $p=.915$] and $CAM_{positivity}$ [group: $F(2,122)=0.53$, $p=.588$; MI: $F(1,122)=0.04$, $p=.833$; group x MI: $F(2,122)=0.05$, $p=.955$]. Furthermore, there were no significant results for $CAM\text{-}Neg_{prop}$ [group: $F(2,116)=0.22$, $p=.800$; MI: $F(1,116)=0.14$, $p=.706$; group x MI: $F(2,116)=0.24$, $p=.787$], $CAM\text{-}Pos_{prop}$ [group: $F(2,116)=0.85$, $p=.431$; MI: $F(1,116)=0.27$, $p=.607$; group x MI: $F(2,116)=0.50$, $p=.606$], and $CAM\text{-}Neut_{prop}$ [group: $F(2,116)=0.54$, $p=.586$; MI: $F(1,116)=0.32$, $p=.570$; group x MI: $F(2,116)=0.62$, $p=.542$].

*Confidence and Simplicity Ratings for ToM decoding*

When examining the confidence ratings to see if they differed among the three different groups depending on whether or not a MI occurred, the following were found: There was a significant main effect of group [$F(2,122)=3.43$, $p=.036$, $\eta_G^2=.047$], but no significant effects of MI [$F(1,122)=0.01$, $p=.906$] and group x MI [$F(2,122)=1.09$, $p=.340$]. To compare the three groups with

each other, three post hoc ANOVAS were performed with group and MI as predictors. For HC vs. aUD, group difference was significant [$F(1,80)=6.48$, $p=.013$, $\eta_G^2=.068$]. HC ($M$ = 68.01, $SD$ = 13.82) revealed significant higher confidence ratings than aUD ($M$ = 61.16 , $SD$ = 11.71). For HC vs. rUD, groups did not differ significantly from each other [$F(1,81)=0.54$, $p=.466$]. Confidence ratings for HC were equal to that of rUD ($M$ = 65.96 , $SD$ = 13.09). For aUD vs. rUD, there was a marginal significant group effect [$F(1,83)=3.59$, $p=.062$, $\eta_G^2=.036$]. Confidence ratings were in trend higher in rUD compared to aUD.

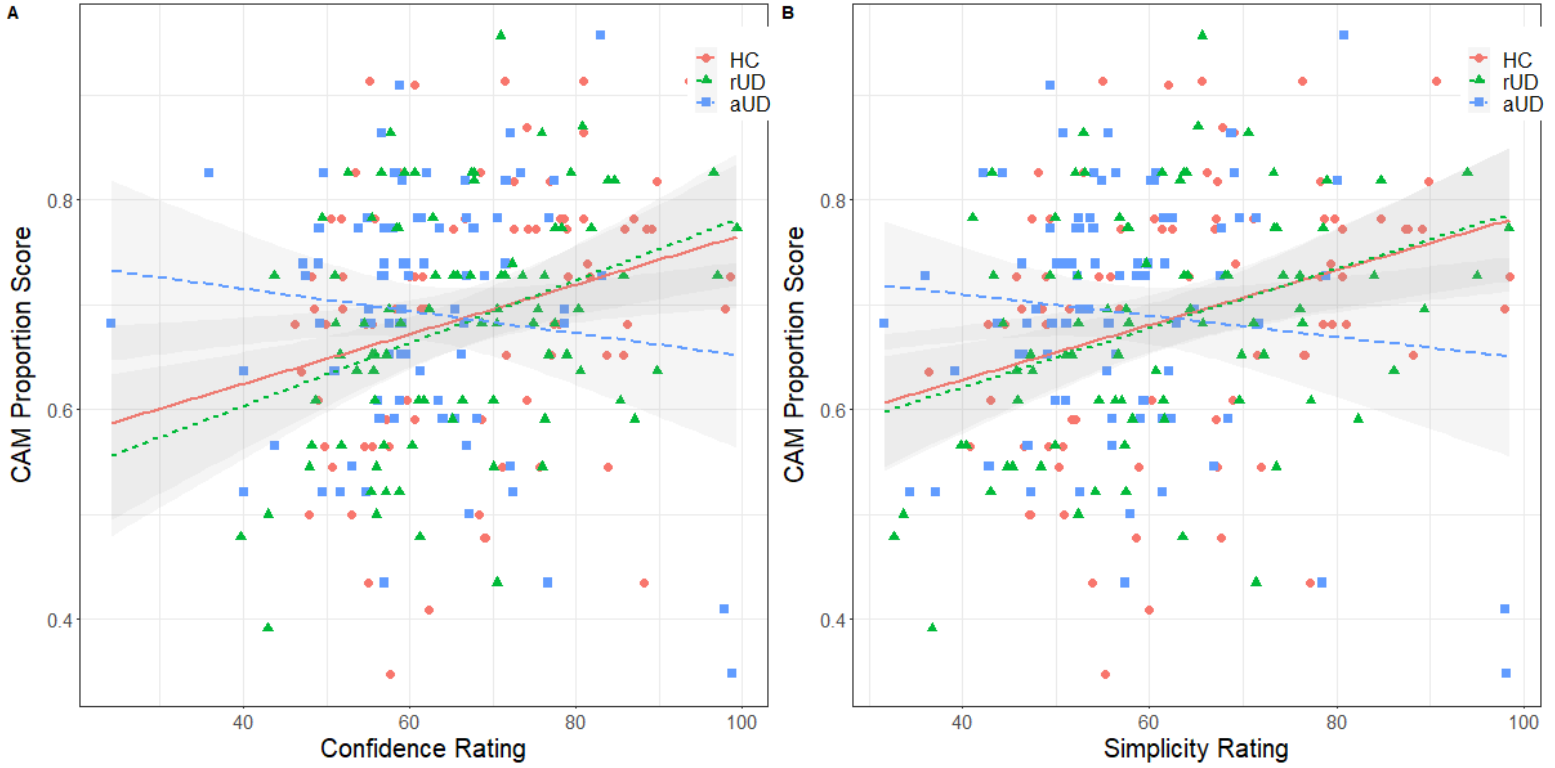
Concerning the simplicity ratings, a similar picture emerges as for the confidence ratings. For Simp, there was a significant main effect of group [$F(2,122)=3.12$, $p=.047$, $\eta_G^2=.044$] but no significant effect of MI [$F(1,122)=0.11$, $p=.744$] and group x MI [$F(2,122)=0.59$, $p=.554$]. Post hoc tests revealed that for HC vs. aUD, there was a significant group effect [$F(1,80)=6.31$, $p=.014$, $\eta_G^2=.066$]. HCs ($M$ = 63.94 , $SD$ = 15.01) simplicity ratings were significantly higher than those of aUDs ($M$ = 56.82, $SD$ = 11.82). For HC vs. rUD, group did not differ significantly from each other [$F(1,81)=0.75$, $p=.389$]. Simplicity ratings in HC were equal to that of rUD ($M$ = 61.29, $SD$ = 14.31). For aUD vs. rUD, there was a marginal significant group effect [$F(1,83)=2.78$, $p=.099$]. Confidence ratings were in trend higher in rUD compared to aUD.

When examining whether accuracy in CAM was significantly related to the confident ratings, the following emerged: For $CAM_{prop}$, there was a significant main effect of Con [$F(1,116)=4.34$, $p=.039$, $\eta_G^2=.020$] as well as a significant interaction effect of group and Con [$F(2,116)=3.91$, $p=.023$, $\eta_G^2=.040$]. No other effects reached significance [group: $F(2,116)=0.17$, $p=.844$; MI x Con: $F(1,116)=3.41$, $p=.068$; group x MI x Con: $F(2,116)=0.75$, $p=.474$]. Post hoc tests revealed that there was a significant positive correlation between $CAM_{prop}$ and Con within HC [$r=.257$, $t(78)=2.34$, $p=.022$, $d=6.889$] and within rUD [$r=.347$, $t(84)=3.39$, $p=.001$, $d=7.050$], and a non-significant negative relationship within aUD [$r=-.106$, $t(82)=-0.96$, $p=.340$]. Results are visualized in Figure 1A.

Also for the simplicity ratings, accuracy in CAM was significantly related to the simplicity ratings, as a function of group. For $CAM_{prop}$, there was a significant main effect of Simp [$F(1,116)=6.29$, $p=.014$, $\eta_G^2=.030$] as well as a significant interaction effect of group and Simp [$F(2,116)=3.91$, $p=.023$, $\eta_G^2=.040$]. No other effects reached significance [group*: $F(2,116)=0.17$, $p=.846$; MI x Con: $F(1,116)=0.93$, $p=.337$; group x MI x Con: $F(2,116)=0.50$, $p=.607$]. Post hoc tests revealed that there was a significant positive correlation between $CAM_{prop}$ and Simp within HC [$r=.307$, $t(78)=2.85$, $p=.006$, $d=5.959$] and within rUD [$r=.357$, $t(84)=3.50$, $p<.001$, $d=5.988$], and a non-significant negative relationship within aUD [$r=-.101$, $t(82)=-0.92$, $p=.360$]. Results are shown in Figure 1B.

**Figure 1**

*Correlation between CAM proportion score and A) the confidence ratings or B) the simplicity ratings, separately for each group (aUD versus rUD versus HC)*




*Note.* CAM = Cambridge Mindreading Face-Voice Battery (Golan et al., 2006), aUD = patients with acute unipolar depressive disorder, rUD = patients with remitted unipolar depressive disorder, HC = healthy control group

*ToM Reasoning*

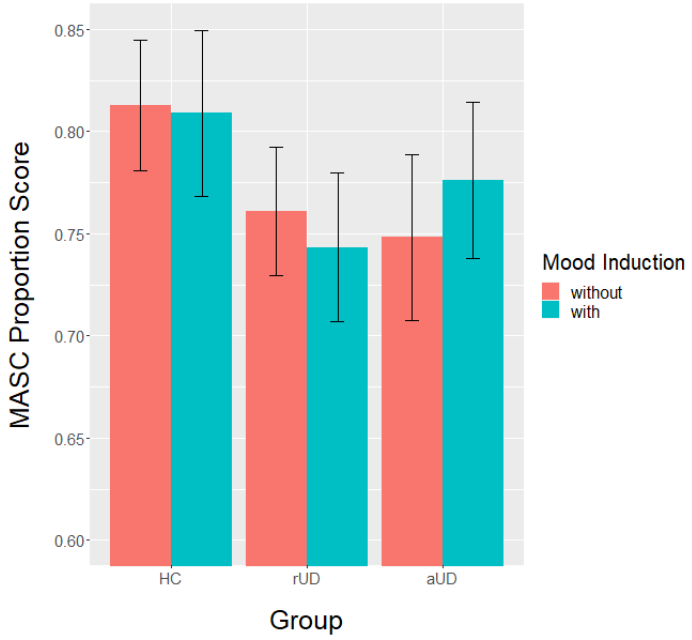
For $MASC_{prop,}$ we found a significant group effect [$F(2,122)=3.87$, $p=.023$, $\eta_G^2=0.046$ ], no significant effect of MI [$F(1,122)=0.04$, $p=.842$] and no interaction of group and MI [$F(2,122)=1.76$, $p=.177$]. Post hoc ANOVAS revealed a significant difference between aUD and HC [$F(1,80)=4.25$, $p=.043$, $\eta_G^2=0.040$] and between rUD and HC [$F(1,81)=8.01$, $p=.006$, $\eta_G^2=0.066$], but not between aUD and rUD [$F(1,83)=0.19$, $p=.662$]. The HC group ($M = 0.811$, $SD = 0.114$) scored significantly higher in $MASC_{prop}$ than the aUD group ($M = 0.762$, $SD = 0.127$) and the rUD group ($M = 0.752$, $SD = 0.110$). The results are visualized in Figure 2 separately for group and MI.

Furthermore, we analyzed the MASC error types. No significant effects were found for $MASC_{no}$ [group: $F(2,122)=2.30$, $p=.105$; MI: $F(1,122)=0.04$, $p=.850$; group x MI: $F(2,122)=0.83$, $p=.437$] and $MASC_{exceeded}$ [group: $F(2,122)=0.95$, $p=.388$; MI: $F(1,122)=1.86$, $p=.176$; group x MI: $F(2,122)=0.69$, $p=.504$]. For $MASC_{less}$, there was a significant group effect [$F(2,122)=3.07$, $p=.049$, $\eta_G^2=0.031$], no significant main effect of MI [$F(1,122)=0.65$, $p=.420$] and no significant interaction [$F(2,122)=0.65$, $p=.524$]. Additional analyses showed a significant difference between aUD and HC [$F(1,80)=4.02$, $p=.048$, $\eta_G^2=0.029$] and between rUD and HC [$F(1,81)=6.36$, $p=.014$, $\eta_G^2=0.045$] but not between aUD and rUD [$F(1,83)=0.18$, $p=.668$]. The HC group ($M = 0.071$, $SD = 0.057$) scored significantly lower in $MASC_{less}$ than the aUD group ($M = 0.095$, $SD = 0.078$) and the rUD group ($M = 0.101$, $SD = 0.079$).

In addition, we also looked at the two different mental state modalities: cognitive ToM ($MASC_{cog}$) and affective ToM ($MASC_{emo}$). For $MASC_{emo}$, there were no significant results [group: $F(2,122)=1.96$, $p=.146$; MI: $F(1,122)=0.08$, $p=.784$; group x MI: $F(2,122)=0.42$, $p=.657$]. However, for $MASC_{cog}$, we found a significant group effect [$F(2,122)=3.90$, $p=.023$, $\eta_G^2=0.042$], whereas MI [$F(1,122)=0.19$, $p=.666$] and the interaction of group and MI [$F(2,122)=2.25$, $p=.109$] did not reach significance. In the post hoc ANOVAS, the difference in $MASC_{cog}$ between aUD and HC [$F(1,81)=4.22$, $p=.043$, $\eta_G^2=0.037$] and between rUD and HC [$F(1,81)=7.82$, $p=.006$, $\eta_G^2=0.060$] reached significance, whereas that between aUD and rUD [$F(1,83)=0.20$, $p=.654$] did not. The HC group ($M=0.828$, $SD=0.130$) scored significantly higher in $MASC_{cog}$ than the aUD group ($M=0.774$, $SD=0.145$) and the rUD group ($M=0.764$, $SD=0.126$).

**Figure 2**

*Effect of group (aUD versus rUD versus HC) on the MASC proportion score separately for conditions with versus without negative mood induction*




*Note.* aUD = patients with acute unipolar depressive disorder, rUD = patients with remitted unipolar depressive disorder, HC = healthy control group, MASC = Movie for the Assessment of Social Cognition (MASC; Dziobek et al., 2006)

### 3.1.4 Discussion

This is one of the first studies using ecologically valid stimulus material to examine whether acute and remitted UD patients are impaired in complex FER (ToM decoding). Regarding ToM reasoning, we aimed to substantiate the findings by Wolkenstein et al. (2011) and Zwick and Wolkenstein (2017). Beyond that, this is the first study to investigate the influence of a sad mood induction and third variables (age and gender) on ToM performance that might help to explain the previously reported inconclusive results concerning ToM in UD.

We found no difference among the three investigated groups (aUD, rUD and HC) in FER performance. This is contrary to what we expected and against some studies, which found HCs outperform UDs in basic FER (e.g., Demenescu, Kortekaas, den Boer, & Aleman, 2010) as well as in complex FER (ToM decoding), measured by the RMET (e.g., L. Lee et al., 2005; Nejati et al., 2012; Wang et al., 2008). With an extended perspective, our results can nevertheless be brought into line with the previous literature. There are also several studies, which did not find any difference between UDs and HCs in ToM decoding, measured by the RMET (e.g., Kettle et al., 2008; Purcell et al., 2013; Wolkenstein et al., 2011). Moreover, most studies that have found a reduced FER ability in UDs have used static images. However, static images are of low ecological validity and have a

relatively low information content compared to dynamic stimuli (Garrido-Vásquez et al., 2011). Already in non-clinical samples, dynamic displays of emotions are recognized more accurately than static displays (e.g., Weyers, Mühlberger, Hefele, & Pauli, 2006). Beyond that, in their EEG experiment, Mayes, Pipingas, Silberstein, and Johnston (2009) found evidence that the processing of dynamic facial displays is faster and more efficient than that of static displays. Therefore, FER deficits in UDs might be limited to static stimuli due to their low information content. Schaefer et al. (2010), Kan et al. (2004) and Zwick and Wolkenstein (2017) also did not find any overall deficits in UD patients when using dynamic stimuli of basic FER displays. The results of the present study suggest that this also applies to complex FER tasks with dynamic displays. In line with this, there are some studies which indicate that the facilitating effect of dynamic displays on FER performance is especially true in the case of subtle or non-prototypical emotion displays in non-clinical samples (e.g., Ambadar, Schooler, & Cohn, 2005; Bould & Morris, 2008). Hence, it seems that the more ecologically valid the stimulus material or the richer the information content is (Garrido-Vásquez et al., 2011), the smaller the group differences between UDs and HCs might be. Another difference from previous studies is the *item difficulty*. Item difficulty is quiete high in complex and not fully developed emotion displays, as used in our study. It could be that a high item difficulty compared to a low item difficulty improves the performance of the depressed group, as a challenging experimental task (Frodl et al., 2009) provides sufficient distraction to suppress rumination and projection of negativity onto it (Garrido-Vásquez et al., 2011).

Some studies indicate that there is only a difference between UDs and HCs when looking at FER performance of very *specific emotions* or *emotions of a certain valence*. For example, Joormann and Gotlib (2006) found that UDs required significantly greater intensity to identify happy expressions correctly, and significantly less intensity to identify sad expressions than angry expressions, compared to HCs. In addition, Zwick and Wolkenstein (2017) observed that aUD patients recognized happy faces less accurately than HCs when using dynamic video clips of basic facial expressions. These results are in line with the explanation by Weightman et al. (2014), who assumed a mood congruent bias in UD patients, improving the performance for negatively valenced stimuli whilst deteriorating the performance for positive or neutral stimuli. As a result, the difference in the overall performance in FER could not reach statistical significance, and the true difference in FER could be masked if the valence of the stimuli is not taken into account. By using the RMET as a measure of complex FER (ToM decoding), Szily and Kéri (2009) demonstrated that aUDs displayed an impaired complex FER performance for negative social emotions. However, in the present study, we also considered ToM decoding performance separately by valence and found no significant difference among the groups. Therefore, again, it seems that the anomalies in ToM decoding in UD

patients are not found when using dynamic and ecologically valid material with high information content.

Studies regarding deficits in decoding basic emotions in UD patients already indicate that this effect is not only very small (e.g., Bora & Berk, 2016; Dalili et al., 2015), but that there are potentially many moderating or mediating factors that influence whether or not a deficit even exists – e.g., being on pharmacotherapy (Weightman et al., 2014). Several studies examining ToM decoding by means of the RMET also indicate that the presence of certain clinical characteristics may influence whether or not a ToM decoding impairment is present or may intensify this impairment. Some of these influences could be the presence of psychosis risk (Szily & Kéri, 2009), the current presence of psychotic symptoms (Wang et al., 2008), or the presence of a suicide attempt in the past (Szanto et al., 2012). Moreover, as Wolkenstein et al. (2011) had already mentioned with reference to the results of L. Lee et al. (2005), it could be that only severely depressed patients are significantly less accurate than HCs. The influence of the severity of the depressive disorder on the differences in decoding between UDs and HCs might be even more complex, since there is evidence that mildly depressed individuals are even better than HCs at decoding mental states (Harkness et al., 2010). The sample in our study contains all patients with major depression or dysthymia, including also mildly and moderately depressed patients. In future studies, this question should be explicitly addressed by categorizing patients according to the severity of their depression and including this as a factor in the analyses. This applies also to the other specific features mentioned above. A failure to consider these factors contributes to inaccurate results, so these factors should be taken into account when studying the decoding of complex emotions. Unfortunately, we were not able to consider these factors in the present study because the sample size was too small to allow for subsample comparisons that were not already fixed a priori in the study design. Moreover, we did not define these variables a priori and clearly classified the participants in these regards.

However, although there were no differences between aUDs, rUDs and HCs in ToM decoding accuracy there are prominent findings regarding the confidence and simplicity ratings of the CAM stimuli. HCs were more confident in their judgements and found it less difficult than aUDs. rUDs scored between HCs and aUDs with a non-significantly lower score than HCs and marginal significantly higher score than aUDs. Furthermore, there was only for HCs and rUDs a significant dependence between CAM performance and confidence or simplicity ratings, respectively. These results suggest that only HCs and rUDs provided confident and simplicity ratings related to their actual performance. aUD patients were not only more insecure in their social judgements, but their ratings also did not match their actual ToM decoding ability. This is of great importance, because increased irritability in relation to the emotional expression of a counterpart can lead to difficulties in social situations, especially when one is affected oneself and when the situation is more complex in

the overall context of everyday life. In line with this, Marton, Connolly, Kutcher, and Korenblum (1993) concluded that depressed adolescents exhibit unique deficits in social self-evaluation that contribute to ineffective social behavior and the maintenance of dysphoric affect. Zwick and Wolkenstein (2017), who investigated FER difficulties separately for each basic emotion, also found only limited evidence for impairments in FER accuracy scores. However, the confidence and simplicity ratings appear to be more sensitive when it comes to finding difficulties in FER of certain basic emotions (happiness, anger and fear). The authors concluded that their findings disprove the existence of a general negativity bias in UD and point to specific impairment in certain basic emotions of aUD and rUD patients. However, the present study suggests that, when it comes to complex FER (ToM decoding), there seems to be a general uncertainty and perceived difficulty in FER only within aUD patients. Thinking forward, uncertainties in ToM decoding linked to other contextual information could lead to difficulties in social reasoning.

In accordance with our hypotheses, both aUDs as well as rUDs showed a lower performance in *ToM reasoning* compared to HCs. But contrary to our assumption, the performance of aUDs and rUDs were equal. We would first like to compare these results with those of the two studies of Wolkenstein et al. (2011) and Zwick and Wolkenstein (2017) that also examined depressed patients with the MASC. By examining a larger sample, we substantiate the trend level findings by Wolkenstein et al. (2011) who found decreased reasoning abilities in aUDs on a trend level and replicated the finding that there are more $MASC_{less}$ mistakes made by UDs than HCs. This suggests that although UD patients make inferences about mental states, these inferences are insufficient. Thus, this helps to clarify the question posed by Wolkenstein et al. (2011) as to whether this error pattern proves to be a stable finding and thus suggests that this is characteristic for UD patients. We also partly replicated the results by Zwick and Wolkenstein (2017) who also found aUDs were worse in reasoning than HCs but, in contrast to our findings, that rUDs were not worse. The different results of the studies cannot be attributed to different sample sizes or methodological differences in diagnosing and separating the two patient groups, as these aspects were comparable. However, the studies differ in the way the MASC was conducted. Zwick and Wolkenstein (2017) presented the MASC in one session. In contrast, in our study the MASC was divided into two parts and performed on different days. In addition, one part of the MASC was performed under sad MI. However, the fact that post hoc tests in which performance with versus without MI and performance in MASC part 1 and MASC part 2 were considered separately cannot explain the differences. Another explanation could be due to the study design. MASC was conducted after the CAM. It can be assumed that performing the CAM was somewhat exhaustive (sitting in a dark cabin, with electrodes attached onto the face, performing a relatively difficult task). Perhaps this mental exhaustion lowered the rUDs to the level of the aUDs.

One explanation for ToM reasoning deficits in UDs refers to the *Cognitive Theory of Depression* (Beck, 1979, 1983). A person's cognitions are based on attitudes or assumptions (schemata), which in turn have arisen from previous experiences. In depressed patients, these schemata are largely dysfunctional and lead to automated and stereotyped negative thoughts. Since these negative thoughts and biases can be triggered easily in social situations, this can be well reconciled with our findings that in aUD and rUD groups there are deficits in ToM reasoning.

Regarding the sub-aspects of ToM reasoning, in contrast to the study by Wolkenstein et al. (2011), who did not find any group differences for the two mental state modalities $MASC_{cog}$ and $MASC_{emo}$, we indeed found group differences for $MASC_{cog}$. This was probably due to the larger sample size. Healthy people showed better performance in cognitive reasoning than aUDs and rUDs. In accordance with our findings on ToM decoding, one possible explanation could be that the impact of ToM decoding on emotional ToM reasoning is greater than on cognitive ToM reasoning. This would mean that since there is no impairment in ToM decoding performance, there is also no impairment in emotional ToM reasoning.

In our study, there was no evidence that age and gender had any moderating role in the association between ToM performance and group membership (aUD vs. rUD vs. HC). Nevertheless, there are significant main effects that are worth briefly highlighting again here. The negative correlation of $CAM_{prop}$ and $MASC_{prop}$ with age suggests that, in both patients and healthy individuals, ToM decoding performance as well as ToM reasoning performance decreases with age. This is consistent with studies that have already found this to be the case for performance on basic FER (e.g., Horning et al., 2012) and ToM decoding and reasoning, mostly measured with static stimuli (Henry et al., 2013). Our study therefore suggests that these results can also be replicated with highly ecologically valid measurement procedures that capture the two ToM components. Regarding gender, there were no significant effects with respect to $CAM_{prop}$, but there was a significant main effect of gender for $MASC_{prop}$. According to this finding, women are better at ToM reasoning than men, regardless of group membership. In some points, these results contradict the results from studies that have used low ecologically valid and static material. Namely, these found better FER performance in women compared to men for both basic FER (e.g., Kessels et al., 2014) and complex FER (e.g., Ibanez et al., 2013). For ToM reasoning, however, only a trend result was found (e.g., Turkstra et al., 2020). In our study with dynamic-visual material, the picture was reversed: While there were no gender differences in complex FER performance, we found better performance by females than males in ToM reasoning.

Contrary to our hypothesis and other studies (e.g., Chepenik et al., 2007; Feyerabend et al., 2018) we did not find that a sad MI influenced ToM performance in any way. Feyerabend et al. (2018) (see Chapter 3.2) examined patients with rBD and a corresponding HC, using an identical

study design as in the present study. They found that MI negatively affected $MASC_{emo}$ scores, not only in rBDs, but also in HCs. Failure to find an effect of MI in the present study may be related to the characteristics of the samples. Firstly, in the present study, the percentage of women was higher (around 80.0% women in the present study vs. 63.1% women in the study by Feyerabend et al. (2018)). Secondly, the average age in the present study was lower (42.9 years versus 46.7 years). Thirdly, the present study additionally examined participants who were experiencing an acute depressive episode (aUD). Since ToM performance is a sensitive measure that is influenced by numerous variables, among other things age, gender and affective state (discussed again later, in Chapter 4), these differences in demographic and clinical variables could have influenced whether effects of MI on ToM were found or not. However, it could also be that it is not only because the effect of MI on ToM differed, but that sad mood per se was not induced equally in the two study samples. And indeed, at first glance and without any calculations, the comparisons of the absolute values of affective status (before and after second MI) seem to support this assumption. For example, if we look at the PANAS variable *sad*, although both studies showed a significant main effect of time (before versus after second MI) in the total sample, the effect was markedly higher in the rBD study (*d*=1.250) than in the current study (*d*=0.904). This was also true for all other variables (worried, bad, happy), except the variable pleased. Also, in the first mood induction, the effect sizes for *time* were higher in the rBD paper, for all PANAS variables without exception. Moreover, the absolute values (before versus after second MI) and the absolute change in the PANAS variable *sad* are comparable between the rBD (Pre=1.23, Post=2.36, change: 1.13) and rUD samples (Pre=1.20, Post=2.34, change: 1.14) and also between the HC group of the rBDs (Pre=1.13, Post=1.93, change: 0.8) and the HC group of the rUDs (Pre=1.10, Post=1.80, change: 0.7). However, the aUD group is somewhat out of line, having higher absolute *sad* scores than all other groups, both before and after MI, while being more comparable to the HC groups in the amount of absolute change (Pre=1.95, Post=2.79, change: 0.84). In particular, the response to MI, and perhaps a small or absent effect of MI on ToM performance in the aUD sample, could account for the different results between the Feyerabend et al. paper in the present study. However, these are only very vague assumptions based on the apparent consideration of PANAS values in the two studies. Therefore, the scope of future studies should cover comparison of the response of different groups of patients (e.g., UD vs. BD) in different states (acute vs. remitted) to a sad MI, and to examine whether this also differently affects different aspects of ToM performance.

Considering the results of CAM and MASC together, it appears that UD patients are not at all or hardly affected in early sensory processes (ToM decoding), but rather in later processing stages (ToM reasoning). Garrido-Vásquez et al. (2011) also question deficits in early emotional processing stages in UDs and point to clearer evidence for deficits in later processing stages. One explanation of

Garrido-Vásquez et al. (2011) is that in UDs, anomalies in emotional processing occur primly at later stages of processing, during which *biases* also arise and on which *dysfunctional beliefs* operate. It could be argued that for ToM reasoning, as in the MASC task and in everyday situations, perceivers are confronted with a more complex situation in terms of the need to process contextual information and to involve multiple sensory channels in receiving the input. In this regard, there is an increased risk for biases and dysfunctional beliefs to develop which would influence processing and reasoning. Furthermore, as mentioned above, the uncertainty and the perceived higher difficulty of aUD patients with ToM decoding linked to further contextual information and multiple sensory input, could lead to ToM reasoning deficits. This could also therefore explain deficits in MASC performance of aUD. However, rUD patients are also impaired in ToM reasoning, although their confidence and simplicity ratings are similar to those of HCs. This makes the explanation of the connection to uncertainty and more perceived difficulty in ToM decoding questionable. On the other hand, rUDs absolutely scored between HC and aUD in their confidence and simplicity ratings. Although these differences did not become significant, they could become so with a larger sample and should be tested in future studies. Altogether, this suggests that for rUDs and aUDs, biases and misinterpretations operate on ToM reasoning independently of or in addition to ToM decoding. Zwick and Wolkenstein (2017) have also interpreted their results in this direction, namely that basic FER deficits and reasoning deficits in aUDs exist independently from each other. Furthermore, the assumption of independence of ToM decoding and ToM reasoning is in accordance with Sabbagh (2004), who points to imaging studies that suggest that decoding and reasoning are two distinct processes: Decoding depends on contributions from the right orbitofrontal/medial temporal circuit, and reasoning relies on left medial frontal regions.

The following limitations of the present study should be mentioned: First, because of the sample size and aprior design as well as the apriori classification of subgroups, it was not possible to compare additional subsamples. In these comparisons, additional clinical or demographic characteristics (e.g., severity of depression, being on pharmacotherapy, general neurocognitive impairment etc.) could have been taken into account and could have helped to clarify the conflicting results in previous studies regarding the association between ToM performance and depression. Future studies should therefore explore this association with these clinical characteristics in mind. Secondly, the proportion of women in the study was particularly high (76.7%), and thus the investigated sample was not representative of the general population. The study should therefore be repeated with a sample consisting of equal numbers of female and male participants. Since ToM reasoning differs by gender, this could also be given more consideration.

In summary, our findings suggest that – when using ecologically valid stimuli – aUDs and rUDs are not impaired with respect to ToM decoding (complex FER), but that aUDs and rUDs are impaired

with respect to ToM reasoning. Especially the cognitive mental state modality is impaired. Although there are no ToM decoding deficits in accuracy, aUD patients felt less confident and found it more difficult to judge the expressions presented. The ability to infer the mental state of other people correctly and confidently contributes decisively to appropriate social responses, and thus to satisfactory social interaction. Therefore, it is conceivable that the low level of social functioning known to accompany depression can be ascribed partially to ToM deficits in people affected. Consequently, it would be important to target deficits in ToM reasoning in the treatment of depression and in relapse prevention to improve the understanding and interpretation of social information. Beyond that, it could be prove to be valuable to incorporate psychosocial treatments, social skills training, and ToM training into standard treatment protocols.

## 3.2 Study 2: Theory of mind in remitted bipolar disorder: Only younger patients struggle with task of high ecological validity (Feyerabend et al., 2018)[6]

To date, research concerning ToM in rBD has yielded inconclusive results. As in the context of UD, this may be a result of methodological shortcomings and the failure to consider relevant third variables. Furthermore, studies using ecologically valid stimuli are rare. This study examines ToM decoding and ToM reasoning in rBD patients, using ecologically valid stimuli. Additionally, the effects of sad MI as well as of age and gender are considered.

### 3.2.1 Introduction

Poor social functioning and interpersonal problems have been frequently reported in patients with BD (e.g., Depp et al., 2010; Hoertnagl, Oberheinricher, & Hofer, 2014; MacQueen, Young, & Joffe, 2001). One possible explanation for these findings may relate to social cognition deficits (Samamé, 2013). Indeed, studies have found impaired basic FER, complex FER (ToM decoding) and ToM reasoning in BD patients in acute depressive or manic mood states as well as in the remitted or euthymic state of the disease. However, findings are inconsistent and there are a lot of studies who did not find differences between BDs and HCs in basic FER, complex FER (ToM decoding) and ToM reasoning. An overview of the previous literature is provided in Chapter 1.5. In summary, studies examining FER and ToM in rBD lack ecological validity because they mostly use static pictures or verbally presented material (e.g., storytelling).

Interestingly, there are only a few studies investigating the recognition of complex emotions (ToM decoding). These few studies usually used the RMET and also yielded inconsistent findings. Again, the ecological validity of these studies is limited, since unlike in everyday experience, the RMET only presents the eye region and uses static rather than dynamic stimuli. Thus, one aim of the present study is to examine ToM decoding (by complex FER ability) in rBD using more ecologically valid stimuli that a) are dynamic and b) display real persons who in turn display c) complex emotional expressions in d) different developmental stages (ranging from subtle to fully developed).

In their recent meta-analysis, Bora and Berk (2016) conclude that there are significant but modest-sized ToM reasoning dysfunctions in rBD which are more severe during acute depressive or manic episodes. However, similarly to studies of ToM decoding, few studies have used highly ecologically valid measures. Demands on highly ecologically valid tasks assessing ToM reasoning are: a) using dynamic stimuli, b) showing real persons, and c) displaying everyday life situations to ensure

[6] Large parts of this Chapter 3.2 were taken verbatim from the original paper.

the possibility of using contextual information to draw conclusions. So far, in the context of BD, only few studies have addressed this need, using the MASC. In the study by Montag et al. (2010), MASC performance of rBD was impaired as compared to HC when looking at affective ToM as opposed to cognitive ToM. In addition, Santos et al. (2017) also found rBD patients and their first-degree relatives being worse in MASC than HC. A further aim of the present study is to replicate these findings.

Besides issues concerning stimulus material, one explanation for the inconsistent findings concerning FER and ToM in rBD is the presence (or absence) of a sad mood. Some studies indicate an influence of the current mood on emotion recognition in rBD (e.g., McKinnon et al., 2010). Therefore, another goal of the present study is to experimentally investigate whether complex FER (ToM decoding) and ToM reasoning performance differs between rBD and HCs as a function of whether a negative MI takes place or not.

In search of further possible reasons for the above-mentioned inconsistencies concerning ToM in BD, additional variables have been considered. This also includes variables assumed to generally influence social cognition – regardless of the presence or absence of a mental disorder – such as gender or age. As already mentioned in Chapter 1.4 and Chapter 1.5, it has become evident that associations present in non-clinical samples may be different, or even absent, in BD. Vaskinn, Sundet, Friis, Simonsen, Birkenæs, et al. (2007), for example, found that healthy male participants performed worse than their female counterparts in FER, whereas no gender difference was observed within BD. Another variable that might be of interest in this context is *age*, because it is not yet clear whether the significant decrease in FER performance with age in non-clinical samples (e.g., Mill et al., 2009; Richter et al., 2011) also exists in rBD. If the relationship between age/gender and ToM were different for BDs than for HCs, even when matching for age, it would be relatively likely to overlook differences between BDs and HCs. Therefore, the current study not only matched for age and gender, but also explored whether the relationship between these variables and ToM decoding/reasoning was the same in rBD as in HCs.

In summary, the present study tests the following hypotheses:

1) First, rBD show deficits in ToM decoding (FER) and ToM reasoning compared to HCs when using tasks of high ecological validity.
2) Second, these deficits are more pronounced under negative MI.
3) Furthermore, we aimed to explore the influence of age and gender on decoding and reasoning in rBD compared to HCs.

## 3.2.2 Methods

### 3.2.2.1 Sample

*N* = 84 participants (rBD: *N* = 44; HC: *N* = 40) were recruited by advertisements in inpatient and outpatient clinics as well as on the internet.

*Inclusion criteria* for the rBD group were: a) lifetime bipolar I or II disorder or cyclothymia according to the DSM-IV; b) currently remitted as defined by a HDRS17 score below 8, an IDS-C30 score below 12 and a *Young Mania Rating Scale* score below 12 (YMRS; Muehlbacher et al., 2011). Healthy participants were included if there was no evidence of current or lifetime mental disorder. Trained interviewers conducted the SICD-I and SCID-II in both groups. The groups were matched with regard to gender, level of education, and age (+/- 5 years).

*Exclusion criteria* for all participants were: a) insufficient knowledge of the German language, b) diseases affecting the central nervous system, c) age below 18 or above 69 years, and d) neurocognitive impairments as defined by an IQ-score below 85 according to the MWT-B or extreme outlier downwards (< 3 * interquartile range) in either the VLMT or the TMT.

Further *exclusion criteria* for the rBD group were: a) current alcohol or substance abuse and/or dependency or lifetime dependency if abstinence period < 3 years, b) acute or lifetime psychotic symptoms except mood congruent delusions during affective episodes, c) current anorexia nervosa (Body Mass Index ≤ 18 kg/m²), d) comorbid schizoid, schizotypical, paranoid, antisocial and/or borderline personality disorder according to DSM IV.

Table 3A shows the sample characteristics. rBDs met criteria for remitted bipolar I (54.55 %) or bipolar II disorder (45.45 %). There were no participants with cyclothymia in the study sample. Of the rBDPs, 40.9% had one or more co-morbid mental illness(es) and 95.45% were on medication at the time of testing.

**Table 3**

*Demographic, clinical and neurocognitive variables*

| **Variable** | **rBD** ***N* = 44** | **HC** ***N* = 40** | ***df*** | ***χ²*** | ***p*** |
|---|---|---|---|---|---|
| *A) Demographic variables* | | | | | |
| Level of education (Abitur) in % | 63.6 | 70.0 | 1 | 0.15 | .699 |
| Gender (female) in % | 63.6 | 62.5 | 1 | 0.00 | 1.00 |
| Current living condition in % | | | 3 | 1.15 | .766 |
| living alone | 34.09 | 25.00 | | | |
| living together with life partner | 52.27 | 55.00 | | | |
| living together with parents/relatives | 9.09 | 12.50 | | | |
| other | 4.55 | 7.50 | | | |
| Current social contact in % | | | 4 | 2.97 | .563 |
| more than 1 time per week | 54.55 | 71.79 | | | |
| not more than 1 time per week | 25.00 | 12.82 | | | |
| once in two weeks | 11.36 | 7.69 | | | |
| once per month | 6.82 | 5.13 | | | |
| only in hallway or at work | 2.27 | 2.56 | | | |
| | **M (SD)** | **M (SD)** | ***df*** | ***t*** | ***p*** |
| Age | 46.98 (12.39) | 46.03 (12.69) | 80.83 | 0.35 | .729 |
| *B) Clinical variables* | | | | | |
| HDRS17-Score | 2.52 (2.43) | - | | | |
| IDS-C30-Score | 4.95 (5.13) | - | | | |
| YMRS-Score | 0.84 (1.60) | - | | | |
| BDI-Score | 5.02 (5.25) | 2.38 (2.77) | 60.99 | 2.85 | .006 |
| MSS-Score | 2.74 (4.28) | 1.58 (2.05) | 59.47 | 1.58 | .119 |
| *C) Neurocognitive variables* | | | | | |
| MWT-B (IQ-score) | 117.36 (17.17) | 122.75 (15.35) | 81.98 | -1.52 | .133 |
| TMT-A (seconds) | 31.48 (11.73) | 27.48 (9.50) | 80.97 | 1.72 | .088 |
| TMT-B (seconds) | 75.64 (37.58) | 61.83 (18.22) | 63.43 | 2.17 | .034 |
| *VLMT* | | | | | |
| short-term verbal retentiveness | 56.48 (8.74) | 59.55 (7.28) | 81.41 | -1.76 | .083 |
| memory loss after delayed recall | 11.38 (2.73) | 12.80 (2.42) | 81.95 | -2.52 | .014 |
| corrected recognition | 14.43 (0.79) | 14.48 (1.01) | 73.63 | -0.22 | .829 |

*Note*. rBD = patients with remitted bipolar disorder, HC = healthy control group, HDRS17 = Hamilton Depression Rating Scale, IDS-C30 = Inventory of Depressive Symptomatology, YMRS = Young Mania Rating Scale, BDI = Beck Depression Inventory, MSS = Self-Report Manic Inventory, MWT-B = Multiple Choice Word Fluency Test, TMT = Trail Making Test, VLMT = Verbal Learning and Memory Test

### 3.2.2.2 Materials & Procedure

*Assessment of symptoms and neurocognitive functioning*

In the rBD group we used the SIGHD-IDS to determine both the HDRS17 score and the IDS-C30 score. In order to ensure the absence of (hypo-)manic symptoms in the rBD group we used the YMRS. Additionally, self-report measures were applied to assure the absence of symptoms in both groups: The BDI and the *Self-Report Manic Inventory* (MSS; Krüger, Bräunig, & Shugar, 1997).

In order to control for significant neurocognitive impairments, we used the *VLMT*, the *TMT-A and -B* and the *MWT-B* (described in more detail in Chapter 3.1.2.2). Table 3B & C presents the clinical and neurocognitive variables separately for rBDs and HCs, showing significant group differences for some of them.

*Assessment of ToM decoding*

We used the face task of the CAM to asses ToM decoding. For the purpose of exploring the influence of sad MI on FER, we subdivided the CAM into two parts to present one part with and the other part without sad MI (same procedure as in Study 1, see Chapter 3.1.2.2: *Assessment of ToM decoding*). Film clips were presented in a randomized order.

We used the *CAM proportion scores* ($CAM_{prop}$), defined as "number of correct answers in the respective CAM part / total number of items in the respective CAM part", as an index of FER performance.

*Assessment of ToM reasoning*

Participants were presented with the MASC in order to evaluate ToM reasoning. For the present study we divided the MASC into two parts (same procedure as in Study 1, see Chapter 3.1.2.2: *Assessment of ToM reasoning*). We calculated *proportion scores* ($MASC_{prop}$) for each part. Additionally, in line with the study by Montag et al. (2010), we calculated proportion scores for two different mental state modalities: *cognitive* ($MASC_{cog}$): "What is X thinking or intending", and *emotional* ($MASC_{emo}$): "What is X feeling?".

*Negative mood induction*

To induce negative mood we applied a 10-minute and a 5-minute version of Peer Gynt suite no. 1 op. 46 "The Death of Ase" from Grieg while participants were asked to reflect on a sad personal experience (see Chapter 2.4).

*Positive and Negative Affect Schedule (PANAS)*

To verify the success of the mood induction we used 5 items of the PANAS: *worried*, *elated*, *bad*, *content*, *sad* administered at multiple time points. Participants were asked to rate the extent to which they experienced these emotions at the present moment on a 5-point Likert Scale ranging from "very slightly" to "very much".

*Procedure*

Participants provided informed consent at the beginning of the diagnostic session (t0). Afterwards, an anamnestic interview, the clinical assessments and the neuropsychological tests were applied[7]. Included participants took part in two testing sessions. A minimum time span of 24 hours between t1 and t2 was required. In the event of a time period longer than 2 weeks between t0 and t1 and/or t2, SIGHD-IDS, YMRS, BDI and MSS were assessed again to ensure sustained euthymia in rBDPs.

In both test sessions one part of the CAM was administered followed by one part of the MASC. The CAM part and the mood induction condition (yes versus no) were fully balanced and randomly assigned to the participants at t1 and t2.

In the mood induction session the music piece was presented immediately before the CAM and again before the MASC. Participants answered the PANAS items immediately before and after the mood inductions. In the session without mood induction, PANAS had to be completed three times: Before the CAM as well as before and after the MASC.

Due to concomitant eye tracking recordings[8], participants were requested to sit on a fixed chair. Furthermore, EMG recordings[9] required electrodes to be attached to the participant's face. Since no EMG or eye tracking recordings took place during the MASC, electrodes were removed beforehand and the sitting position could be adjusted more comfortably.

rBDPs had a 9-month follow-up session (t3) to assess the course of the illness.[10]

### 3.2.2.3 Statistical analyses

We performed the statistical analyses with R, Version 4. 1. 2 (R Core Team, 2021). Level of significance was set to $\alpha < .05$.

*Pre-Analyses*: To decide whether age and gender should be included in the main analyses as additional factors, we first analyzed if the relationship between the dependent variables and these two variables (age or gender) was significantly different between the groups (rBD vs. HC). We planned to include age or gender in the main analysis of any dependent variable if the respective relation was significantly different for one group as for the other. To test that for age, we conducted ANCOVAS with $CAM_{prop}$ and $MASC_{prop}$ as dependent variables and group and age as independent variables. To test that for gender, we conducted separate ANOVAs with gender and group as

[7] Besides the material described in this section we used additional questionnaires in this study that are, however, not relevant to the question that is examined here. Participants were asked to complete the following questionnaires at home: Inventory of Interpersonal Problems (Horowitz et al., 2000); Social support questionnaire (Fydrich et al., 2007); Cognitive Emotion Regulation Questionnaire (Garnefski & Kraaij, 2007); Emotion regulation questionnaire (Gross & John, 2003); Empathy Quotient (Baron-Cohen & Wheelwright, 2004). The respective results are reported elsewhere.

[8] This is outside the scope of this work, and the respective results are reported elsewhere

[9] This is outside the scope of this work, and the respective results are reported elsewhere

[10] This is outside the scope of this chapter, and the respective results are reported in chapter 3.3

between-subject factors for each of the dependent variables. We planned to include age or gender in the main analysis, if there were a significant interaction between group and age/gender.

For the *main analyses*, we conducted mixed-factor ANOVAs and mixed-factor ANCOVAs with $CAM_{prop}$, $MASC_{prop}$, $MASC_{cog}$ and $MASC_{emo}$ respectively as dependent variables. We defined group as a between-subjects factor and *MI* as a within-subject factor. Based on the aforementioned pre-analyses, we included age as an additional continuous variable as well as its interactions for all the dependent variables. Post-hoc tests were performed when needed. For all analyses, we centered the age variable. For the ANOVAs and ANCOVAs, we report $\eta_G^2$ as indicators of effect size. For the t-tests we report *Cohen's d* (*d*).

### 3.2.3 Results

*Manipulation check of mood induction*

A MANOVA was conducted, with a four-level *time* factor (pre_1 = *before first MI*, post_1 = *after first MI*, pre_2 = *before second MI*, post_2 = *after second MI*) and the group factor (rBD vs. HCs) as predictors and the item scores (worried, elated, bad, content and sad) as dependent variables.

We found the expected main effect of time [*V*=.519, *F*(3,331)=13.58, *p*<.001, $\eta_G^2$=.173]. Mean scores before the first mood induction varied significantly from mean scores after the first MI in the expected direction [worried: *t*(113.75)=-8.77, *p*<.001, *d*=1.353; elated: *t*(150.42)=6.74, *p*<.001, *d*=1.040; bad*: t*(94.29)=-10.12, *p*<.001, *d*=1.561; content: *t*(162.76)=8.48, *p*<.001, *d*=1.308; sad*: t*(96.28)=-10.98 , *p*<.001, *d*=1.695]. Also, mean scores before the second MI varied significantly from mean scores after the second MI [worried*: t*(126.64)=-6.56, *p*<.001, *d*=1.013; elated*: t*(139.65)=3.81, *p*<.001, *d*=0.589; bad: *t*(134.49)=-5.97 , *p*<.001, *d*=0.921; content: *t*(161.44)=3.60 , *p*<.001, *d*=0.556; sad*: t*(105.55)=-8.10, *p*<.001, *d*=1.250].

There was also a significant group effect [*V*=.078, *F*(1,331)=5.51, *p*<.001, $\eta_G^2$=.078] indicating that irrespective of time*,* rBD scored higher in worried [*t*(79.68)=3.50*, p*<.001, *d*=0.754], bad [*t*(74.69)=3.30*, p*=.002, *d*=0.706] and sad mood [*t*(81.41)=2.27*, p*=.026, *d*=0.491], compared to HCs. There was no significant interaction between time and group [*V*=.056, *F*(3,331)=1.24, *p*=.232].

Means and standard deviations are presented separately for rBD and HCs in Table 4.

**Table 4**

*Means and standard deviations for the PANAS-items before and after first and second mood induction*

| *PANAS-items* | Remitted bipolar group (*N* = 44) | | | | Healthy control group (*N* = 40) | | | |
|---|---|---|---|---|---|---|---|---|
| | First mood induction | | Second mood indcuction | | First mood induction | | Second mood induction | |
| | Pre *M* (*SD*) | Post *M* (*SD*) | Pre *M* (*SD*) | Post *M* (*SD*) | Pre *M* (*SD*) | Post *M* (*SD*) | Pre *M* (*SD*) | Post *M* (*SD*) |
| worried | 1.18 (0.45) | 2.59 (1.00) | 1.27 (0.59) | 2.23 (1.01) | 1.15 (0.43) | 1.78 (0.80) | 1.15 (0.36) | 1.68 (0.73) |
| elated | 2.05 (0.89) | 1.20 (0.51) | 1.66 (0.99) | 1.16 (0.43) | 2.08 (0.83) | 1.38 (0.70) | 1.85 (1.00) | 1.38 (0.77) |
| bad | 1.09 (0.29) | 1.98 (1.02) | 1.33 (0.61) | 1.77 (0.99) | 1.05 (0.22) | 1.33 (0.53) | 1.20 (0.46) | 1.30 (0.65) |
| content | 3.57 (0.82) | 2.18 (0.90) | 3.16 (0.91) | 2.48 (1.09) | 3.63 (0.90) | 2.63 (1.03) | 3.15 (0.89) | 2.75 (1.03) |
| sad | 1.11 (0.32) | 2.82 (1.23) | 1.23 (0.42) | 2.36 (1.10) | 1.08 (0.35) | 2.28 (1.06) | 1.13 (0.33) | 1.93 (0.92) |

*Note.* PANAS = *Positive and Negative Affect Schedule* (PANAS; Crawford & Henry, 2004)

*Pre-Analyses: Identification of additional relevant factors*

We next considered whether age or gender should be added as additional covariates in the main analyses. For $Cam_{prop}$, there was a significant main effect of age [$F(1,80)=13.24$, $p<.001$, $\eta_G^2=.142$] and furthermore there was a significant interaction between age and group [$F(1,80)=4.13$, $p=.045$, $\eta_G^2=.049$]. Post hoc correlation tests revealed, that there was a significant negative correlation between $CAM_{prop}$ and age for HCs ($r=-0.51$, $t(38)=-3.61$, $p=.001$, $d=.131$) but not for rBDs ($r=-0.19$, $t(42)=-1.27$, $p=.210$). Furthermore, for $MASC_{prop}$, there was a significant main effect of age [$F(1,80)=11.35$, $p=.001$, $\eta_G^2=.124$] but no significant interaction between age and group [$F(1,80)=1.93$, $p=.168$]. Post hoc correlation tests revealed, that there was a non-significant negative correlation between $MASC_{prop}$ and age for HCs ($r=-0.21$, $t(38)=-1.33$ $p=.190$) and a significant negative correlation for rBD (r=-0.48, $t(42)=-3.53$, p<.001, $d=.04$).

On $CAM_{prop}$, ANOVAs neither revealed a significant effect of gender [$F(1,80)=0.01$, $p=.919$] nor of the interaction of gender and group [$F(1,80)=0.33$, $p=.570$]. Moreover, on $MASC_{prop}$, there was no significant effect of gender [$F(1,80)=0.20$, $p=.654$] and also no significant interaction between gender and group [$F(1,80)=0.78$, $p=.380$]. Across both groups, female participants ($M = 0.775$, $SD = 0.096$) scored equal to the male participants ($M = 0.763$, $SD = 0.123$) in $MASC_{prop}$.

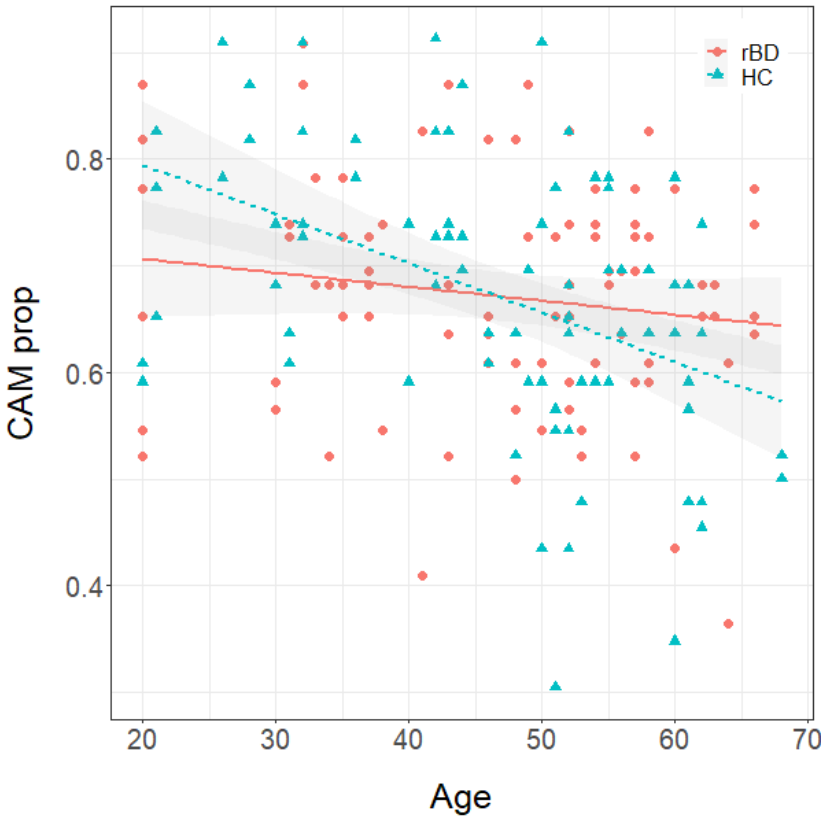
Based on these analyses, we decided to include age as additional independent variable in the main analyses for all dependent variables and to refrain from including gender as additional factor.

*Facial emotion recognition*

For $CAM_{prop}$ there was neither a significant effect of group [$F(1,80)=0.00$, $p=.982$] nor of MI [$F(1,80)=0.03$, $p=.856$]. Furthermore, there was no significant interaction between group and MI [$F(1,80)=0.27$, $p=.607$]. However, there was a significant main effect of age [$F(1,80)=13.24$, $p<.001$, $\eta_G^2=.099$], which was qualified by a significant interaction between group and age [$F(1,80)=4.13$, $p=.045$, $\eta_G^2=.033$ (see Figure 3)].

**Figure 3**

*Interaction effect of group (rBD versus HC) and age on the CAM proportion score ($Cam_{prop}$)*



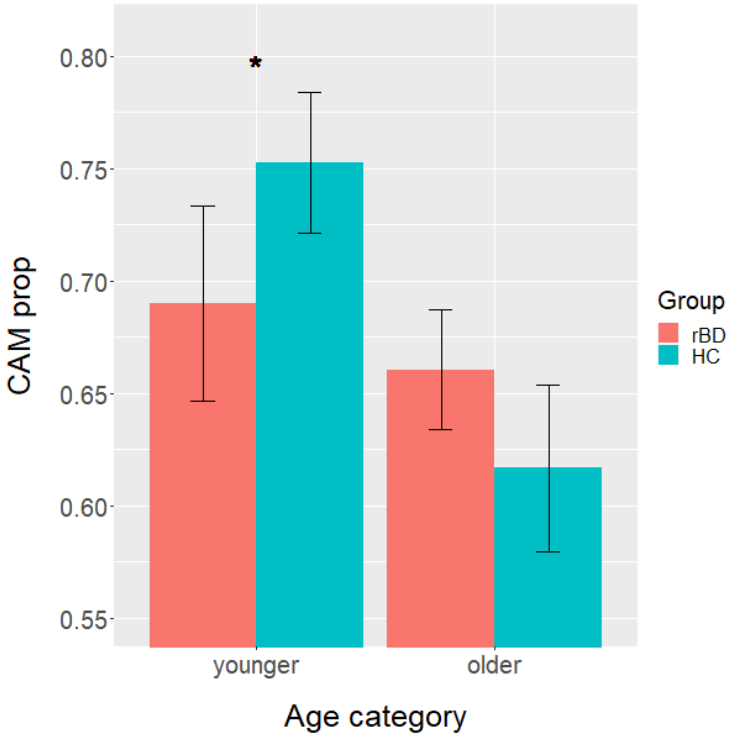

*Note.* CAM = Cambridge Mindreading Face-Voice Battery (Golan et al., 2006), rBD = patients with remitted bipolar disorder, HC = healthy control group

To further examine this interaction we created a new factor *age category* with the help of a nonparametric regression estimator based on local linear regression with adaptive bandwidths (Friedman, 1984). This estimation revealed that the participants could empirically be divided into two age categories: *younger* (<45 years) and *older* (≥45 years). An ANOVA with age category as the predictor instead of the centered age variable revealed similar results: Again, neither the main effects of group [$F(1,80)=0.23$, $p=.636$] and MI [$F(1,80)=0.01$, $p=.937$] nor the interaction between group and MI [$F(1,80)=0.08$, $p=.779$] were significant. However, the main effect of age category was significant [$F(1,80)=17.08$, $p<.0001$, $\eta_G^2=.12$]. Again, this was qualified by a significant interaction

between group and age category [$F(1,80)=7.05$, $p=0.010$, $\eta_G^2=.048$]. We conducted further mixed ANOVAs for the younger and the older age category with group and MI as predictors and $CAM_{prop}$ as the outcome variable. In neither of the two age categories was there a significant main effect of MI [younger: $F(1,31)=0.18$, $p=.671$; older: $F(1,49)=0.39$, $p=.537$] or a significant interaction between MI and group [younger: $F(1,31)=0.37$, $p=.546$; older: $F(1,49)=1.32$, $p=.256$]. There was, however, a significant main effect of group in the younger category [$F(1,31)=4.21$, $p=.049$, $\eta_G^2=.083$] that was not present in the older category [$F(1,49)=2.94$, $p=.093$, $\eta_G^2=.038$]. T-tests indicated that only within the younger age group was $CAM_{prop}$ significantly higher in HCs ($M=.75$, $SD=.09$) than in rBDs ($M=.69$, $SD=.12$) (see Figure 4, and Table 6).

**Figure 4**

*Mean CAM proportion score ($CAM_{prop}$) separately for rBD versus HC within the younger (< 45 years) vs. the older (≥ 45 years) age category*




*Note.* Figure 4 shows that only within the younger age category is the difference in $CAM_{prop}$ between BD vs. HC significant. CAM = Cambridge Mindreading Face-Voice Battery (Golan et al., 2006), rBD = patients with remitted bipolar disorder, HC = healthy control group.

*ToM reasoning*

In terms of $MASC_{prop,}$ there was no significant main effect of group [$F(1,80)=0.04$, $p=.850$] or MI [$F(1,80)=1.68$, $p=.198$] but there was an effect of age [$F(1,80)=11.35$, $p=.001$, $\eta_G^2=.097$]. Furthermore, there were no significant interaction effects between group and MI [$F(1,80)=.03$, $p=.864$] or between group and age [$F(1,80)=1.93$, $p=.168$]. A post-hoc correlation showed that MASC performance decreased significantly with increasing age across the whole sample ($r=-0.35$, $p=.001$).

In terms of $MASC_{cog,}$ we also found no significant main effect of group [$F(1,80)=0.14$, $p=.71$] or MI [$F(1,80)=0.01$, $p=.93$] but an effect of age [$F(1,80)=4.62$, $p=.035$, $\eta_G^2=.038$]. Furthermore, there were no significant interaction effects between group and MI [$F(1,80)=1.45$, $p=.231$] or group and age [$F(1,80)=1.15$, $p=.238$]. Again, $MASC_{cog}$ performance decreased significantly with increasing age within the whole sample ($r=- 0.24$, $p=.030$).

For $MASC_{emo}$ we found no significant main effect of group [$F(1,80)=0.60$, $p=.439$] but an effect of age [$F(1,80)=14.85$, $p<.001$, $\eta_G^2=.103$] and MI [$F(1,80)=3.98$, $p<.049$, $\eta_G^2=.019$]. The interaction effects between group and MI [$F(1,80)=0.80$, $p=.374$] as well as between group and age [$F(1,80)=1.72$, $p=.193$] were not significant. The main effect of age again indicated that $MASC_{emo}$ performance decreased significantly with increasing age in the whole sample ($r=-0.39$, $p<.001$). The main effect of MI indicated that participants performed better without MI ($M=.787$, SD=.156) than with MI ($M=.745$, SD=.184).

Means and standard derivations are displayed in Table 5 and Table 6.

**Table 5**

*Means and standard deviations for the characteristic values of CAM and MASC, separately for the both mood induction conditions within rBD, HCs and the total sample*

| | rBD (*N* = 44) | | | HC (*N* = 40) | | | Total (*N* = 84) | |
|---|---|---|---|---|---|---|---|---|
| | MI | no MI | total | MI | no MI | total | MI | no MI |
| **CAM** | | | | | | | | |
| $CAM_{prop}$ | .669 (.115) | .674 (.101) | .671 (.108) | .679 (.120) | .670 (.141) | .674 (.130) | .674 (.117) | .672 (.121) |
| **MASC** | | | | | | | | |
| $MASC_{prop}$ | .762 (.121) | .780 (.120) | .771 (.120) | .763 (.129) | .776 (.114) | .770 (.121) | .763 (.124) | .778 (.117) |
| $MASC_{cog}$ | .762 (.132) | .780 (.122) | .771 (.127) | .792 (.148) | .771 (.145) | .782 (.146) | .776 (.140) | .776 (.133) |
| $MASC_{emo}$ | .762 (.193) | .786 (.171) | .774 (.181) | .726 (.174) | .787 (.140) | .757 (.160) | .745 (.184) | .787 (.156) |

*Note*. rBD = patients with remitted bipolar disorder, HC = healthy control group, MI = condition with sad mood induction, CAM = Cambridge Mindreading Face-Voice-Battery, $CAM_{prop}$ = CAM proportion score, MASC = Movie for the Assessment of Social Cognition, $MASC_{prop}$ = MASC proportion score, $MASC_{cog}$ = MASC proportion score for the cognitive state modality, $MASC_{emo}$ = MASC proportion score for the affective state modality

**Table 6**

*Means and standard deviations for the characteristic values of CAM and MASC in the younger and the older subsample, separately for rBD and HCs and both mood induction conditions*

| | **Younger subsample (*N* = 33)** | | | | | | | | **Older subsample (*N* = 51)** | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **rBD (*N* = 16)** | | | **HC (*N* = 17)** | | | **total (*N* = 33)** | | **rBD (*N* = 28)** | | | **HC (*N* = 23)** | | | **total (*N* = 51)** | |
| | MI | no MI | total | MI | no MI | total | MI | no MI | MI | no MI | total | MI | no MI | total | MI | no MI |
| **CAM** | | | | | | | | | | | | | | | | |
| $CAM_{prop}$ | .692 (.137) | .688 (.107) | .690 (.121) | .741 (.082) | .764 (.10) | .753 (.090) | .717 (.113) | .727 (.108) | .655 (.101) | .665 (.099) | .660 (.099) | .633 (.124) | .600 (.126) | .617 (.125) | .645 (.111) | .636 (.116) |
| **MASC** | | | | | | | | | | | | | | | | |
| $MASC_{prop}$ | .814 (.093) | .831 (.120) | .823 (.106) | .812 (.134) | .791 (.101) | .801 (.117) | .813 (.114) | .810 (.111) | .732 (.126) | .751 (.112) | .742 (.119) | .727 (.115) | .766 (.124) | .747 (.120) | .730 (.120) | .758 (.117) |
| $MASC_{cog}$ | .805 (.108) | .825 (.097) | .815 (.101) | .832 (.155) | .795 (.143) | .814 (.148) | .819 (.133) | .810 (.122) | .737 (.140) | .754 (.129) | .745 (.134) | .763 (.138) | .754 (.147) | .758 (.141) | .749 (.138) | .754 (.136) |
| $MASC_{emo}$ | .833 (.119) | .850 (.194) | .841 (.159) | .791 (.173) | .784 (.104) | .788 (.141) | .811 (.149) | .816 (.155) | .721 (.216) | .750 (.147) | .736 (.184) | .678 (.161) | .789 (.165) | .734 (.171) | .702 (.193) | .768 (.155) |

*Note*. rBD = patients with remitted bipolar disorder, HC = healthy control group, MI = condition with sad mood induction, CAM = Cambridge Mindreading Face-Voice-Battery, $CAM_{prop}$ = CAM proportion score, MASC = Movie for the Assessment of Social Cognition, $MASC_{prop}$ = MASC proportion score, $MASC_{cog}$ = MASC proportion score for the cognitive state modality, $MASC_{emo}$ = MASC proportion score for the affective state modality

### 3.2.4 Discussion

This study is the first to examine whether rBD is associated with complex FER impairments using ecologically valid stimulus material. Regarding ToM reasoning, we aimed to replicate the findings by Montag et al. (2010) concerning $MASC_{cog}$ and $MASC_{emo}$ and beyond that, to compare the overall performance in MASC between rBD and HC ($MASC_{prop}$). Furthermore, it is the first study to investigate variables that might help to explain the previously reported inconclusive results concerning ToM in rBD – namely the influence of a sad mood induction, age, and gender.

The rBD group as a whole did not perform worse than HC group in FER as assessed by the CAM. However, we found differences between the younger subsample of rBDs and the younger subsample of HCs. This is based on a significant decrease in CAM performance with age that is only present in HCs but not in rBDs. There were no significant group differences in MASC scores – neither independently nor in interaction with age. However, we found that mental state reasoning performance ($MASC_{prop,}$ $MASC_{emo,}$ $MASC_{cog}$) decreased significantly with increasing age and that reasoning concerning emotions ($MASC_{emo}$) was impaired by negative MI in both rBDs and HCs. While age and MI both appear to influence ToM, gender did not affect decoding or reasoning.

The finding that there was no overall difference in CAM performance between rBDs and HCs is in line with several other studies (e.g., Bora et al., 2005; Hulvershorn et al., 2012; Shamay-Tsoory, Harari, et al., 2009; Vaskinn, Sundet, Friis, Simonsen, Birkenæs, et al., 2007; Venn et al., 2004). However, other studies found HCs to be superior in FER compared to rBDs (e.g., Bio, Soeiro-de-Souza, Otaduy, Machado-Vieira, & Moreno, 2013; Derntl et al., 2009; Hoertnagl et al., 2011; Lahera et al., 2012; Neves et al., 2015). Of note, there is one fundamental difference between the current study and former studies that limits comparability: while the other studies used static photographs of basic facial emotional expressions, we used dynamic displays of complex emotions. Thus, even though the current study suggests that rBDPs do not differ from HCs in decoding dynamic complex emotions, as long as one ignores the age of the participants, rBDPs may still have difficulties in correctly identifying static expressions of basic emotions – irrespective of their age.

It is known that not only does the decoding of basic emotions differ from that of complex emotions (Mill et al., 2009; Takahashi et al., 2004), but also that the decoding of dynamic stimuli differs from that of static stimuli with respect to information content, processing requirements, recruited brain areas, and realism (Adolphs, Tranel, & Damasio, 2003; Weyers et al., 2006). Furthermore, it is disputable as to how relevant these formerly reported deficits are in the everyday life of patients. Another difference between the current study and preceding studies is the extent to which the FER task places demands on processing speed. In some previous studies, which showed an overall superiority of HCs in FER, participants were instructed to identify the emotions as fast as possible (e.g., Derntl et al., 2009; Hoertnagl et al., 2011) and/or were confronted with the facial

expressions only for short time periods (Almeida et al., 2010; Hoertnagl et al., 2011). In contrast, the videos in this study lasted comparably longer, and there was no time constraint. This might be relevant as there are not only studies indicating impairments in processing speed in (remitted) BDs (e.g., Bora, Yücel, et al., 2009; Calhoun & Mayes, 2005; Mur, Portella, Martinez-Aran, Pifarré, & Vieta, 2008; Torres, Boudreau, & Yatham, 2007), but also one study showing that the basic FER ability of BDs decreases under time constraints (Lawlor-Savage et al., 2014).

Interestingly, we found impaired complex FER performance in the subgroup of younger rBDs as compared to younger HCs. To our knowledge, there is only one other study to date that has taken the interaction between group and age into account. In this study, Wegbreit et al. (2015) also found that basic FER impairments were more evident in younger than in older BDs. Thus, one might conclude that younger rBDs are impaired in FER but as they age, no longer differ from HCs. However, it should be kept in mind that the age categorization in the current study was undertaken empirically, i.e., without prior theoretical assumptions. Therefore, this age categorization cannot be generalized. Nevertheless, it provides a promising starting point for future studies that are necessary to systematically investigate the influence of age on the decoding of complex emotional facial expressions in BDs. Alternatively, there is also another possible explanation for the result that only younger rBDPs differ from HCs, but not older patients: Figure 3 indicates that while ToM decoding ability deteriorates in HCs as age progresses there is no such deterioration in rBDs. This raises the question whether there might be factors preventing rBDs from showing a normal age-related decline in FER. Thus, future studies are not only required to replicate that only younger rBDs differ from HCs with respect to ToM decoding, but also whether it can be replicated that rBDs do not show the typical age-related deterioration of FER found in HCs and whether there are specific factors preventing rBDs from showing this age-related decline.

We did not find reasoning deficits in rBDs as compared to HCs. Neither irrespective nor in relation to age and MI. This is comparable with the finding of Donohoe et al. (2012) that BD have difficulties with ToM decoding but not with ToM reasoning. They concluded that the decoding deficits in BD are only mild and can therefore be compensated by contextual information, resulting in adequate reasoning performance. This is a possible explanation for the results of our study as well. However, this finding is not only contrary to a number of other studies that have used different tasks (e.g., Lahera et al., 2008; Martino et al., 2011), but also to studies that also used the MASC. Results are available from three other studies (Aidelbaum & Goghari, 2022; Montag et al., 2010; Santos et al., 2017) that have also used the MASC to investigate ToM reasoning deficits in rBD. Thus, including our study, there are two studies that found no impairment in MASC performance in rBD patients (Aidelbaum & Goghari, 2022; Feyerabend et al., 2018) and two studies that found impairment in rBD patients (Montag et al., 2010; Santos et al., 2017) compared to HC. We proposed differences in the

sample characteristics between the study of Montag et al. (2010) and our study as an explanation for the lack of significant effects (Feyerabend et al., 2018). While the sample of the current study consisted of patients with bipolar I and bipolar II disorder, Montag et al. (2010) as well as Santos et al. (2017) exclusively investigated bipolar I patients. Given that the latter have been shown to be more severely impaired than patients with bipolar II disorder with respect to cognitive domains (Torrent et al., 2006) and social functioning (Judd et al., 2008), the different findings of the two studies are perhaps not that surprising. Aidelbaum and Goghari (2022) also discussed this explanation but rather rejected it because their sample consisted almost entirely of rBD I patients (BD I = 24; BD II = 2) and yet this sample had comparable MASC scores to HCs. Nevertheless, it should be mentioned that in view of the multifactorial nature of whether a ToM deficit is present or not, a moderating effect of the diagnostic category (BD I vs. BD II) cannot be definitely excluded. Our alternative explanation, that has also more support from Aidelbaum and Goghari (2022) refers to differences in the functioning of the patients. For there is evidence of relations between social and nonsocial functioning within rBD (e.g., Lahera et al., 2012; Wolf et al., 2010), high levels of functioning may correlate with more preserved social or nonsocial cognitive abilities (Aidelbaum & Goghari, 2022). Unfortunately, Montag et al. (2010), did not report the functioning levels of the patients. However, the comparison of the other three studies regarding functioning supports this explanation. The sample of Santos et al. (2017) showed significantly reduced cognitive functioning compared to HC. In contrast, the majority of our rBD sample had a high level of social functioning when applying the cutoff suggested by Martinez-Aran et al. (2007) to the *General Assessment of Functioning scale* (GAF; American Psychiatric Association, 1994), and above this, the overall impairments in functional abilities (performance on cognitive tasks) were relatively low. The sample of Aidelbaum and Goghari (2022) also had relatively high levels of general functioning, as patients reported average levels of functioning based on their *Functional Assessment Short Test* (FAST) scores (Rosa et al., 2007) that were just above the thresholds for functional impairment (i.e., Total FAST score > 11; Rosa et al., 2007). Lahera et al. (2012) explicitly investigated and reported an association between social cognition and global functioning in euthymic BD. Due to a rather homogeneous BD group in terms of the level of psychosocial functioning, we were not able to investigate whether the missing group difference may be explained by the high level of social functioning. Thus, future studies are needed to investigate whether the findings of Lahera et al. (2012) can be replicated for ecologically valid stimulus material and whether bipolar I and bipolar II patients differ with respect to their ToM abilities.

Furthermore, it should also be noted that studies on ToM with highly ecologically valid material have already been carried out in schizophrenic patients (e.g., Martinez et al., 2017; Montag et al., 2011). In contrast to the existing findings, a clear restriction of performance was found in this

patient group. For future studies in rBD, it is therefore recommended to distinguish between rBD with and without psychotic symptoms.

In accordance with our hypothesis and with other studies (Chepenik et al., 2007; McKinnon et al., 2010), we found that MI negatively influenced ToM reasoning. However, this was only the case for $MASC_{emo}$ scores. Furthermore, this effect was not specific to rBDs, but also present in HCs. Nonetheless it is important to know that negative mood impairs the ability of rBDs to correctly identify what other people feel in certain situations as this might not only lead to social misunderstandings but may thereby also be one contributing factor to a depressive downward spiral.

Contrary to Vaskinn, Sundet, Friis, Simonsen, Birkenæs, et al. (2007), we found no influence of gender, neither in HCs nor in rBDs. Again, this might be due to different tasks that have been used and that Vaskinn et al. studied basic FER while we studied complex FER, i.e., ToM decoding. In line with our findings, Golan, Baron-Cohen, Hill, and Rutherford (2007), who used the RMET (ToM decoding), also found no gender differences. This supports the idea that there is no difference between male BD and female BD when it comes to complex emotion recognition (ToM decoding). However, future studies are needed to replicate this finding as there have been few studies which have investigated complex FER and even less studies doing so using ecologically valid stimulus material. Surprisingly, there was also no significant effect of gender for ToM reasoning ($MASC_{prop}$), either as a main effect or as an interaction effect. This is noteworthy for several reasons. First, other studies that have used the MASC have also found gender differences (better performance for females) in non-clinical or clinical samples (e.g., Isaksson et al., 2019). Second, in our Study 1 (see Chapter 3.1), we found that females outperform males on MASC scores, equally in the UD samples and in the HC sample. This could be due to differences in sample characteristics. Firstly, in the present study, the percentage of women was lower (around 63.1% women in the present study vs. 80.0% women in the study by Feyerabend et al. (2018)). Secondly, the average age in the present study was higher (46.7 years versus 42.9 years). Since ToM performance is a sensitive measure that is influenced by numerous variables, these differences in demographic variables could influence whether effects on ToM are found or not.

In summary, our findings suggest that – when using ecologically valid stimuli – rBDs are not impaired with respect to ToM reasoning, but that younger rBDs are impaired with respect to the decoding of dynamic facial expressions of complex emotions. This is of high relevance since a number of important social steps have to be undertaken before the age of 45 (e.g., career, family formation and the establishment of a stable social network). Thus, it would be important to investigate whether impaired ToM decoding during these years has an influence on the social functioning and/or the course of the illness of BDs. In fact, for patients with major depression it has already been shown that

ToM deficits are a risk of relapse (Inoue et al., 2006). If this is also shown in BDs, it would be important to target ToM deficits in psychotherapy and relapse prevention.

## 3.3 Study 3: Theory of Mind in remitted bipolar disorder: Decoding predicts relapse[11]

Even though BD has a high relapse rate, only few reliable risk factors have been identified so far. This study investigates whether ToM decoding and ToM reasoning as well as mood-linked changes in ToM predict the course of BD.

### 3.3.1 Introduction

BD has a high relapse rate (e.g., Gitlin, Swendsen, Heller, & Hammen, 1995), and the number of episodes is associated with poorer cognitive performance (e.g., López-Jaramillo et al., 2010) and impaired inter-episode functioning (e.g., MacQueen et al., 2000). Hence, there is a strong need to identify predictors of relapse. So far, only few reliable risk factors have been identified (e.g., Inoue et al., 2006; Yamada et al., 2015). Most of them are associated with the psychopathology itself including *residual symptoms during remission* (Altman et al., 2006; Judd et al., 2008; Perlis et al., 2006; Treuer & Tohen, 2010), *number of previous depressive episodes* (Altman et al., 2006; Judd et al., 2008), and *comorbidity* (e.g., Treuer & Tohen, 2010). However, *education* as well as *perceived social support* has also been shown to predict the course of illness (Altman et al., 2006; Cohen, Hammen, Henry, & Daley, 2004; Johnson, Lundström, Åberg-Wistedt, & Mathé, 2003).

Interestingly, BD is associated with deficits in ToM decoding and ToM reasoning – not only during acute episodes but also during remission (see Chapter 1.5 and Chapter 3.2). It is assumed that this causes misunderstandings and problems in interpersonal relationships, which in turn can cause social distress (Inoue et al., 2006; Yamada et al., 2015). Since social distress predicts the course of BD (e.g., Miklowitz, Simoneau, Sachs-Ericsson, Warner, & et al., 1996), and might be associated with social cognition, we hypothesize that deficits in social cognition affect the clinical outcome of patients with rBD.

Even though ToM decoding and ToM reasoning appears to be deficient in rBD (see Chapter 1.5), there are only very few studies investigating ToM as potential predictor of the course of illness (see Chapter 1.5.3). In brief, regarding decoding, there is only one study by Bouhuys et al. (1999), who investigated basic FER in depressive UD and BD inpatients by using static, schematic line drawings of positive and negative emotional facial expressions in different developmental stages and found that the attribution of negative emotions to ambiguous faces (negativity bias) was associated with relapse. As to ToM reasoning, rUD and rBD patients who failed to correctly answer a second order false belief question (this requires predicting that one person has a false belief about what another person beliefs) had more relapses in a 1-year follow-up than patients who answered

[11] Large parts of this Chapter 3.3 were taken verbatim from the original submitted paper.

correctly (Inoue et al., 2006). Of note, both Bouhuys et al. (1999) and Inoue et al. (2006) did not conduct seperate anlayses for patients with UD and with BD, it cannot be ruled out that these results can be traced back only to the UD subgroup. Furthermore, Bouhuys et al. (1999) used static drawings of positive, negative, and ambiguous emotional facial expressions instead of dynamic and subtle displays of complex emotional expressions shown by real persons. Similarly, Inoue et al. (2006) used stimulus material that showed drawings rather than real people, was not dynamic, and did not contain multiple sources of information (e.g., facial expressions, spoken content, knowledge). Thus, the main goal of this study is to expand these findings by using more ecologically valid measures of ToM to examine its role in the course of BD. Our primary hypothesis is that relapse in BD is predicted by ToM decoding and ToM reasoning deficits. In addition, we also consider results of previous studies revealing perceptual biases in patients with affective disorders. Hence, the second goal of this study is to investigate whether decoding biases (i.e., positivity and negativity bias) predict relapse in rBD.

Current (induced) negative mood has been shown to influence ToM both in healthy samples (Chepenik et al., 2007) and in rBD (e.g., Feyerabend et al., 2018; McKinnon et al., 2010). In addition, in the context of UD it has been shown that mood-linked cognitive changes predict relapse in recovered patients. Segal and colleagues, for example, found that UD patients who showed a heightened cognitive reactivity (e.g., presence of dysfunctional cognitions or attitudes) to a mood induction had a higher relapse rate (Segal, Gemar, & Williams, 1999; Segal et al., 2006). Thus, we further want to investigate whether the extent of mood-linked changes in ToM predict relapse in rBD. Comparable to depressed patients (Segal et al., 2006) it is possible that it is not the ToM deficit post recovery per se that characterizes the vulnerability of rBD, but the ease with which it can be aggravated by negative mood. Therefore, we hypothesize that the extent of mood-linked changes in ToM (including negativity/positivity bias) is also predictive of the course of rBD.

Finally, we investigate whether the predictive value of ToM and mood-linked changes in ToM is still present when controlling for the above mentioned risk factors of relapse that have already been shown to be relevant for the course of BD. In summary, in the present study we made the following hypotheses:

1) Relapse in BD is predicted by ToM decoding and ToM reasoning deficits when using ecologically valid measures of ToM.
2) Decoding biases (i.e., positivity and negativity bias) predict relapse in rBD.
3) The extent of mood-linked changes in ToM (including negativity/positivity bias) is also predictive of the course of rBD.
4) The predictive value of ToM and mood-linked changes in ToM is still present when controlling for the risk factors of relapse that have already been shown to be relevant for the course of BD.

## 3.3.2 Methods

### 3.3.2.1 Sample

*N* = 44 rBD were recruited. This sample is the identical sample as in Study 2. However, for this study, we excluded four participants from data analyses as they did not attend the follow-up examination resulting in final sample size of *N* = 40. Information on the inclusion and exclusion criteria can be found in the methods section of the second study (see Chapter 3.2.2.1). The criteria for remitted bipolar I disorder were met by 57.5% of the patients, that of remitted bipolar II disorder by 42.5%. No participants met criteria for cyclothymia. At the time of testing 42.5% had one or more co-morbid mental illness(es) and 95.0% were on various medication(s).

During the study, patients were divided into two *groups* depending on whether they had a relapse (manic or depressive episode) or not during the follow-up period. See Table 7A for sample characteristics, separately for rBD patients who relapsed and rBD patients who did not.

### 3.3.2.2 Materials & Procedure

*Assessment of symptoms and neurocognitive functioning*

To assess the absence of depressive and (hypo-)manic symptoms as well as the absence of significant neurocognitive impairments, we proceeded as described in Chapter 3.2.2.2. There were no significant group differences in the clinical and neurocognitive characteristics (see Table 7B & C).

*Assessment of ToM decoding*

The assessment of ToM decoding ability was performed via the CAM face task (see Chapter 2.1). To investigate the influence of MI on decoding, we divided the CAM into two parts (22 and 23 film clips, respectively) to present one part with, and one without negative MI. Also, we assigned the items to these two parts with respect to item difficulty, valence and emotion concept based on the results of a pre-test (Dolde, 2012). Five film clips showing poor discriminatory power were presented as practice trials. Film clips were presented in a randomized order. We calculated the percentage of correct answers for the parts presented *without* ($CAM_{NMI}$) and *with MI* ($CAM_{MI}$). Furthermore, we calculated the percentage of the *positivity* and the *negativity bias* for both MI-conditions (CAM-$Positivity_{MI}$; CAM-$Negativity_{MI}$; CAM-$Positivity_{NMI}$; CAM-$Negativity_{NMI}$). Positivitiy and negativity biases were calculated based on the number of incorrectly chosen distractors that – according to a preliminary study (Choudhery, 2012) – were more positive or more negative than the correct answer.

**Table 7**

*Demographic, clinical and neurocognitive variables, separately for BD patients with relapse and BD patients without relapse*

| Variable | Patients with relapse *N* = 13 | Patients without relapse *N* = 27 | *df* | $\chi^2$ | *p* |
|---|---|---|---|---|---|
| *A) Demographic variables* | | | | | |
| Level of education (Abitur) in % | 46.15 | 70.37 | 1 | 2.20 | .169 |
| Gender (female) in % | 61.54 | 62.96 | 1 | 0.01 | 1.00 |
| Current living condition in % | | | 3 | 4.84 | .206 |
| living alone | 38.46 | 33.33 | | | |
| living together with life partner | 38.46 | 55.56 | | | |
| living together with parents/relatives | 23.08 | 3.70 | | | |
| other | 0.00 | 7.41 | | | |
| Current social contact in % | | | 4 | 1.75 | .865 |
| more than 1 time per week | 53.85 | 59.26 | | | |
| not more than 1 time per week | 30.77 | 18.52 | | | |
| once in two weeks | 7.69 | 14.81 | | | |
| once per month | 7.69 | 3.70 | | | |
| only in hallway or at work | 0 | 3.70 | | | |
| | **M (SD)** | **M (SD)** | ***df*** | ***t*** | ***p*** |
| Age | 45.54 (14.59) | 47.81 (11.85) | 19.91 | 0.49 | .629 |
| *B) Clinical variables* | | | | | |
| HDRS17-Score | 3.46 (2.60) | 2.07 (2.30) | 21.36 | -1.64 | .116 |
| IDS-C30-Score | 6.69 (6.07) | 4.26 (4.60) | 18.87 | -1.28 | .217 |
| YMRS-Score | 0.85 (1.82) | 0.85 (1.57) | 21.16 | 0.00 | 1.00 |
| BDI-Score | 7.50 (6.23) | 3.92 (4.43) | 16.58 | -1.79 | .092 |
| MSS-Score | 4.15 (4.08) | 1.74 (2.97) | 18.35 | -1.90 | .072 |
| *C) Neurocognitive variables* | | | | | |
| MWT-B (IQ-score) | 117.15 (20.03) | 118.59 (15.05) | 18.77 | 0.23 | .821 |
| TMT-A (seconds) | 32.77 (13.63) | 31.44 (11.56) | 20.61 | -0.30 | .766 |
| TMT-B (seconds) | 64.38 (24.93) | 82.93 (43.38) | 36.60 | 1.71 | .096 |
| *VLMT* | | | | | |
| short-term verbal retentiveness | 59.08 (5.79) | 55.78 (9.64) | 35.88 | -1.34 | .187 |
| memory loss after delayed recall | 12.46 (1.81) | 11.04 (2.90) | 35.22 | -1.90 | .066 |
| corrected recognition | 14.31 (0.85) | 14.52 (0.75) | 21.27 | 0.76 | .456 |

*Note*. BD = bipolar disorder, HDRS17 = Hamilton Depression Rating Scale, IDS-C30 = Inventory of Depressive Symptomatology, YMRS = Young Mania Rating Scale, BDI = Beck Depression Inventory, MSS = Self-Report Manic Inventory, MWT-B = Multiple Choice Word Fluency Test, TMT = Trail Making Test, VLMT = Verbal Learning and Memory Test

*Assessment of ToM reasoning*

We used the MASC (see Chapter 2.2) to evaluate subtle mindreading difficulties. To investigate the effect of MI, we divided the MASC into two parts. The first part contained the first 23, the second part the remaining 22 items. To avoid memory effects, the first part was shown again as an introduction at the second test session. For MASC, we calculated percentage scores for each part, defined as number of correct answers in the respective MASC part divided by the total number of items in the respective MASC part multiplied by 100. Depending on whether MI had been performed in the respective MASC part or not, we assigned the MASC-scores to the conditions *with* ($MASC_{MI}$) and *without MI* ($MASC_{NMI}$).

*Negative mood induction*

To induce negative mood we applied a 10-minute and a 5-minute version of Peer Gynt suite no. 1 op. 46 "The Death of Ase" from Grieg while participants were asked to reflect on a sad personal experience (see Chapter 2.4).

*Positive and Negative Affect Schedule (PANAS)*

For verification of a successful MI, we evaluated 5 items of the PANAS repeatedly: *worried*, *elated*, *bad*, *content*, *sad*. Participants were asked to evaluate the extent to which they are experiencing these affects at that moment on a Likert Scale (1 = *very light*, 5 = *very strong)*.

*Assessment of control variables*

Based on the literature we controlled for: 1) *presence of residual symptoms during remission* (ResSym) as assessed by HDRS17 score, 2) *number of previous depressive episodes* (NumDep) as assessed by the SCID interview, 3) *number of comorbidity* (Com) as assessed by the SCID interview, 4) *education* (Edu), i.e., whether or not participants have a secondary degree (Abitur vs. no Abitur), assessed by an anamnestic interview and 5) *perceived social support* (pSoSup), as assessed by the *social support questionnaire* (F-SozU; Fydrich et al., 2007). The HDRS17 and the F-SozU, the methods used to evaluate ResSym and pSoSup, are described in more detail in the following:

HDRS17: The 17-item *Hamilton Depression Rating Scale* ($HDRS_{17}$) is a diagnostic tool for determining the severity of depression in adults, based on symptoms experienced in the past week. It was introduced by Hamilton (1960) and has been revised several times, most recently by Hamilton (1980). The measurement is based on external assessment by a diagnostician during an interview with the patient. It is commonly used in clinical trials but also in clinical practice (Cusin, Yang, Yeung, & Fava, 2009). There are also other versions with 21 ($HDRS_{21}$) and 24 ($HDRS_{24}$) items. Diagnosticians are to rate the items on a point scale of 0 to 4 or 0 to 2 to indicate how severe a particular symptom is. These include, for example, querying mood, feelings of guilt, suicidal thoughts, insomnia, agitation or retardation, anxiety, weight loss, and somatic symptoms. The sum of the 17 individual scores is the total score, which ranges from 0 to 54. A total score between 0 and 8 is considered normal,

without clinical signs, or remitted. Total scores between 9 and 16 indicate mild depression, between 17 and 24 indicate moderate depression, and 25 and above indicate severe depression. The internal consistency (Cronbach's α) of $HDRS_{17}$ is 0.83 (Rush et al., 2003) and the inter-rater reliability is also very high (0.80-0.98) for the HDRS total scores (Cusin et al., 2009). Reported concurrent validity with global measures of depression severity range from 0.65 to 0.90 and is also high for clinician-rated measures (e.g., IDS) (Hamilton, 2000).

F-SozU: The F-SozU is a self-assessment tool to measure the subjective conviction of receiving support and help form one's social network when needed, as well as the assessment of being able to draw on resources from the social environment. The F-SozU consists of two parts – part A and part B. Part A of the 54-item standard version includes 4 main scales: a) *Emotional Support* (EU), b) *Practical Support* (PU), c) *Social Integration* (SI), and d) *Social Distress* (including Overwhelm, Overprotection, and Rejection) (Bel). Scoring of all items from the EU, PU and SI scales form an *overall measure of perceived social support* (WasU). Part A also contains 3 additional scales: *Reciprocity* (Rez), *availability of confidants* (Vert), *satisfaction with social support* (Zuf). The items are phrased as statements (e.g., "If I am ever deeply depressed, I know who I can go to."). Participants use a 5-point Likert scale to indicate their level of agreement with a particular statement. Part B asks about specific people perceived as supportive or burdensome, thus providing structural information about the participant's social network (e.g., "Who can you trust completely?"). Internal consistency (Cronbach's α) ranges from α = .81 to α = .93 for main scales and from α = .70 to α = .84 for addional scales. Retest-reliability is .84. Good factorial, differential, convergent, and discriminant validity has been reported, including correlations with psychopathological symptoms and also with personality factors and social competence (e.g., Franke, 1994; Sommer & Fydrich, 1991). Short versions with 22 items (Fydrich, Sommer, Menzel, & Höll, 1987) and 14 items (Fydrich, Sommer, Tydecks, & Brähler, 2009) are also available. In the present study, only the WasU scale (pSoSup) was used.

*Assessment of the course of the illness*

We conducted the LIFE by Keller et al. (1987) about 39 weeks after the assessment of ToM performance. In the present study, we determined whether a patient experienced a relapse within the follow-up period. Relapse was defined as the presence of all DSM-IV criteria for a depressive episode or a (hypo-)manic episode after a period of complete remission (i.e., no or hardly any symptoms for at least 8 weeks) since t0. See Chapter 2.3 for detailed information about the LIFE interview.

*Procedure*

This longitudinal study has a follow-up-interval of 39 weeks. At the diagnostic session (t0) an anamnestic interview, clinical assessments and neuropsychological tests were conducted.

Furthermore, participants filled out the F-SozU.[12] Two testing sessions (t1 & t2) followed on separate days within two weeks after t0. In case of a time period longer than two weeks between t0 and t1 and/or t2, we assessed SIGHD-IDS, YMRS, BDI and MSS again to ensure sustained remission.

In both testing sessions, one part of the CAM was administered, followed by one part of the MASC. The CAM part and the MI condition (yes versus no) were completely balanced and randomly assigned to the participants. The number of patients to be included was determined in advance.

In the MI session we presented the music piece immediately before the CAM and again before the MASC. Participants answered the PANAS items immediately before and after the MIs. In the session without MI, participants completed the PANAS before the CAM as well as before and after the MASC.

The follow-up session (t3) was performed 39 weeks after t2 (*M*=40.08 weeks, *SD*=2.00 weeks, min=39 weeks, max=49 weeks). If t3 was carried out later, only the 39 weeks after t2 were evaluated in the LIFE-interview. During t3, we assessed SIGHD-IDS, YMRS, BDI, MSS and LIFE.

#### 3.3.2.3 Statistical analyses

We performed statistical analyses with R, Version 4. 1. 2 (R Core Team, 2021). Level of significance was $\alpha$<.05. We also provide 95% confidence intervals (95% CI) of the effect size. We report $\eta_G^2$ for the ANOVAs and ANCOVAs, as indicators of effect size.

Firstly, we calculated change scores for CAM, CAM-Positivity, CAM-Negativity and MASC. As the variability among these scores in the condition without MI is a problem to the calculation of simple change scores, we calculated *residualized change scores* in accordance with Segal et al. (2006). Therefore, we used linear regression models in which the ToM scores assessed without MI predict the ToM scores with MI so that the variability among residuals can be considered independent from the scores in the condition without MI. We saved *standardized residuals* (*ZCAM*, *ZCAM-Positivity*, *ZCAM-Negativity* and *ZMASC*). For better interpretability of the respective analyses, we provide the differences of $CAM_{MI}$ minus $CAM_{NMI}$– or $MASC_{MI}$ minus $MASC_{NMI}$, respectively – as descriptives instead of the standardized residuals.

Secondly, to reduce the number of variables to be included in the regression models, we carried out *Welch's two sample t-tests* as preliminary analyses to examine whether ToM performance per se and/or mood-linked changes in ToM performance are associated with relapse. Therefore, we conducted t-tests with Welch's correction to examine whether patients with relapse differed from patients without relapse with respect to the CAM and MASC variables. For variables

[12] Besides the material described in this section we used additional questionnaires in this study that are, however, not relevant to the question that is examined here. Participants were asked to complete the following questionnaires at home: Inventory of Interpersonal Problems (Horowitz et al., 2000); Cognitive Emotion Regulation Questionnaire (Garnefski & Kraaij, 2007); Emotion regulation questionnaire (Gross & John, 2003); Empathy Quotient (Baron-Cohen & Wheelwright, 2004). The respective results are reported elsewhere.

that did not meet the assumption of normal distribution we performed the *Wilcoxon rank-sum test* instead of the Welch t-test. We report *Cohen's d* (*d*) for the t-tests, as indicators of effect size.

Since univariate analysis does not take into account that individual variables that are only weakly associated with the outcome can make a significant contribution when combined in a regression model (Chowdhury & Turin, 2020; Hosmer, David, Lemeshow, & Sturdivant, 2013) we decided to also include variables that only reached marginal significance ($.05 \leq p < .10$).

Thirdly, we conducted logistic regression models to predict relapse to examine whether ToM decoding and/or reasoning still have a predictive value with respect to the course of BD when controlling for the factors that have reliably been found to predict the clinical outcome of BD. In a first step of the regression models, we included the already known predictors: ResSym, NumDep, Edu, Com, pSoSup. In the second step, we extended the regression models by including those variables for which we found that patients who relapsed differed (marginal) significant from patients who did not relapse in the afore described t-tests. Given the widely varying scales, we centered $CAM_{NMI}$, CAM-$Positivity_{NMI}$, NumDep, NumMan, Com and pSoSup for all analyses.

### 3.3.3 Results

Table 8 presents LIFE- and clinical variables for the overall sample and separately for patients with and without relapse.

**Table 8**

*Relevant LIFE- and clinical variables for the overall sample and separately for BD patients who had a relapse versus BD patients without relapse*

| | **Overall *N* = 40** | **Patients with relapse *N* = 13** | **Patients without relapse *N* = 27** | ***df*** | ***t*** | ***p*** |
|---|---|---|---|---|---|---|
| Percentage of relapse | 32.5% | | | | | |
| Ø-percentage (number) of weeks with residual depressive symptoms | 17.69% (7.08) | 19.81% (7.92) | 16.67% (6.67) | 27.64 | -0.38 | .709 |
| Ø-percentage (number) of weeks with full clinical picture of depression | 9.44% (3.78) | 29.04% (11.62) | 00.00% (0) | 12.00 | -4.04 | .002 |
| Ø-percentage (number) of weeks with residual manic symptoms | 3.93% (1.57) | 4.23% (1.69) | 3.80% (1.52) | 22.69 | -0.16 | .877 |
| Ø-percentage(number) of weeks with full clinical picture of mania | 2.08% (0.83) | 6.35% (2.54) | 00.00% (0) | 12.00 | -2.13 | .054 |
| Ø-percentage (number) of symptomatic weeks | 33.13% (13.25) | 59.43% (23.77) | 20.48% (8.19) | 23.92 | -4.10 | <.001 |
| percentage (number) of comorbidity | 42.5% (17) | 61.54% (8) | 33.33% (9) | | | |
| Number of previous depressive episodes | 13.74 | 18.42 | 11.58 | 19.58 | -0.77 | .449 |
| Number of previous manic episodes | 7.92 | 10.25 | 6.85 | 15.20 | -0.58 | .572 |
| Mean HDRS17-Score at t0 | 2.53 | 3.46 | 2.07 | 21.36 | -1.64 | .116 |
| Mean pSoSup-score | 3.63 | 3.37 | 3.75 | 19.76 | 1.50 | .149 |

*Note*. This table includes the characteristics of the t-tests, for the mean value comparisons between patients with and patients without relapse.

LIFE = Longitudinal Interval Follow-up Evaluation (LIFE; Keller et al., 1987), BD = bipolar disorder, HDRS17 = Hamilton Depression Rating Scale, pSoSup = perceived social support score assessed by the social support questionnaire, GAF = Global Assessment of Functioning.

*Manipulation check of MI*

A MANOVA was conducted, with *time* as a four-level within subject factor (pre_1 = *before first MI*, post_1 = *after first MI*, pre_2 = *before second MI*, post_2 = *after second MI*) and relapse as between subject factor and the PANAS item scores (worried, elated, bad, content and sad) as dependent variables.

We found the expected main effect of time [*V*=.831, *F*(3,113)=8.51, *p*<.001]. Mean scores before the first MI varied significantly from mean scores after the first MI in the expected direction [worried: *t*(48.87)=-8.77, *p*<.001, *d*=1.960; elated: *t*(63.17)=5.30, *p*<.001, *d*=1.184; bad*: t*(45.32)=-9.28, *p*<.001, *d*=2.075; content: *t*(77.65)=7.17, *p*<.001, *d*=1.602; sad*: t*(43.04)=-8.82, *p*<.001, *d*=1.972]. Similarly, mean scores before the second MI varied significantly from mean scores after the second MI [worried*: t*(68.16)=-5.15, *p*<.001, *d*=1.151; elated*: t*(55.47)=2.87, *p*=.006, *d*=0.64; bad: *t*(69.70)=-4.73, *p*<.001, *d*=1.059; content: *t*(75.83)=2.62, *p*=.010, *d*=0.587; *sad: t*(51.05)=-6.35, *p*<.001, *d*=1.420].

There was no significant effect of relapse [*V*=.258, *F*(1,37)=2.30, *p*=.067] and no significant interaction between time and relapse [*V*=.180, *F*(3,113)=1.42, *p*=.136].

Means and standard deviations are presented separately for participants with and without relapse in Table 9.

**Table 9**

*Means and standard deviations for the PANAS-items before and after first and second MI separately for participants with and without relapse*

| *PANAS-items* | Participants who relapsed (*N* = 13) | | | | Participants who did not relapse (*N* = 27) | | | |
|---|---|---|---|---|---|---|---|---|
| | First mood induction | | Second mood indcuction | | First mood induction | | Second mood induction | |
| | Pre *M* (*SD*) | Post *M* (*SD*) | Pre *M* (*SD*) | Post *M* (*SD*) | Pre *M* (*SD*) | Post *M* (*SD*) | Pre *M* (*SD*) | Post *M* (*SD*) |
| worried | 1.23 (0.44) | 2.69 (0.95) | 1.38 (0.87) | 2.54 (1.05) | 1.07 (0.27) | 2.41 (0.93) | 1.22 (0.42) | 1.96 (0.76) |
| elated | 1.92 (0.64) | 1.23 (0.44) | 1.69 (1.11) | 1.15 (0.37) | 2.19 (1.00) | 1.22 (0.58) | 1.63 (0.88) | 1.19 (0.48) |
| bad | 1.15 (0.38) | 2.23 (0.93) | 1.62 (0.87) | 2.00 (0.71) | 1.04 (0.19) | 1.74 (0.94) | 1.19 (0.40) | 1.44 (0.64) |
| content | 3.46 (0.66) | 1.85 (0.80) | 3.15 (0.99) | 2.08 (1.04) | 3.63 (0.93) | 2.33 (0.92) | 3.15 (0.86) | 2.81 (1.00) |
| sad | 1.23 (0.44) | 3.38 (1.12) | 1.31 (0.48) | 2.92 (1.19) | 1.00 (0.00) | 2.44 (1.09) | 1.15 (0.36) | 2.00 (0.78) |

*Note.* PANAS = *Positive and Negative Affect Schedule* (PANAS; Crawford & Henry, 2004)

*Mean difference between patients who relapsed and patients who did not regarding ToM decoding and ToM reasoning*

In the Welch's two sample t-test patients who relapsed compared to patients who did not relapse differed significantly in the mean values of $CAM_{NMI}$ [$t$(30.56)=-2.83, $p$=.008, $d$=0.868, 95% CI [0.157, 1.579]], CAM-$Positivity_{NMI}$ [$t$(35.57)=2.10, $p$=.043, $d$=-0.602, 95% CI [-1.299, 0.095]] and differed marginally significant in the mean value of ZCAM [$t$(26.03)=-1.86, $p$=.074, $d$=0.607, 95% CI [-0.090, 1.304]]. Participants who relapsed ($M$=73.70, $SD$=7.88) scored significantly higher in $CAM_{NMI}$ than participants without relapse ($M$=65.33, $SD$=10.37). Furthermore, patients who relapsed ($M$=15.10, SD=6.41) showed significantly less positivity bias (CAM-$Positivity_{NMI}$) than patients who did not ($M$=20.75, SD=10.49). Relapsed and non-relapsed participants also differed marginally significant in their mood-linked changes in decoding (ZCAM ): Both groups showed a mood-linked impairment in overall decoding (ZCAM) which, was in trend more pronounced amongst patients with relapse (M=-0.27, SD=12.32) than amongst patients without relapse (M=-0.19, SD=15.41). We found no significant difference in the mean values of ZCAM-Positivity [$W$=226.5, $p$=.145], CAM-$Negativity_{NMI}$ [$t$(35.39)=1.23, $p$=.227] and ZNEG [$t$(27.36)=0.90, $p$=.376] between patients who relapsed and patients who did not relapse. Regarding the MASC variables there were no group differences – neither for $MASC_{NMI}$ [$W$=125, $p$=.147] nor for ZMASC [$t$(33.01)=-0.25, $p$=.802]. Based on these analyses, we decided to include $CAM_{NMI}$, ZCAM and CAM-$Positivity_{NMI,}$ in the extended regression model to predict relapse and to refrain from including ZCAM-Positivity, CAM-$Negativity_{NMI}$, ZCAM-Negativity, $MASC_{NMI}$ and ZMASC as additional factors (see below).

*Prediction of relapse*

The results of the regression analyses and their comparisons are tabulated in Table 10. Since the two predictors $CAM_{NMI}$ and CAM-$Positivity_{NMI}$ correlate significantly negatively with each other [$r$=-.628, $t$(38)=-4.98, $p$<.001] and the *variance inflation factors* (*VIF*) were high (VIF>10; Kutner, Nachtsheim, & Neter, 2004) pointing to strong linear relationships among these predictors, we decided to calculate two different extended regression models for relapse. In the first step, we included the following predictor variables into the regression model: 1) ResSym, 2) NumDep, 3) Edu, 4) Com, 5) pSoSup. This basal regression model (*Basal model*) was marginal significant but none of the included variables reached significance (see Table 10, *Basal model*). In the second step, we evolved two extended models from the *Basal model*. In the first extended regression model (*CAM model*), we added $CAM_{NMI}$ and ZCAM as sixth and seventh predictors. *CAM model* explained the data significantly better than the *Basal model* and reached significance. Only $CAM_{NMI}$ and ZCAM reached (marginal) significance, whereas the other predictors did not (see Table 10, *CAM model*). In accordance with the above-described t-tests, this indicates that relapse is associated with higher $CAM_{NMI}$ scores as well as tending to higher ZCAM scores. The latter means that relapse tends to be

more likely the more CAM performance decreases following a MI as compared to NMI. In the second extended regression model (*POS model*), we added CAM-Positivity$_{NMI}$ as the sixth predictor. However, *POS model* did not explain the data significantly better than the *Basal model* and did not reach significance. None of the included variables were significant predictors (see Table 10, *POS model*).

**Table 10**

*Results of the regression analyses for the prediction of relapse as well as the comparisons between the basal regression model and the CAM model or POS model respectively*

| | Basal model | | | | CAM model | | | | POS model | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\chi^2(df)$ | 9.47 (5)(*) | | | | 19.59 (7) ** | | | | 11.19 (6)(*) | | | |
| $R^2$ | 0.211 | | | | 0.387 | | | | 0.244 | | | |
| $\Delta R^2(df)$ | | | | | 10.12 (2)** | | | | 1.73 (1) | | | |
| | **β (SE)** | **95% CI for odds ratio** | | | **β (SE)** | **95% CI for odds ratio** | | | **β (SE)** | **95% CI for odds ratio** | | |
| | | **Lower** | **Odds Ratio** | **Upper** | | **Lower** | **Odds Ratio** | **Upper** | | **Lower** | **Odds Ratio** | **Upper** |
| ResSym | 0.19 (0.17) | 0.87 | 1.22 | 1.72 | 0.27 (0.20) | 0.88 | 1.31 | 1.84 | 0.18 (0.17) | 0.84 | 1.20 | 1.72 |
| NumDep | 0.00 (0.02) | 0.97 | 1.00 | 1.04 | 0.04 (0.03) | 0.98 | 1.04 | 1.10 | 0.01 (0.02) | 0.97 | 1.00 | 1.04 |
| Edu | -1.06 (0.88) | 0.05 | 0.35 | 1.92 | -0.64 (1.12) | 0.06 | 0.53 | 5.18 | -0.84 (0.91) | 0.06 | 0.43 | 2.59 |
| Com | 0.60 (0.43) | 0.82 | 1.82 | 4.70 | 0.61 (0.59) | 0.65 | 1.84 | 6.94 | 0.64 (0.44) | 0.65 | 1.90 | 5.27 |
| pSoSup | -0.40 (0.63) | 0.18 | 0.67 | 2.33 | -0.75 (0.78) | 0.09 | 0.47 | 2.10 | -0.38 (0.63) | 0.18 | 0.68 | 2.34 |
| $CAM_{NMI}$ | not included | | | | 0.14* (0.07) | 1.03 | 1.15 | 1.35 | not included | | | |
| ZCAM | not included | | | | 1.33(*) (0.75) | 1.08 | 3.79 | 22.98 | not included | | | |
| CAM-$Positivity_{NMI}$ | not included | | | | not included | | | | -0.06 (0.05) | 0.83 | 0.94 | 1.03 |

*Note.* ResSym = presence of residual symptoms during remission, NumDep = number of previous depressive episodes, Edu = education, Com = number of comorbidity, pSoSup = perceived social support, CAM = face task of the Cambridge Mindreading Face-Voice Battery, $CAM_{NMI}$ = percentage of correct answers for the CAM parts presented without mood induction, ZCAM = residualized change scores for CAM proportion score with versus without mood induction, CAM-$Positivity_{NMI}$ = Percentage of CAM positivity bias in the condition without mood induction, ZCAM-Positivity = residualized change scores for CAM positivity bias with versus without mood induction, Basal model = basal regression model only including ResSym, NumDep, NumMan, Com, pSoSup, CAM model = extended regression model, in which we added $CAM_{NMI}$ and *ZCAM* to the basal regression model, POS model = extended regression model, in which we added *CAM-*$Positivity_{NMI}$ and *ZCAM-Positivity* to the basal regression model.

*p < .05, **p < .01, ***p < .001; (*)p < .10

### 3.3.4 Discussion

To our knowledge, this is the first study to examine whether ToM performance, assessed by an ecologically valid method, predicts relapse of rBD patients within a 9-month follow-up period. It is also the first study investigating the influence of mood-linked changes in ToM on relapse in rBD.

While we found no evidence of ToM reasoning predicting the course of the illness, we found that the ability to decode emotional facial expressions indeed predicts relapse in rBD. However, contrary to our hypothesis, a higher rather than a worse decoding performance was associated with a higher probability of relapse. Moreover, we found partial support for our third hypothesis: Decoding reactivity to MI (*ZCAM*) predicts relapse on a trend level. Furthermore, decoding and mood-linked decoding changes still predict relapse in rBD when controlling for already known risk factors and appear to be even more relevant to the course of BD than the latter.

At first sight, the finding that better decoding performance is associated with relapse is not what we expected. However, this finding is put into perspective when considering the positivity bias: Even though CAM-Positivity$_{NMI}$ is not a significant predictor in the regression model, t-tests showed that a positivity bias is associated with the absence of relapse.

A reduced positivity bias in the relapse-group can be interpreted on the background of studies concerning positivity bias in human cognition. For example, healthy people show a positivity bias when it comes to self-perception and the perception of their environment (Mezulis, Abramson, Hyde, & Hankin, 2004) and it is assumed that this allows for the maintenance of mental health (Mezulis et al., 2004; Taylor & Brown, 1988). Indeed, it has been suggested that the positivity bias is reduced or even absent in people suffering from mental illness, especially depression (Joormann & Gotlib, 2007; Strunk, Lopez, & DeRubeis, 2006; Sweeney, Anderson, & Bailey, 1986). Our results suggest that differences in positivity bias might not only exist between healthy and mentally ill people but also between patients with and patients without relapse in a specific time period. It is possible that a positivity bias in ToM decoding maintains mental health also in people with a lifetime history of BD. One possible explanation could be that a reduced positivity bias in ToM decoding causes interpersonal problems (e.g., less positively evaluated interactions) and thus promotes social stress and consequently a heightened risk of relapse. On the other hand, the presence of a positivity bias might cause more positive evaluations and experiences in relationships as well as more positive interaction styles. This in turn might be associated with less social stress and a reduced risk of relapse. However, it has to be kept in mind that this finding was only present in the preliminary t-tests whereas the CAM positivity bias was no significant predictor of relapse when we took other factors into account in the regression model (*POS model)*. Hence, it only allows for cautious conclusions. First, studies are needed to replicate this trend level finding in a bigger sample. Second,

assumptions of how a reduced positivity bias might lead to an increased risk of relapse have to be investigated in studies that are specifically designed to answer this question.

Interestingly there were no significant results regarding the negativity bias. This contradicts the results of Bouhuys et al. (1999). In their study, affective patients who had relapsed previously showed higher levels of perception of negative emotions for ambiguous faces. However, the ambiguous faces conveyed equal amounts of positive and negative emotions meaning that there was no correct or incorrect answer. Thus, the alleged negativity bias in the study by Bouhuys et al. could also be interpreted as a reduced positivity bias. Future studies are required to clarify this question.

Even though it was only on a trend-level (.05> p <.01), we found that besides decoding performance post-recovery per se, mood-linked changes in decoding might also predict the course of illness. Similar to studies in the context of UD (Segal et al., 2006), we found that a heightened cognitive reactivity to MI is associated with relapse in trend. Specifically, patients who relapsed showed a more pronounced impairment in overall CAM performance following a negative MI. Again, these results have to be treated cautiously as they were only marginally significant. Still, they suggest that the ease with which dysfunctional social cognitions can be brought back to mind under MI is associated with a negative course of BD. This is in line with explanatory models of the development and maintenance of BD (Beck, 1996; C. F. Newman, Leahy, Beck, Reilly-Harrington, & Gyulai, 2002; Scott, 2001), suggesting that activated schemata and modes influence information processing, affect, and behavior by directing individuals towards information consistent with the activated schema.

We did not find differences in the reasoning performance between patients with and without relapse. This contradicts Inoue et al. (2006) who found that patients who relapse are more likely to fail a second order false belief question than patients without relapse. One reason for the differing findings may be the sample characteristics. While we only included rBDs, Inoue et al. included rBD as well as UD patients. Thus, it cannot be ruled out, that their finding can be traced back to the UD subgroup. Moreover, while Inoue et al. used stimulus-material showing drawn, static pictures that did not include multiple sources of information we used ecologically valid material – dynamic displays of real persons in everyday life situations – with the possibility to use context information. Hence, although the current study points out that rBDs who relapse do not differ from patients without relapse in the ability to draw social conclusions using ecologically valid material, there might still be differences when using reasoning tasks of lower ecological validity. In addition, it is possible that group differences have been undetected due to small sample size and low power of the study.

Interestingly, none of the empirically well-founded predictors (i.e., presence of residual symptoms during remission, number of previous depressive episodes, education, number of comorbidity and, perceived social support) revealed significance. Most probably this is due to

differences in sample size and a low test power of the present study as compared to former studies investigating these factors (e.g., Cohen et al., 2004; Perlis et al., 2006).

Our results must be interpreted in consideration of some limitations. Firstly, the data on relapse rely on retrospective statements that are often inaccurate and subject to various biases, both in general (Greenberg & Beck, 1989; Hassan, 2006; J. M. G. Williams, Teasdale, Segal, & Soulsby, 2000) as well as specifically when recalling affect (Ben-Zeev, Young, & Madsen, 2009; Kardum & Daskijević, 2001; Wirtz, Kruger, Scollon, & Diener, 2003) and symptom severity (Schrader, Davis, Stefanovic, & Christie, 1990). Therefore, they should be treated with caution. Such retrospective biases may even provide an alternative explanation of our results: It could be that participants who show an increased positivity bias report fewer relapses only because they underestimate their own symptomatology in retrospect due to this bias. Secondly, only 13 participants relapsed within this study. However, this relapse rate (32.5%) is comparable with the results of other studies that surveyed the relapse rate over a similar time period (e.g., Vázquez, Holtzman, Lolich, Ketter, & Baldessarini, 2015). Nevertheless, this study should be replicated with a substantially larger sample. Thirdly, we did not discriminate between manic and depressive relapses, as the number of relapses was too small. Consequently, we were not able to investigate whether a reduced positivity bias is associated with the risk for a depressive relapse specifically or with the risk of depressive and manic relapses likewise. It could well be that a reduced positivity bias is a risk factor specifically for a depressive relapse, whereas a heightened positivity bias might be a risk factor for mania. Fourthly, we did not investigate the influence of positive MI on ToM performance. Given some evidence that positive mood induction leads to alterations in decision making and attentional biases in rBD even with transient and subtle changes in mood (Roiser et al., 2009), future replications should also examine effects of positive mood induction. And last but not least, some results were only marginally significant or significant only in the preliminary t-tests but not in the regression model. Especially this and the small sample size point to the urge of a replication of these first findings concerning the predictive value of ToM decoding.

In summary, relapse can be predicted by the ability to correctly decode complex emotional facial expressions and by mood-linked changes in ToM decoding in a sample of rBD patients. If these findings can be replicated interventions for the treatment of BD should begin to focus on social cognition to prevent further affective episodes.

# 4 General Discussion

The work presented here aimed to investigate whether aUD, rUD, and rBD are associated with impairments in ToM decoding and ToM reasoning when ecologically valid stimulus material is used. As ecologically valid stimulus material, we chose the CAM face task to operationalize ToM decoding and the MASC to operationalize ToM reasoning. Another object of this work was to examine whether ToM performance predicts relapse of patients with rBD within a 9-month follow-up period. In addition, we were interested in whether a sad MI affects ToM performance differently in healthy individuals and patients, or if there is any influence at all. The impact of a mood-linked change in ToM on relapse in patients with rBD was also investigated.

Our findings suggest that there are no differences among aUD, rUD or HC individuals in ToM decoding performance. However, aUD patients were less confident in ToM decoding and found it more difficult than did HCs and rUDs. Furthermore, only the confidence and simplicity ratings of HCs and rUDs, but not of aUDs, are related to their ToM decoding performance. Moreover, both aUDs as well as rUDs showed a lower performance in ToM reasoning compared to HCs. The performance of aUDs and rUDs were equal. We found group differences for the cognitive mental state modality ($MASC_{cog}$) but not for the affective mental state modality ($MASC_{emo}$). These results suggest that in aUD patients as well as rUD patients, difficulties in ToM arise in later processing stages when it comes to integrating contextual information.

Regarding ToM in bipolar patients, we found that rBD patients were not impaired in ToM reasoning, but that younger rBD patients were impaired in ToM decoding. This is related to our finding that in rBDs, unlike in HCs, there was no age-related decrease in ToM decoding. Furthermore, in rBDs, relapse appeared not to be associated with ToM reasoning performance. However, relapse could be predicted by ToM decoding ability and by mood-linked changes in decoding. More precisely, in rBDs, better ToM decoding performance was associated with an increased probability to relapse, which was likely due to the observation that patients who had relapsed had a reduced positivity bias, which in turn led to a lower error rate. Moreover, a more pronounced reduction of the decoding performance following MI was in trend associated with relapse. These results suggest that the ease with which dysfunctional social cognitions can be brought back to mind under MI is associated with a negative course of BD.

From the comparative view of our results in UD and BD patients, it appears that euthymic BD patients have more abnormalities in early stages of emotional processing which are reflected in distinctive performances in tasks of ToM decoding. With the help of contextual information and subsequent unobtrusive emotional processing, these early abnormalities seem to be compensated for, so that rBDs do not show abnormalities in ToM reasoning. The integration of contextual

information seems to proceed in unbiased fashion and without interpretation errors. In contrast, within acutely depressed as well as remitted UDs, normal scores in ToM decoding suggest that these patients have no difficulties in early emotional processing. However, they show reduced ToM reasoning performance, which would indicate that they would have difficulties in later stages of emotional processing where cognitive interpretation biases take effect. Therefore, insufficient interpretations and biases seem to hinder the adequate integration of contextual influences. Taken together, these findings seem to indicate that ToM decoding and ToM reasoning are at least partially independent. As discussed in Chapter 1.2, this assumption is consistent with Sabbagh (2004), who points to imaging studies suggesting that ToM decoding and ToM reasoning are two distinct processes: Decoding depends on contributions from the right orbitofrontal/medial temporal circuit within the right hemisphere, whereas reasoning relies on left medial frontal regions.

Chapter 1.3 has already pointed out the differences between static and dynamic FER stimuli. In the joint consideration of our results on UD and BD – namely that anomalies in FER (ToM decoding) are only present in rBDs but not in aUDs or rUDs – our results fit well with the findings of previous studies that also used dynamic FER stimuli. For example, using the dynamic Emotional Expression Multimorph Task, which presents basal facial emotions in gradations from neutral to 100% emotional expression, Schaefer et al. (2010) and found anomalies in the aBD (acute depressive) group but not in the aUD group. It is also important to highlight that they found no difference between the HC group and the patient groups in percentage of correct responses. Instead, they found that BDs require a more intense facial expression before correctly identifying the emotion, indicating that BDs only have FER problems with dynamic emotional displays of low intensity (subtle emotions). In our study, subtle emotions were also used, which could be responsible for the difficulties of BD patients. Together with the findings of Schaefer et al. (2010) and other studies that used dynamic stimuli (e.g., Kan et al., 2004), the results of our study suggest that deficits in FER are not clearly found when dynamic and ecological valid material is used. We found FER performance comparable to HCs in rUD and aUD patients, and deficits were also relatively small in the rBD group, and were only observed in a younger subgroup. According to Kan et al. (2004) and Schaefer et al. (2010), dynamic facial expressions convey enough information for patients to accurately identify emotions, in contrast to static pictures. Therefore, the possibility of using more information content in the stimuli we used could account for the inconsistency between our findings and the previous studies using static stimuli with less information content. A further explanation, which however also fits to the explanation just given, refers to the simulation theory. This theory states that the activation of mirror neurons by observing others supports the generation of a ToM (see Chapter 1.2.2). Thus, this also implies that dynamic visual material should be particularly conducive to the activation of mirror neurons to enable the actual observation of others' activities or feelings.

However, these conditions are absent in static material or in tasks with verbal material (e.g., static pictures).

Certain aspects are discussed in more detail below. It should be mentioned at the outset that when assessing the results, the lack of studies using ecologically valid material makes discussion difficult in some cases. For the most part, a comparison is only possible with studies that did not use ecologically valid material. The conclusions that can be drawn when standard material of low ecological validity is used might be invalid when highly ecologically valid material is used, and vice versa.

## 4.1 ToM in major depression

When stimulus material with low ecological validity is used, previous studies of UD patients have found that deficits in ToM reasoning seem to be larger than deficits in FER, whereas FER seems to be only slightly impaired (Milders et al., 2010). Only small effect sizes are reported here (Dalili et al., 2015). Therefore, this finding seems to be reinforced when ecologically valid material is used. As a result, UDs no longer seem to show deficits in FER, but still in ToM reasoning. In line with this, Kupferberg, Bicks, and Hasler (2016) reported that that ToM impairment in aUDs is specific to the domain of ToM reasoning and that ToM decoding from observable cues was generally preserved in aUD patients. Beyond that, the results of our study suggest that this conclusion can be extended to the field of rUD as well. In the investigation of additionally rUD patients, our study offers a particularly valuable contribution, because in fact very few studies with remitted UD patients have been performed in the past.

Of particular interest is that, in Study 1 (see Chapter 3.1), although aUD patients decode complex emotions as successfully as the other two groups (rUD and HC) studied, they perceive the task as more difficult, are more uncertain in their choice of emotion, and are the only one of the three groups whose confidence and simplicity ratings do not correlate with their actual performance. These observations suggest a disturbance in self-appraisal, which is a fundamental feature of depression, in which patients see themselves in a poor light, e.g., as "defective" and "inadequate" (Beck, 1979), and where patients show a lower overall self-appraisal compared to HCs (Mathews, Williams, & Nedeljkovic, 2020). These considerations fit well with the study by Blankstein, Flett, and Johnston (1992). They administered a college student version of the Means Ends Problem Solving Procedure (MEPS; Platt & Spivack, 1975) with measures of perceived problem-solving ability to depressed and non-depressed students. They used this method to determine whether differences exist in both problem-solving ability and problem-solving appraisal. In the MEPS, participants are given descriptions of a series of interpersonal conflict situations and asked to list the steps they would take to achieve the positive outcomes also described in the test. Although the depressed

participants had more negative expectations and rated their problem-solving ability lower, the groups did not differ in the actual quality of their behavioural solutions to interpersonal, intrapersonal, and emotional problem situations. This kind of deficient self-appraisal in depression, which has no relation to the actual performance, was also observed in the aUD sample of our study for the domain of ToM decoding.

## 4.2 ToM in bipolar disorder

Our second study (see Chapter 3.2) indicates an anomaly in rBDs' ToM decoding. This is associated with the absence of an age-related decline in the performance on CAM task. However, this is noteworthy, in that an age-related decline in FER performance occurred both in the HC group examined here and in samples of other studies that have examined FER performance as a function of age (e.g., Study 1 in Chapter 3.1; more examples are listed in Chapter 3.2.4). Nevertheless, the overall performance in CAM (ToM decoding) was equal between rBDs and HCs, and there was no indication that rBDs were impaired or abnormal in any facet of ToM reasoning, including overall ToM, cognitive ToM, affective ToM and error types. Therefore, these results contradict previous research that reported general ToM impairments when using other ToM measures (e.g., Ibanez et al., 2012; Lahera et al., 2012; Martino et al., 2011; Wolf et al., 2010). As mentioned in Chapter 1.5.2, these inconsistencies could be attributed to variations in task modality and their moderating effect on whether ToM deficits can be identified (Aidelbaum & Goghari, 2022). For example, the most commonly used tasks, such as the Faux Pas task, False Belief task or Hinting task employ verbal formats. Moderate to large effects were consistently found when using these tasks, whereas studies that presented visual tasks, such as the tasks used in this study, reported only small effects, if any impairment was present at all (Samamé, 2013).

Furthermore, with respect to performance in CAM, we were able to identify an empirical age cut-off at 45 years (performance in CAM in younger vs. older rBD patients). As visible in Figure 1, this is also the area where the two regression lines of rBDs and HCs cross. Interestingly, Horning et al. (2012) found the same age cut-off. Precisely, they conducted analyses investigating the influence of cognition on basic FER in a sample that was not specifically a patient sample (a convenience sample of *N* =732 participants between the ages of 5 to 89). They used the facial expression recognition task (Murray, 2000) in which participants were shown dynamic facial photographs of the six basic emotions. Furthermore, they assessed participants' fluid intelligence by the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999), processing speed by a computerized reaction time task (Teng, 1990) and memory performance by the Rey Auditory Verbal Learning Test (RAVLT; Rey, 1964). They could demonstrate that cognitive abilities contributed to emotion recognition performance, especially for participants over age 45. Although Horning et al. (2012) used these results to explain

why FER performance declines with age – namely, due to the decline in overall cognitive performance – these results may also further explain why there was no difference between rBDs and HCs in the older sample of our rBD study. It is possible that the older rBD sample might have been able to compensate for its FER deficit with cognitive skills. Therefore, we recommend that in the future, the influence of age on social cognition in persons with affective disorder, especially BD, be reviewed as a main question.

In Study 3 (see Chapter 3.3), it turned out that the ToM decoding performance of patients who relapsed compared to that of patients who did not relapse is exactly the opposite of what we expected. Although the following explanatory approach by Berecz et al. (2016) for such a phenomenon was related to risk factors for the development of depression, it could also serve as an approach to explain our findings regarding risk factors for relapse in an rBD sample. According to Berecz et al. (2016), both impaired and enhanced ToM could be a risk factor for the development of depression, as both could lead to excessive rumination on subtle social cues, resulting in a tendency toward social withdrawal. Our findings with the reduced positivity bias of rBD patients who were going to relapse would be consistent with this hypothesis. Reduced positivity bias could also contribute to more negative interpretation of social situations, unfavorable reactions, social withdrawal, and eventually the development of a depressive phase. This is also consistent with our observation that almost all rBD patients who relapsed had a depressive episode (*N* = 12) and only very few additionally had a (hypo-)manic episode (*N* = 3) or only a manic episode (*N* = 1). An alternative explanation, which can also be complemented with the previous one, arises again from previous studies with UD patients. It was found that rUDs or individuals with subclinical depressive symptoms or dysphoria actually performed better on ToM decoding tasks (RMET) than healthy participants (this effect was not found on cognitive ToM tasks) (e.g., Berecz et al., 2016). Van Neerven et al. (2021) explained this with the fact that in these subclinical patients a negativity bias (e.g., Kupferberg et al., 2016; Weightman et al., 2014) or a reduced positivity bias is present only to a small extent, resulting in better recognition of stimuli with neutral and negative content. Moore and Fresco (2012) also refer to this as a “depressive realism” effect. In our study with rBD patients, it could be that the better ToM decoding performance of rBD patients who relapsed reflects the "negative realism" effect, especially since the tendency toward a reduced positivity bias was also observed. Remitted patients who experienced a depressive relapse within the next 9 months may have already had a tendency to detect negative and neutral stimuli at test time better. It might be, that the patients who did not relapse were presumably "robust" rehabilitants with little or no subsyndromal symptomatology, so there was no "negative realism" effect. Assuming that there was a “depressive realism” effect, the results of our study also suggest that it seems to influence only ToM decoding and not ToM reasoning, which is consistent with findings in UD patients. Moreover,

CAM consists of significantly more negative than positive and neutral stimuli (12 negative, 5 positive and 3 neutral emotion concepts), which can make the "negative realism" effect particularly evident. These considerations would suggest that the increased ToM decoding performance (and thus reduced positivity bias) in BD patients is dependent upon the current state of the patient rather than a trait marker. This may further explain the inconsistent results of previous studies: Accordingly, it could be that a subgroup of remitted patients, namely those who relapse soon, is better than HCs, whereas "robustly" remitted patients are equally good or perhaps worse than HCs. This would probably also be moderated by the stimulus material (e.g., amount of positive and negative emotional stimuli). However, these are considerations and research questions that cannot be answered by the studies presented here and should therefore be investigated further in future studies.

## 4.3 ToM in major depression, bipolar disorder and schizophrenia

Broad patterns of social dysfunction have been reported not only for UD and BD, but also for *schizophrenia* (SZ) (Porcelli et al., 2019; Velthorst et al., 2017). SZ, BD, and UD have overlapping clinical symptoms, particularly in the mood and psychosis spectrum (Kempf, Hussain, & Potash, 2005; Pearlson, 2015; Van Neerven et al., 2021). Especially for BD and SZ, there is evidence that these two disorders share substantial genetic and familial vulnerability (Lichtenstein et al., 2009; Owen, Craddock, & Jablensky, 2007). Therefore, it may be revealing to consider findings related to ToM deficits in these three mental disorders in comparison to each other.

ToM deficits as a trait marker versus state-dependent phenomenon

In their review, Van Neerven et al. (2021) concluded that ToM deficits may be a core characteristic of both psychotic and affective psychopathology. They also concluded that ToM deficits are more pronounced in SZ patients than in UD patients, with BD patients holding an intermediate position. They additionally inferred that ToM deficits persist beyond the symptomatic phase in both SZ and BD (albeit in an attenuated form), thus assuming that ToM deficits are a trait marker in these disorders. In contrast, ToM impairments in UD patients are more linked to the presence of acute depressive symptoms and thus are more likely to occur state-dependent than to be a trait marker. Our results are inconsistent with these conclusions in some respects. First of all, we found very limited evidence for abnormalities in ToM in rBD patients (only for the younger subsample and also only for ToM decoding). Therefore, it seems that the results of a review based on studies using mainly static, verbal and not very ecologically valid material cannot be generalized to our studies in which we used dynamic, visual and highly ecologically valid material. Accordingly, this would weaken the argument that ToM deficits are definitely composed of a trait component and

represent an endophenotype in BD, as suggested by Van Neerven et al. (2021). However, it cannot be excluded that even when using ecologically valid material, ToM deficits may be found in BD patients who are in an acute depressive or manic phase, where it would then be a state-dependent characteristic of the disease. In accordance with this, Van Neerven et al. (2021) reported that ToM deficits in BD are mainly associated with the presence of high levels of manic symptoms. Another point in which our results are contrary to what Van Neerven et al. (2021) stated is the conclusion that in UD, impairments of ToM are largely state-dependent. This does not seem to be the case when highly ecologically valid material is used, as in our study. Whereas there seemed to be no constraints in ToM decoding, ToM reasoning appeared to be clearly deficient in both aUDs and rUDs, suggesting a trait rather than a state-dependent impairment. However, there was at least partial accordance that UD patients also exhibit state-dependent ToM abnormalities, specifically in the ToM decoding domain. According to our results, although ToM decoding performance was unrestricted, only aUD patients indicated that they found it more difficult and were more uncertain in their judgements than rUDs and HCs.

The influence of non-social cognitive abilities

The inconsistency between our results and those from previous studies may also be related to the fact that the highly ecologically valid tasks we used were significantly less confounded with general non-social cognitive abilities. The tasks most commonly used in previous research to assess ToM reasoning are the False Beliefs Task, the Hinting Task and the Faux-Pas Task, all tasks that not only have low ecological validity, but are also predominantly verbal and rely more heavily on general cognitive skills (Westby, 2014). Given that significant (non-social or general) cognitive impairments have been identified in both BD (e.g., Vreeker et al., 2016) and SZ (Van Haren, Van Dam, & Stellato, 2019), the observation from previous research – namely, that the cognitive domain of ToM and higher-order ToM facilities appear to be more readily affected than the affective domain and first-order ToM processes (e.g., Bora, Yucel, & Pantelis, 2009; Healey, Bartholomeusz, & Penn, 2016; McKinnon, Cusi, & MacQueen, 2013; Samamé, 2013) – may simply reflect the presence of more generalized cognitive deficits in these disorders (Van Neerven et al., 2021). Even in SZ, performance in visually presented tasks is less affected than in verbal tasks (e.g., Healey et al., 2016), possibly as a result of the relatively higher cognitive demands placed on the interpretation of verbal information (Van Neerven et al., 2021). Furthermore, consistent with the finding mentioned above, that acutely manic BD patients were most affected by ToM impairments, S. Gruber, Rathgeber, Bräunig, and Gauggel (2007) also found in the domain of non-social cognitive abilities that attention and inhibitory control deficits were most pronounced in rBD patients who were recently manic, compared with rBD patients who were recently depressed and rUD patients. Therefore, the fact that CAM and MASC are less associated with general cognitive abilities, and that our rBD sample had comparatively few

constraints in cognitive and general functionality, may explain why no impairments in ToM decoding (overall sample) and ToM reasoning were found in our BD sample.

In this context, the results of Bora, Veznedaroğlu, and Vahip (2016), who performed a latent class analysis to identify neuropsychological subtypes in BD and SZ patients, might also be interesting. One of the authors' goals was to identify similarities and differences in the pattern of cognitive deficits, namely executive functions and ToM, between SZs and BDs. Although BDs and SZs have genetic similarities, each of the disorders also has unique familial and genetic risk factors (Hamshere et al., 2011; Lichtenstein et al., 2009). There are also major differences in the long-term outcome of the two disorders. Differences in the pattern of cognitive impairment may help to understand these differences. Bora, Veznedaroğlu, et al. (2016) investigated a sample of euthymic patients with BD and SZ and a HC group. They used the Stroop Color Word Test (SCWT; Stroop, 1935) and the Wisconsin card sorting test (WCST; Grant & Berg, 1993) to assess executive functions. They used the RMET and the Hinting tasks to assess ToM decoding and ToM reasoning. BD and SZ patients performed worse than HC, with BD performance intermediate between SZ and HC. Latent class analyses revealed four neurocognitive clusters, in which both patient groups were represented in each of the 4 groups. One cluster (Class I) included patients with no significant differences in ToM and executive functions when compared with HC. This group was therefore designated as the "neuropsychologically normal" cluster. Compared with SZ, BD patients were overrepresented in Class I (25.6% vs. 9.3%). In contrast, the „very severe cognitive impairment" cluster (Class IV) was relatively specific to SZ, and only a small minority of BD patients was member of this cluster (9.3% vs. 27.8%). The prevalence for the membership to the other two clusters was not significantly different between SZ and BD, suggesting that the cognitive profiles of approximately 60% of the individuals in both those patient groups were very close to each other. About 20% of the patients belonged to the "selective ToM" cluster (Class II), in which only ToM but not the executive functions were impaired. The remaining 40% of the patients belonged to Class III, which is characterized by impairment of both ToM and executive functions. The authors concluded that the existence of a "selective ToM impairment" cluster supports the notion that defcitis in ToM end executive functions are separable. At this point, however, it should be mentioned that it is not clear to what extent ToM and executive functions can be separated from each other, to what extent the two are related within (and also between) clusters, and to what extent they have moderating or mediating influences on each other. Assuming the validity of the cluster distribution found by Bora, Veznedaroğlu, et al. (2016), the results of our rBD study (Study 2) could be explained by the fact that, by chance, individuals from Class I were overrepresented in our BD sample, followed by individuals from Class II. But there were hardly any representatives from Class III and especially Class IV in our study. It would thus fit that in our rBD sample, some extent of deficits in ToM decoding were found, which is consistent with Bora,

Veznedaroğlu, et al. (2016) findings that Class II representatives are mainly impaired in ToM decoding (RMET). Thus, our sample would be a non-representative rBD sample. This would also be supported by the comparatively high general functionality and only relatively small deficits in neurocognitive tasks in our rBD sample. However, it is unclear whether the clusters and their composition in the groups of SZ and BD patients as defined by Bora, Veznedaroğlu, et al. (2016), can be replicated and thus are representative. It should also be mentioned in particular that it is questionable whether the same clusters and/or comparable cluster distribution would show up if the analyses were performed using ecologically valid ToM stimuli. It could well be that a different picture would then emerge. Indeed, on average, the more ecologically valid, dynamic, and less impaired by non-social cognitive abilities the stimulus material is, the smaller the differences to the HC group have been shown to be. In summary, in our rBD studies (Study 2 and 3, Chapter 3.2 and 3.3), several factors come together that have already been associated with less pronounced rather than more clear ToM reasoning deficits (or cognitive ToM deficits) in previous studies: 1) BD in remission and the absence of manic symptoms, 2) the use of visual rather than verbal material, 3) the use of tasks that are less dependent on general (non-social) cognitive skills, 4) the use of dynamic and highly ecological valid rather than static material and 5) the investigation of a BD sample of comparably good cognitive and general functionality.

In UD too, neurocognitive impairments have been reported (e.g., Porter, Robinson, Malhi, & Gallagher, 2015), e.g., in executive functioning (Marazziti, Consoli, Picchetti, Carlini, & Faravelli, 2010). There is some evidence in the literature that BD patients have higher levels of cognitive impairment than UD patients. Cotrena, Branco, Shansis, and Fonseca (2016) compared the executive function impairments in subsyndromal UD and BD. They found that BD patients had widespread impairment compared with UD patients and that BD patients performed worse in measures of sustained attention and inhibitory control. In addition, rBD patients exhibited a specific executive deficit that was independent of attentional impairment and did not occur in rUD patients (Stoddart, Craddock, & Jones, 2007). MacQueen and Memedovich (2017) reported that some studies found that rBD patients were more impaired on tests of verbal memory, set shifting, and inhibitory control, whereas UD patients were impaired only on set shifting. The observations of our studies are in line with this: Compared with HCs, aUDs and rUDs showed significant impairment only in working memory and task-switching ability (both measured by TMT-B). Patients with rBD additionally tended to be impaired in visual perceptual abilities, such as processing speed (measured by TMT-A), had a tendency to be impaired in verbal short-term retention ability (VLMT), and showed significantly more memory loss after delayed recall (VLMT). Although UD patients are significantly less impaired in non-social cognitions than rBD patients, we observed that they had significant deficits in ToM reasoning compared to HC, whereas this was not the case with rBDs. This suggests at least partial

independence of social and non-social cognitions when using ecologically valid ToM stimuli. However, exactly how this is the case must be the subject of future studies.

When these considerations are taken together, it could be that the ToM reasoning deficits in rBDs found so far were more a result of general cognitive impairment. Since ToM had been examined with measures that were confounded with non-social cognitions, a putative ToM deficit was measured that no longer showed up when ecologically valid material was used. The previous observation that BD patients were more likely to show impairment in ToM decoding are substantiated by our results with ecologically valid material, at least in the younger subsample. UDs, on the other hand, appear to have an isolated ToM reasoning deficit that is not clearly associated with impaired non-social cognitions. However, these are only initial assumptions that were not investigated in our studies and should therefore be addressed in future studies.

## 4.4 Investigation of mood induction

In the present study, we only examined the influence of a negative (sad) MI on ToM performance, not the influence of a *positive MI*. We decided to consider only the negative MI for two reasons: First, because the study design was already very complex, and consideration of a positive MI would have necessitated an even more extensive study. And second, because the overall findings on negative MI in affective disorders are clearer and more substantial than those on positive MI. Schmid and Schmid Mast (2010) also stated that the influence of a happy mood on emotion recognition is much less researched than the influence of a negative mood. In their own study with healthy individuals, they found clearer effects for the influence of a negative mood induction compared to the influence of a positive mood induction: Participants in sad moods showed a negative bias and recognized sad facial expressions better than happy ones. Participants in happy moods did not recognize happy facial expressions better than sad ones. Nevertheless, there is also evidence that a positive MI can influence social cognition: Schmid and Schmid Mast (2010) found in healthy participants that an induced happy mood was responsible for a reduction in the recognition of sad facial expressions compared to a neutral mood. This suggests that it would also be worthwhile to investigate the influence of positive MI on social cognition in affective disorders, especially when using tasks with high ecological validity. However, future studies examining the effects of MI in affective patients should take into account that reactivity to MI could vary within patient groups, depending on specific characteristics of the disorder. For example, Guhn, Steinacher, Merkl, Sterzer, and Köhler (2019) found a difference in affective reactivity between patients with recurrent depression and patients with persistent depression. The persistent group showed blunted reactivity to a negative MI, whereas the recurrent group showed an affective response comparable to HC: an increase in negative affect and a decrease in positive affect. Thus, blunted affective reactivity was

specifically associated with persistent as opposed to recurrent depression. The possibility of differential reactivity to MI based on certain characteristics of the disease was not considered in the present study and should be addressed in future studies. Furthermore, this could also explain the apparent observation mentioned in Chapter 3.1.4 (Discussion of Study 1), that the induction of sad mood (measured by PANAS items) was more successful in the rBD study than in the UD sample. In our UD study, we did not distinguish between chronic, recurrent, and single-episode UD. Therefore, we can only assume that a certain proportion of chronically depressed patients showed a weaker response to MI, which might explain the small effect sizes regarding the induction of negative mood in the sample of the UD study.

As mentioned above, studies of affective disorders have predominantly focused on negative feelings and moods and elaborated on their association with dysfunctional mechanisms and maladaptive consequences. Hence, most of the studies investigating the influence of negative MI make the assumption that abnormal processes in patients with affective disorder only show up if negative mood was also acutely present in the examination situation. Although in UD primarily negative affect is disturbed, in BD abnormalities in positive affect are also present, which is a cardinal symptom of this disorder. In her groundbreaking review, J. Gruber (2011) points out the lack of attention to harmful positive feelings. In her account of positive-emotion disturbance in BD, called Positive Emotion Persistence (PEP), she claims that BD is associated with heightened and persistent reward- and achievement-focused positive emotions that are present across contexts, including inappropriate ones. PEP also assumes that BD is associated with greater experiential and physiological indicators of reactivity to positive emotions rather than negative emotions or arousal. Therefore, exploring the impact of positive MI is particularly relevant in BD and should be the purpose of future studies. By inducing a positive mood, abnormal emotion-related processes could be triggered in BD patients, which in turn could affect FER and ToM performance.

## 4.5 Limitations

### 4.5.1 Affective versus cognitive ToM

As mentioned earlier, we hypothesized at the outset that a substantial amount of variance in ToM reasoning is due to the ability to decode complex emotions. In the present work, the verification was unnecessary because in the studies, depending on the sample examined, either only decoding or only reasoning, but not both, differed significantly from HCs within a patient sample. It should be noted, however, that, strictly speaking, this question could only be partially investigated with our study material for the following reason: Regarding the division of affective and cognitive ToM described in the introduction (Shamay-Tsoory, Aharon-Peretz, et al., 2009), only affective ToM decoding was measured with CAM. This circumstance is especially relevant for Study 1 (see Chapter

3.1), in which a deficit was found only for cognitive but not for affective ToM reasoning in depressives. Thus, a specific deficit only in cognitive ToM reasoning probably cannot be related to the performance in CAM (affective ToM decoding only).

### 4.5.2 Suitability of the CAM

Another question open for discussion is whether the CAM is suitable for studying a German sample. The tool is based on a lexical approach (see Chapter 1.2.1), which includes all the emotion terms in the English language. Further differentiation was made on the basis of word frequency in the English language and verbal comprehension. We translated the emotion terms of the CAM from English to German. These terms were back-translated (by a native speaker) and discussed if the back-translation did not fit. We provided participants with a manual in which all of the emotion terms used were briefly defined to ensure that all terms could be properly understood. However, some of the study participants mentioned that they experienced the emotion terms of the CAM task as unusual. In addition, some patients used the manual, which included the explanation of emotion terms, with remarkable frequency. Therefore, whether the translated terms were appropriate or seemed rather "artificial" and thus were cognitively challenging cannot be answered unequivocally. Cognitively challenging emotion words could be problematic for the following reason: The concept of *inhibitory control* (e.g., Bartholomew, Heller, & Miller, 2021) or *cognitive modulation* (e.g., Pessoa, Padmala, & Morland, 2005) suggests that activation of the prefrontal cortex, e.g., by a demanding cognitive task, inhibit emotion salience networks, such as the amygdala or hippocampus. Evidence suggests that cognitive modulation comprises a strong factor in determining amygdala responses (Pessoa et al., 2005). Therefore, a somewhat broader conjecture would be that the unfamiliarity of the emotion terms, those used in our study, triggered more cognitive activation and concomitant inhibition of emotional responses. This could have also altered emotional processing.

We chose MASC as a realistic measure of ToM reasoning and CAM as a realistic measure of ToM decoding. One goal was also to relate performance in the two tasks. However, while the MASC is a very broad, multimodal, and all-encompassing measure of ToM reasoning, the CAM refers only to a very specific facet (facial expressions only). In future research, a realistic measurement method of ToM decoding should also incorporate observable cues from various sources of information (e.g., tone of voice, gestures, posture).

## 4.6 Future directions

### 4.6.1 Homogenous subgroups of patients

The inconsistent results for social cognition in affective disorders may have arisen not only from methodological shortcomings but also from the heterogeneity of the disease (Berecz et al.,

2016; Samamé, 2013) and the heterogeneity regarding demographic variables (see Chapter 1.5). Our results indicate that the *subgroups of affective disorders* - individuals with UD versus individuals with BD - show abnormalities in different domains of social cognition. Furthermore, this work indicates that the *age* of the participants should be considered, at least in BD samples. We therefore agree with Berecz et al. (2016) who propose, too, that it might make sense to examine ToM abilities in different homogenous subgroups of patients with affective disorder. Specifically, from this work it can be deduced that UD and BD should not be mixed in the future, and we also recommend forming homogeneous groups with respect to certain demographic variables, such as age, which have also been associated with social cognition. A homogeneous study group should possibly also be formed according to specific characteristics of the disease, such as severity (mild vs. moderate vs. severe; Bora, Bartholomeusz, et al., 2016; Van Neerven et al., 2021) and the presence of certain features such as suicidality (e.g., Szanto et al., 2012), general cognitive deficits, or a chronic symptomatology.

### 4.6.2 Use of uniform and ecologically valid measurement methods

To date, research on ToM in affective disorders has primarily used static visual stimulus material (especially in ToM decoding) or verbally presented or drawn (especially in ToM reasoning) stimulus material. This material shows a lack of ecological validity and is also strongly confounded with other, non-social cognitions. This is problematic, because some patients with affective disorders have deficits in general cognitive abilities. Therefore, the extent of non-social neurocognitive impairment should also be considered in the respective study group, and ToM measures that are relatively unconfounded with general cognitive abilities should be used. The ecologically valid material to be used is presented visually and displays real individuals naturally acting emotions or social situations. For social situations, the measurement procedure must also ensure that the participants are presented with contextual factors that are as close to everyday life as possible, so that they have the opportunity to integrate them. To increase the comparability of results, the measures used should be uniform in these characteristics.

### 4.6.3 Other aspects of social cognition in affective disorders

The present work is primarily concerned with deviations in ToM as an explanation for problems in the social functioning of patients with affective disorder. However, social functioning also depends on social cognitions other than ToM. For example, social perception and empathy are also important components of social cognition, and there are studies indicating that these components are abnormal or deficient in some individuals with UD (e.g., Schreiter, Pijnenborg, & aan het Rot, 2013) and BD (e.g., Derntl, Seidel, Schneider, & Habel, 2012). Our conclusions about the requirements for ToM research can be applied to these domains as well: Care should also be taken when measuring other aspects of social cognition to ensure that the stimulus material is ecologically

valid and has little interference with non-social cognition, that uniform patient groups are studied, and that particular demographic or clinical characteristics are taken into account.

### 4.6.4 Self-reference bias

By using the CAM, one focus of the present study assesses FER of complex or social emotions, which, according to Adolphs (2002), involve the representation of the self to a greater extent than basic emotions. The representation of the self, in turn, may be subject to specific biases, particularly in affective disorders (e.g., Lyon, Startup, & Bentall, 1999; Schwert, Stohrer, Aschenbrenner, Weisbrod, & Schröder, 2018; Smith, Reynolds, Orchard, Whalley, & Chan, 2018), which are themselves associated with core or secondary features of the affective disorder, such as worthlessness and self-blame (e.g., Zahn et al., 2015), negative basic assumptions about the self and the self in relation to others as well as maladaptive schemes (e.g., J. E. Young, Klosko, Weishaar, & Kierdorf, 2005). It is likely that these self-reference biases come into effect primarily in those situations or tasks in which we are involved as social interaction partners by, among other things, influencing ToM performance accordingly. However, during the execution of the CAM task and the MASC task, the participants are merely observing a person like a spectator. Therefore, the influence of a self-referential bias, which is presumably relatively large, could not be tested in this study. In future studies, ToM and FER performance should be investigated in social situations that are personally relevant to the participants, e.g., when they have to engage as social counterparts.

This idea fits with the article by Schilbach et al. (2013), which states that social cognition is fundamentally different when we're interacting with others than when we're just observing them, as the dynamics of social interaction contribute to and even constitute our perception of other minds. According to this, a social partner's reaction and her or his ToM depend, among other things, on her or his neural processes, which have been partly shaped by personal experience. Furthermore, the actor's reaction depends on his intentions, abilities and habits. These influencing factors may be disrupted in individuals suffering from affective disorders (e.g., Stephan, 2013) and consequently lead to abnormalities in social cognition. In order for these influencing factors to come into play in ToM performance and make investigation possible, tasks must be used in which the test person participates as an active social partner in a social interaction.

### 4.6.5 Practical implications

The practical implications from studies of social cognition in affective disorders concern two main issues (Weightman et al., 2014). First, they can be used to further differentiate the UD and BD phenotype and to derive explanations for the problems in social functioning. Second, they can be used to identify treatment options by targeting social cognition for improvement, and they can further be used to monitor improvements.

Regarding the first point, it can be concluded from our results that BD and UD differ in terms of deficits in ToM when measurements with high ecological validity are used. BDs appear to have more difficulties in correctly recognizing complex facial emotion expressions (ToM decoding), and the occurrence of these deficits seems to depend on age. UDs, on the other hand, have difficulty with cognitive ToM reasoning, and this is equally true for aUDs and rUDs, suggesting that deficits in ToM reasoning may be a trait marker in UD.

Regarding the second point - treatment options - the question arises whether already existing treatments could be extended by further components. As already discussed, the difficulties in ToM reasoning in UD seem to be mainly based on cognitive biases. The already established therapy of *cognitive restructuring* or *interpersonal psychotherapy* in depression seems to be a helpful approach to counteract these deficits. There are already special therapy methods, such as the *Cognitive Behavioral Analysis System of Psychotherapy* (CBASP; McCullough Jr, 2003), which among other things also aims to train chronically depressed people to interpret correctly the intentions or mental states of others and their own. Because only a small proportion in our depressed sample were chronically depressed, and yet clear ToM reasoning deficits were identified, the CBASP method could also be helpful for individuals with first-episode, recurrent, or remitted depression and could become part of treatment. Furthermore, Bateman and Fonagy (2008) developed and drafted a manual for the *Mentalization-Based Treatment* (MBT), originally for individuals with borderline personality disorder. MBT is designed to enable people to better understand their own desires, thoughts, and beliefs, as well as those of others. In other people and in oneself, mental and emotional processes that underlie one's actions are to be recognized and understood. According to our results, these treatments (CBASP and MBT) could also be used to treat depressed and bipolar patients. However, both the CBASP and the MBT methods are relatively complex, time-consuming and lengthy treatments. Furthermore, these psychoanalytically oriented methods are not practiced very widely. Future studies could investigate whether the basic training in ToM, for example as (computer-based) training, might already have an effect. If this were the case, it would be a relatively time-efficient and cost-effective alternative. Following the model of social cognition training for schizophrenics (see the last section of this chapter), appropriate training and methods can be developed and adapted, and their effects on ToM performance can be investigated in future studies.

Equivalently, basic ToM decoding training for BD patients could be developed and first investigated to see if it contributes to treatment success or at least promotes better social outcome in BD patients before being included as standard in the treatment regimen for these patients.

In Chapter 4.3, a connection between affective disorders and schizophrenia has already been demonstrated. Deficits in social cognition are considered a core feature of schizophrenia (Green, Horan, & Lee, 2015) and a number of social cognition training programs have been developed for

schizophrenia patients. The recent meta-analysis by d'Arma et al. (2021) provides an overview of the state-of-the-art of social cognition trainings in SZ and their efficacy on both cognitive and affective ToM. However, there are only few studies that have examined the effectiveness of these trainings in UD and BD patients. There is a study by Lahera et al. (2013) that examined whether the *Social Cognition and Interaction Training* (SCIT; Roberts, Penn, & Combs, 2006) is more effective than treatment as usual in BD and SZ patients. SCIT is a training program designed to improve emotion perception, attributional style and ToM abilities. Lahera et al. provided preliminary evidence that SCIT is feasible and may improve social cognition in BD and SZ outpatients. Further studies are needed to look at whether existing social cognition trainings can also improve ToM and social functioning levels in UD and BD patients, and if so, whether it should be included in the standard treatment program.

### 4.6.6 Summary

A broad deduction can be made from the present work, that a highly differentiated approach needs to be taken in research on affective disorders and social cognition. First, the heterogeneity in affective disorders must be taken into account, which has different effects on social cognition (e.g., BD versus UD, younger versus older patients, female versus male patients, severely versus mildly depressed patients). Second, the type of operationalization or stimuli used seems crucial (static vs. dynamic, verbal vs. visual, subtle versus fully developed emotion displays, degree of ecological validity). And third, further differentiation must be made within social cognition as well as within respective aspects of social cognition, as different facets of the same construct may be affected independently of each other (e.g., ToM decoding versus ToM reasoning, general emotion recognition versus specific emotion recognition). This complexity makes exploring this area so difficult and may further explain the numerous inconsistent findings. Thus, this work makes clear the need for a very careful and differentiated research approach.

# References


Amrican Psychiatric Association, 1994. DSM-IV: Diagnostic and Statistic Manual of Mental Disorders. American Psychiatric Association, *Washington DC.*

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

Adolphs, R. (2002). Recognizing emotion from facial expressions: psychological and neurological mechanisms. *Behavioral and Cognitive Neuroscience Reviews, 1*(1), 21-62. doi:10.1177/1534582302001001003

Adolphs, R. (2009). The social brain: Neural basis of social knowledge. *Annual Review of Psychology, 60*(1), 693-716. doi:10.1146/annurev.psych.60.110707.163514

Adolphs, R., Tranel, D., & Damasio, A. R. (2003). Dissociable neural systems for recognizing emotions. *Brain and Cognition, 52*(1), 61-69. https://doi.org/10.1016/S0278-2626(03)00009-5

Aidelbaum, R., & Goghari, V. (2022). A visual task-based assessment of theory of mind and social perception within bipolar disorder. *The Journal of Nervous and Mental Disease, 210*(1), 37-44. doi:10.1097/nmd.0000000000001408

Almeida, J. R. C., Versace, A., Hassel, S., Kupfer, D. J., & Phillips, M. L. (2010). Elevated amygdala activity to sad facial expressions: a state marker of bipolar but not unipolar depression. *Biological Psychiatry, 67*(5), 414-421. https://doi.org/10.1016/j.biopsych.2009.09.027

Altman, S., Haeri, S., Cohen, L. J., Ten, A., Barron, E., Galynker, I. I., & Duhamel, K. N. (2006). Predictors of relapse in bipolar disorder: a review. *Journal of Psychiatric Practice, 12*(5), 269-282. https://doi.org/10.9758/cpn.2018.16.2.168

Ambadar, Z., Schooler, J. W., & Cohn, J. F. (2005). Deciphering the enigmatic face: the importance of facial dynamics in interpreting subtle facial expressions. *Psychological Science, 16*(5), 403-410. doi:10.1111/j.0956-7976.2005.01548.x

Anderson, I. M., Shippen, C., Juhasz, G., Chase, D., Thomas, E., Downey, D., . . . Deakin, J. F. W. (2011). State-dependent alteration in face emotion recognition in depression. *British Journal of Psychiatry, 198*(4), 302-308. doi:10.1192/bjp.bp.110.078139

Andrews, S. C. (2013). *Exploring social cognition and mirror systems in schizophrenia and bipolar disorder.* [Doctoral dissertation, Monash University].

Arnett, J. A., & Labovitz, S. S. (1995). Effect of physical layout in performance of the Trail Making Test. *Psychological Assessment, 7*(2), 220-221. https://doi.org/10.1037/1040-3590.7.2.220

Baez, S., Herrera, E., Villarin, L., Theil, D., Gonzalez-Gadea, M. L., Gomez, P., . . . Ibañez, A. M. (2013). Contextual social cognition impairments in schizophrenia and bipolar disorder. *PLoS ONE, 8*(3). doi:10.1371/journal.pone.0057664

Bagby, R. M., Ryder, A. G., Schuller, D. R., & Marshall, M. B. (2004). The Hamilton Depression Rating Scale: has the gold standard become a lead weight? *American Journal of Psychiatry, 161*(12), 2163-2177. doi:10.1176/appi.ajp.161.12.2163

Barker, A. A., Musso, M. W. & Gouvier, W. D. (n.d.). *ecological validity*. *Encyclopedia Britannica*. https://www.britannica.com/science/ecological-validity. Accessed 28 March 2022.

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

Baron-Cohen, S., Golan, O., Hill, J. J., & Wheelwright, S. (submitted). Developmental changes in the emotion lexicon: a research note.

Baron-Cohen, S., Golan, O., Wheelwright, S., & Hill, J. (2004). *Mind reading: The interactive guide to emotions.* London: Jessica Kingsley.

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

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "reading the mind in the eyes" test revised version: A study with normal adults, and adults with asperger syndrome or high-functioning autism. *The Journal of Child Psychology and Psychiatry and Allied Disciplines, 42*(2), 241-251. doi:10.1017/S0021963001006643

Baron-Cohen, S., Wheelwright, S., Jolliffe, & Therese. (1997). Is there a "language of the eyes"? Evidence from normal adults, and aults with autism or Asperger syndrome. *Visual Cognition, 4*(3), 311-331. doi:10.1080/713756761

Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another advanced test of theory of mind: Evidence from very high functioning adults with autism or Asperger syndrome. *Journal of Child Psychology and Psychiatry, 38*(7), 813-822. doi:10.1111/j.1469-7610.1997.tb01599.x

Barrera, Á., Vázquez, G., Tannenhaus, L., Lolich, M., & Herbst, L. (2013). Theory of mind and functionality in bipolar patients with symptomatic remission. *Revista de Psiquiatría y Salud Mental (English Edition), 6*(2), 67-74. https://doi.org/10.1016/j.rpsmen.2012.07.003

Bartholomew, M. E., Heller, W., & Miller, G. A. (2021). Inhibitory control of emotional processing: Theoretical and empirical considerations. *International Journal of Psychophysiology, 163*, 5-10. https://doi.org/10.1016/j.ijpsycho.2019.03.015

Bateman, A., & Fonagy, P. (2008). *Psychotherapie der Borderline-Persönlichkeitsstörung: ein mentalisierungsgestütztes Behandlungskonzept; mit einem umfangreichen Behandlungsmanual*: Psychosozial-Verlag.

Bazin, N., Brunet-Gouet, E., Bourdet, C., Kayser, N., Falissard, B., Hardy-Baylé, M.-C., & Passerieux, C. (2009). Quantitative assessment of attribution of intentions to others in schizophrenia using an ecological video-based task: A comparison with manic and depressed patients. *Psychiatry Research, 167*(1), 28-35. https://doi.org/10.1016/j.psychres.2007.12.010

Beaupré, M. G., & Hess, U. (2005). Cross-cultural emotion recognition among Canadian ethnic groups. *Journal of Cross-Cultural Psychology, 36*(3), 355-370. doi:10.1177/0022022104273656

Beck, A. T. (1979). *Cognitive therapy of depression*. Guilford press.

Beck, A. T. (1983). Cognitive therapy of depression: New perspectives. In P. J. Clayton, & J. E. Barett (Eds.), *Treatment of depression: Old controversies and new approaches* (pp. 265-284). New York: Raven Press.

Beck, A. T. (1996). Beyond belief: A theory of modes, personality, and psychopathology. In P. M. Salkovskis (Ed.), *Frontiers of cognitive therapy* (pp. 1–25). The Guilford Press.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4*(6), 561. doi:10.1001/archpsyc.1961.01710120031004

Bediou, B., Krolak-Salmon, P., Saoud, M., Henaff, M.-A., Burt, M., Dalery, J., & D'Amato, T. (2005). Facial expression and sex recognition in schizophrenia and depression. *The Canadian Journal of Psychiatry, 50*(9), 525-533.

Beer, J. S., & Ochsner, K. N. (2006). Social cognition: a multi level analysis. *Brain Research, 1079*(1), 98-105. https://doi.org/10.1016/j.brainres.2006.01.002

Ben-Zeev, D., Young, M. A., & Madsen, J. W. (2009). Retrospective recall of affect in clinically depressed individuals and controls. *Cognition and Emotion, 23*(5), 1021-1040. doi:10.1080/02699930802607937

Berecz, H., Tényi, T., & Herold, R. (2016). Theory of mind in depressive disorders: A review of the literature. *Psychopathology, 49*(3), 125-134. doi:10.1159/000446707

Bertoux, M., Delavest, M., de Souza, L. C., Funkiewiez, A., Lépine, J.-P., Fossati, P., . . . Sarazin, M. (2012). Social cognition and emotional assessment differentiates frontotemporal dementia from depression. *Journal of Neurology, Neurosurgery & Psychiatry, 83*(4), 411-416. doi:10.1136/jnnp-2011-301849

Bhagwagar, Z., Cowen, P. J., Goodwin, G. M., & Harmer, C. J. (2004). Normalization of enhanced fear recognition by acute SSRI treatment in subjects with a previous history of depression. *American Journal of Psychiatry, 161*(1), 166-168. doi:10.1176/appi.ajp.161.1.166

Biehl, M., Matsumoto, D., Ekman, P., Hearn, V., Heider, K., Kudoh, T., & Ton, V. (1997). Matsumoto and Ekman's japanese and caucasian facial expressions of emotion (JACFEE): Reliability data and cross-national differences. *Journal of Nonverbal Behavior, 21*(1), 3-21. https://doi.org/10.1023/A:1024902500935

Biele, C., & Grabowska, A. (2006). Sex differences in perception of emotion intensity in dynamic and static facial expressions. *Experimental Brain Research, 171*(1), 1-6. doi:10.1007/s00221-005-0254-0

Bio, D. S., Soeiro de-Souza, M. G., Otaduy, M. C. G., Machado-Vieira, R., & Moreno, R. A. (2013). The impact of limbic system morphology on facial emotion recognition in bipolar I disorder and healthy controls. *Neuropsychiatric Disease and Treatment, 9*, 743-751. doi:10.2147/NDT.S41896

Biyik, U., Keskin, D., Oguz, K., Akdeniz, F., & Gonul, A. S. (2015). Facial emotion recognition in remitted depressed women. *Asian Journal of Psychiatry, 17*, 111-113. http://dx.doi.org/10.1016/j.ajp.2015.08.001

Blair, R. J. R., Colledge, E., Murray, L., & Mitchell, D. G. V. (2001). A selective impairment in the processing of sad and fearful expressions in children with psychopathic tendencies. *Journal of Abnormal Child Psychology, 29*(6), 491-498. doi:10.1023/A:1012225108281

Blankstein, K. R., Flett, G. L., & Johnston, M. E. (1992). Depression, problem-solving ability, and problem-solving appraisals. *Journal of Clinical Psychology, 48*(6), 749-759. https://doi.org/10.1002/1097-4679(199211)48:6<749::AID-JCLP2270480609>3.0.CO;2-V

Bombari, D., Schmid, P. C., Schmid Mast, M., Birri, S., Mast, F. W., & Lobmaier, J. S. (2013). Emotion recognition: The role of featural and configural face information. *Quarterly Journal of Experimental Psychology, 66*(12), 2426-2442. doi:10.1080/17470218.2013.789065

Bonawitz, E. B., van Schijndel, T. J. P., Friel, D., & Schulz, L. (2012). Children balance theories and evidence in exploration, explanation, and learning. *Cognitive Psychology, 64*(4), 215-234. https://doi.org/10.1016/j.cogpsych.2011.12.002

Bora, E., Bartholomeusz, C., & Pantelis, C. (2016). Meta-analysis of theory of mind (ToM) impairment in bipolar disorder. *Psychological Medicine, 46*(2), 253-264. doi:10.1017/S0033291715001993

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

Bora, E., Vahip, S., Gonul, A. S., Akdeniz, F., Alkan, M., Ogut, M., & Eryavuz, A. (2005). Evidence for theory of mind deficits in euthymic patients with bipolar disorder. *Acta Psychiatrica Scandinavica, 112*(2), 110-116. doi:10.1111/j.1600-0447.2005.00570.x

Bora, E., Veznedaroğlu, B., & Vahip, S. (2016). Theory of mind and executive functions in schizophrenia and bipolar disorder: a cross-diagnostic latent class analysis for identification of neuropsychological subtypes. *Schizophrenia Research, 176*(2-3), 500-505. doi:10.1016/j.schres.2016.06.007

Bora, E., Yucel, M., & Pantelis, C. (2009). Theory of mind impairment in schizophrenia: Meta-analysis. *Schizophrenia Research, 109*(1), 1-9. https://doi.org/10.1016/j.schres.2008.12.020

Bora, E., Yücel, M., & Pantelis, C. (2009). Theory of mind impairment: a distinct trait-marker for schizophrenia spectrum disorders and bipolar disorder? *Acta Psychiatrica Scandinavica, 120*(4), 253-264. doi: 10.1111/j.1600-0447.2009.01414.x

Bora, E., & Zorlu, N. (2017). Social cognition in alcohol use disorder: a meta-analysis. *Addiction, 112*(1), 40-48. doi: 10.1111/add.13486

Bornstein, R. A., Baker, G. B., & Douglass, A. B. (1987). Short-term retest reliability of the Halstead-Reitan Battery in a normal sample. *Journal of Nervous and Mental Disease, 175*(4), 229-232. doi:10.1097/00005053-198704000-00007

Bouhuys, A. L., Bloem, G. M., & Groothuis, T. G. G. (1995). Induction of depressed and elated mood by music influences the perception of facial emotional expressions in healthy subjects. *Journal of Affective Disorders, 33*(4), 215-226. doi: 10.1016/0165-0327(94)00092-N

Bouhuys, A. L., Geerts, E., & Gordijn, M. C. M. (1999). Depressed patients' perceptions of facial emotions in depressed and remitted states are associated with relapse: A longitudinal study. *The Journal of Nervous and Mental Disease, 187*(10), 595-602. doi:10.1097/00005053-199910000-00002

Bould, E., & Morris, N. (2008). Role of motion signals in recognizing subtle facial expressions of emotion. *British Journal of Psychology, 99*(2), 167-189. doi:10.1348/000712607X206702

Bower, G. H. (1981). Mood and memory. *American Psychologist, 36*(2), 129-148. https://doi.org/10.1037/0003-066X.36.2.129

Bower, G. H., & Forgas, J. P. (2000). Affect, memory, and social cognition. *Cognition and Emotion*, 87-168. Oxford University Press.

Box, G. E., & Tiao, G. C. (2011). *Bayesian inference in statistical analysis* (Vol. 40). John Wiley & Sons.

Brothers, L. (2002). The social brain: a project for integrating primate behavior and neurophysiology in a new domain. *Foundations in Social Neuroscience, 367*, 385.

Brüne, M. (2003). Social cognition and behaviour in schizophrenia. In M. Brüne, H. Ribbert & W. Schiefenhövel (Eds.), *The Social Brain: Evolution and Pathology*, 277-313.

Budak, E. (2011). *Theory of mind and its relationship with clinical features and social functioning in remitted bipolar patients* [Doctoral dissertation*,* Thesis. Istanbul. Bakırkoy Research and Training Hospital].

Calder, A. J., Keane, J., Manly, T., Sprengelmeyer, R., Scott, S., Nimmo-Smith, I., & Young, A. W. (2003). Facial expression recognition across the adult life span. *Neuropsychologia, 41*(2), 195-202. https://doi.org/10.1016/S0028-3932(02)00149-5

Caletti, E., Paoli, R. A., Fiorentini, A., Cigliobianco, M., Zugno, E., Serati, M., . . . Altamura, A. (2013). Neuropsychology, social cognition and global functioning among bipolar, schizophrenic patients and healthy controls: preliminary data. *Frontiers in Human Neuroscience, 7*(661). doi:10.3389/fnhum.2013.00661

Calhoun, S. L., & Mayes, S. D. (2005). Processing speed in children with clinical disorders. *Psychology in the Schools, 42*(4), 333-343. https://doi.org/10.1002/pits.20067

Campbell, R., Elgar, K., Kuntsi, J., Akers, R., Terstegge, J., Coleman, M., & Skuse, D. (2002). The classification of 'fear' from faces is associated with face recognition skill in women. *Neuropsychologia, 40*(6), 575-584. https://doi.org/10.1016/S0028-3932(01)00164-6

Canli, T., & Lesch, K.-P. (2007). Long story short: the serotonin transporter in emotion regulation and social cognition. *Nature Neuroscience, 10*(9), 1103-1109. doi:10.1038/nn1964

Carlson, S. M., Koenig, M. A., & Harms, M. B. (2013). Theory of mind. *Wiley Interdisciplinary Reviews: Cognitive Science, 4*(4), 391-402. https://doi.org/10.1002/wcs.1232

Carr, L., Iacoboni, M., Dubeau, M.-C., Mazziotta, J. C., & Lenzi, G. L. (2003). Neural mechanisms of empathy in humans: a relay from neural systems for imitation to limbic areas. *Proceedings of the National Academy of Sciences of the United States of America, 100*(9), 5497-5502. doi:10.1073/pnas.0935845100

Chepenik, L. G., Cornew, L. A., & Farah, M. J. (2007). The influence of sad mood on cognition. *Emotion, 7*(4), 802-811. doi:10.1037/1528-3542.7.4.802

Chowdhury, M. Z. I., & Turin, T. C. (2020). Variable selection strategies and its importance in clinical prediction modelling. *Family Medicine and Community Health, 8*(1). doi:10.1136/fmch-2019-000262

Clore, G. L., Gasper, K., & Garvin, E. (2001). Affect as information. In J. P. Forgas (Ed.), *Handbook of affect and social cognition* (pp. 121-144). Lawrence Erlbaum Associates Publishers.

Clore, G. L., & Storbeck, J. (2006). Affect as information about liking, efficacy, and importance. In J. P. Forgas (Ed.), *Affect in social thinking and behavior* (pp. 123–141). Psychology Press.

Cohen, A. N., Hammen, C., Henry, R. M., & Daley, S. E. (2004). Effects of stress and social support on recurrence in bipolar disorder. *Journal of Affective Disorders, 82*(1), 143-147. https://doi.org/10.1016/j.jad.2003.10.008

Corcoran, R., Mercer, G., & Frith, C. D. (1995). Schizophrenia, symptomatology and social inference: investigating "theory of mind" in people with schizophrenia. *Schizophrenia Research, 17*(1), 5-13. doi:10.1016/0920-9964(95)00024-g

Coryell, W., Scheftner, W., Keller, M., Endicott, J., Maser, J., & Klerman, G. L. (1993). The enduring psychosocial consequences of mania and depression. *The American Journal of Psychiatry, 150*(5), 720-727. https://doi.org/10.1176/ajp.150.5.720

Cotrena, C., Branco, L. D., Shansis, F. M., & Fonseca, R. P. (2016). Executive function impairments in depression and bipolar disorder: association with functional impairment and quality of life. *Journal of Affective Disorders, 190*, 744-753. https://doi.org/10.1016/j.jad.2015.11.007

Coyle, T. R., Elpers, K. E., Gonzalez, M. C., Freeman, J., & Baggio, J. A. (2018). General intelligence (g), ACT scores, and theory of mind: (ACT) g predicts limited variance among theory of mind tests. *Intelligence, 71*, 85-91. doi:10.1016/j.intell.2018.10.006

Crawford, J. R., & Henry, J. D. (2004). The Positive and Negative Affect Schedule (PANAS): Construct validity, measurement properties and normative data in a large non-clinical sample. *British Journal of Clinical Psychology, 43*(3), 245-265. doi:10.1348/0144665031752934

Csukly, G., Czobor, P., Szily, E., Takács, B., & Simon, L. (2009). Facial expression recognition in depressed subjects: the impact of intensity level and arousal dimension. *The Journal of Nervous and Mental Disease, 197*(2), 98-103. doi:10.1097/NMD.0b013e3181923f82

Cüceloglu, D. M. (1970). Perception of facial expressions in three different cultures. *Ergonomics, 13*(1), 93-100. doi:10.1080/00140137008931123

Cusi, A. M., MacQueen, G. M., & McKinnon, M. C. (2012). Patients with bipolar disorder show impaired performance on complex tests of social cognition. *Psychiatry Research, 200*(2), 258-264. https://doi.org/10.1016/j.psychres.2012.06.021

Cusi, A. M., Nazarov, A., MacQueen, G. M., & McKinnon, M. C. (2013). Theory of mind deficits in patients with mild symptoms of major depressive disorder. *Psychiatry Research, 210*(2), 672-674. https://doi.org/10.1016/j.psychres.2013.06.018

Cusin, C., Yang, H., Yeung, A., & Fava, M. (2009). Rating scales for depression. In L. Baer & M. A. Blais (Eds.), *Handbook of clinical rating scales and assessment in psychiatry and mental health* (pp. 7–35). Humana Press.

d'Arma, A., Isernia, S., Di Tella, S., Rovaris, M., Valle, A., Baglio, F., & Marchetti, A. (2021). Social cognition training for enhancing affective and cognitive theory of mind in schizophrenia: A systematic review and a meta-Analysis. *The Journal of Psychology, 155*(1), 26-58. doi:10.1080/00223980.2020.1818671

Dalili, M. N., Penton-Voak, I. S., Harmer, C. J., & Munafò, M. R. (2015). Meta-analysis of emotion recognition deficits in major depressive disorder. *Psychological Medicine, 45*(6), 1135-1144. doi:10.1017/S0033291714002591

Damasio, A. R. (1995). Toward a Neurobiology of Emotion and Feeling: Operational Concepts and Hypotheses. *The Neuroscientist, 1*(1), 19-25. doi:10.1177/107385849500100104

Damasio, A. R. (2014). *Descartes' Irrtum: Fühlen, Denken und das menschliche Gehirn*. Ullstein eBooks.

Darke, H., Cropper, S. J., & Carter, O. (2019). A novel dynamic morphed stimuli set to assess sensitivity to identity and emotion atributes in faces. *Frontiers in Psychology, 10*(757). doi:10.3389/fpsyg.2019.00757

Deckers, J. W. M., Lobbestael, J., van Wingen, G. A., Kessels, R. P. C., Arntz, A., & Egger, J. I. M. (2015). The influence of stress on social cognition in patients with borderline personality disorder. *Psychoneuroendocrinology, 52*, 119-129. https://doi.org/10.1016/j.psyneuen.2014.11.003

Demenescu, L. R., Kortekaas, R., den Boer, J. A., & Aleman, A. (2010). Impaired attribution of emotion to facial expressions in anxiety and major depression. *PLoS ONE, 5*(12), Article e15058. doi:10.1371/journal.pone.0015058

Demetriou, A., Spanoudis, G., & Mouyi, A. (2010). The development of mental processing. In W. F. Overton & R. M. Lerner (Eds.), *The handbook of life-span development, Vol. 1. Cognition, biology, and methods* (pp. 306–345). John Wiley & Sons, Inc. https://doi.org/10.1002/9780470880166.hlsd001010

Depp, C. A., Mausbach, B. T., Harvey, P. D., Bowie, C. R., Wolyniec, P. S., Thornquist, M. H., . . . Patterson, T. L. (2010). Social competence and observer-rated social functioning in bipolar disorder. *Bipolar Disorders, 12*(8), 843-850. https://doi.org/10.1111/j.1399-5618.2010.00880.x

Derntl, B., Kryspin-Exner, I., Fernbach, E., Moser, E., & Habel, U. (2008). Emotion recognition accuracy in healthy young females is associated with cycle phase. *Hormones and Behavior, 53*(1), 90-95. https://doi.org/10.1016/j.yhbeh.2007.09.006

Derntl, B., Seidel, E.-M., Kryspin-Exner, I., Hasmann, A., & Dobmeier, M. (2009). Facial emotion recognition in patients with bipolar I and bipolar II disorder. *British Journal of Clinical Psychology, 48*(4), 363-375. doi:10.1348/014466509X404845

Derntl, B., Seidel, E.-M., Schneider, F., & Habel, U. (2012). How specific are emotional deficits? A comparison of empathic abilities in schizophrenia, bipolar and depressed patients. *Schizophrenia Research, 142*(1), 58-64. https://doi.org/10.1016/j.schres.2012.09.020

DeSoto, M. C., Bumgardner, J., Close, A., & Geary, D. C. (2007). Investigating the role of hormones in theory of mind. *North American Journal of Psychology, 9*(3), 535-544.

di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental Brain Research, 91*(1), 176-180. doi:10.1007/BF00230027

Dikmen, S. S., Heaton, R. K., Grant, I., & Temkin, N. R. (1999). Test–retest reliability and practice effects of expanded Halstead–Reitan neuropsychological test battery. *Journal of the International Neuropsychological Society, 5*(4), 346-356. https://doi.org/10.1017/S1355617799544056

Dolde, S.K., 2012. Die Cambridge Mindreading (CAM) Face-Voice Battery: Validierung und erste Befunde zur Dekodierung komplexer Emotionen mit der deutschen Übersetzung der CAM Face Task [Unpublished Diplomathesis, University of Tübingen]..

Domes, G., Heinrichs, M., Michel, A., Berger, C., & Herpertz, S. C. (2007). Oxytocin improves “mind-reading” in humans. *Biological Psychiatry, 61*(6), 731-733. https://doi.org/10.1016/j.biopsych.2006.07.015

Donohoe, G., Duignan, A., Hargreaves, A., Morris, D. W., Rose, E., Robertson, D., . . . Corvin, A. (2012). Social cognition in bipolar disorder versus schizophrenia: comparability in mental state decoding deficits. *Bipolar Disorders, 14*(7), 743-748. doi:10.1111/bdi.12011

Doody, G. A., Götz, M., Johnstone, E. C., Frith, C. D., & Cunningham Owens, D. G. (1998). Theory of mind and psychoses. *Psychological Medicine, 28*(2), 397-405. doi:10.1017/S003329179700648X

Douglas, K. M., & Porter, R. J. (2010). Recognition of disgusted facial expressions in severe depression. *British Journal of Psychiatry, 197*(2), 156-157. doi:10.1192/bjp.bp.110.078113

Duman, T., Ateşçi, F. Ç., Topak, O. Z., Şendur, İ., Tümkaya, S., & Özdel, O. (2019). Theory of mind and executive functions in bipolar disorder patients and their first-degree relatives. *Journal of Clinical Psychology, 22*(4), 396-407. doi:10.5505/kpd.2019.78942

Dziobek, I., Fleck, S., Kalbe, E., Rogers, K., Hassenstab, J., Brand, M., . . . 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

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion, 6*(3-4), 169-200. doi:10.1080/02699939208411068

Ekman, P. (1993). Facial expression and emotion. *American Psychologist, 48*(4), 384-392. doi:10.1037/0003-066X.48.4.384

Ekman, P. (1999). Basic emotions, chapter 3. In In T. Dalgleish & M. J. Power (Eds.), *Handbook of cognition and emotion* (pp. 45–60). John Wiley & Sons, Ltd. https://doi.org/10.1002/0470013494.ch3

Ekman, P., & Friesen, W. V. (1976). *Pictures of facial affect.* Consulting Psychologists Press, Palo Alto, CA.

Ekman, P., & Friesen, W. V. (2003). *Unmasking the face. A guide to recognizing emotions from facial clues*. Los Altos, CA: Malor Books.

Ekman, P., Friesen, W. V., O'sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., . . . Ricci-Bitti, P. E. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology, 53*(4), 712-717. https://doi.org/10.1037/0022-3514.53.4.712

Fernandes, J. M., Cajão, R., Lopes, R., Jerónimo, R., & Barahona-Corrêa, J. B. (2018). Social cognition in schizophrenia and autism spectrum disorders: a systematic review and meta-analysis of direct comparisons. *Frontiers in Psychiatry, 9*, 504. doi:10.3389/fpsyt.2018.00504. eCollection 2018

Fett, A.-K. J., Viechtbauer, W., 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. doi:10.1016/j.neubiorev.2010.07.001

Feyerabend, J., Lüttke, S., Grosse-Wentrup, F., Wolter, S., Hautzinger, M., & Wolkenstein, L. (2018). Theory of mind in remitted bipolar disorder: Younger patients struggle in tasks of higher ecological validity. *Journal of Affective Disorders, 231*, 32-40. https://doi.org/10.1016/j.jad.2018.01.026

Fiedler, K. (2001). Affective influences on social information processing. In In J. P. Forgas (Ed.), *Handbook of affect and social cognition* (pp. 163-185). Lawrence Erlbaum Associates Publishers.

Fiorentini, C., & Viviani, P. (2011). Is there a dynamic advantage for facial expressions? *Journal of Vision, 11*(3), Article 17. https://doi.org/10.1167/11.3.17

Fleck, S. (2007). *Theory of Mind bei Patienten mit paranoider Schizophrenie und Patienten mit Asperger-Syndrom: Untersuchungen mit einem neuen filmischen Testverfahren* [Doctoral dissertation, University of Bochum].

Fletcher, P. C., Happe, F., Frith, U., Baker, S. C., Dolan, R. J., Frackowiak, R. S., & Frith, C. D. (1995). Other minds in the brain: A functional imaging study of" theory of mind" in story comprehension. *Cognition, 57*(2), 109-128. https://doi.org/10.1016/0010-0277(95)00692-R

Foland-Ross, L. C., Bookheimer, S. Y., Lieberman, M. D., Sugar, C. A., Townsend, J. D., Fischer, J., . . . Altshuler, L. L. (2012). Normal amygdala activation but deficient ventrolateral prefrontal activation in adults with bipolar disorder during euthymia. *NeuroImage, 59*(1), 738-744. https://doi.org/10.1016/j.neuroimage.2011.07.054

Foland, L. C., Altshuler, L. L., Bookheimer, S. Y., Eisenberger, N., Townsend, J., & Thompson, P. M. (2008). Evidence for deficient modulation of amygdala response by prefrontal cortex in bipolar mania. *Psychiatry Research: Neuroimaging, 162*(1), 27-37. https://doi.org/10.1016/j.pscychresns.2007.04.007

Fonagy, P., Gergely, G., Jurist, E. L., & Target, M. (2005). Affektregulierung, Mentalisierung und die Entwicklung des Selbst. *Kinderanalyse, 13*(1), 106-109.

Forgas, J. P. (1995). Mood and judgment: The affect infusion model (AIM). *Psychological Bulletin, 117*(1), 39-66. doi:10.1037/0033-2909.117.1.39

Forgas, J. P. (2002). Feeling and doing: affective influences on interpersonal behavior. *Psychological Inquiry, 13*(1), 1-28. doi:10.1207/S15327965PLI1301_01

Forgas, J. P. (2017). Mood effects on cognition: Affective influences on the content and process of information processing and behavior. In M. Jeon (Ed.), *Emotions and affect in human factors and human-computer interaction* (pp. 89-122). Elsevier Academic Press. https://doi.org/10.1016/B978-0-12-801851-4.00003-3

Forgas, J. P., & Bower, G. H. (1987). Mood effects on person-perception judgments. *Journal of Personality and Social Psychology, 53*(1), 53-60. https://doi.org/10.1037/0022-3514.53.1.53

Forgas, J. P., Bower, G. H., & Krantz, S. E. (1984). The influence of mood on perceptions of social interactions. *Journal of Experimental Social Psychology, 20*(6), 497-513. https://doi.org/10.1016/0022-1031(84)90040-4

Franke, G. H. (1994). Testtheoretische Überprüfung des Fragebogens zur sozialen Unterstützung. *Zeitschrift für medizinische Psychologie, 3*(4), 168-177.

Friedman, J. (1984). A variable span smoother. (No. LCS-TR-5). Stanford Univ CA lab for computational statistics.

Frijda, N. H. (1986). *The emotions*. Cambridge University Press ; Editions de la Maison des Sciences de l'Homme.

Frith, C. D. (2008). Social cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences, 363*(1499), 2033-2039. https://doi.org/10.1098/rstb.2008.0005

Frith, C. D., & Corcoran, R. (1996). Exploring 'theory of mind' in people with schizophrenia. *Psychological Medicine, 26*(3), 521-530. doi:10.1017/S0033291700035601

Frith, C. D., & Frith, U. (2005). Theory of mind. *Current Biology, 15*(17), R644-R645. doi:10.1016/j.cub.2005.08.041

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

Frodl, T., Scheuerecker, J., Albrecht, J., Kleemann, A. M., Müller-Schunk, S., Koutsouleris, N., . . . Meisenzahl, E. (2009). Neuronal correlates of emotional processing in patients with major depression. *The World Journal of Biological Psychiatry, 10*(3), 202-208. doi:10.1080/15622970701624603

Fulford, D., Peckham, A. D., Johnson, K., & Johnson, S. L. (2014). Emotion perception and quality of life in bipolar I disorder. *Journal of Affective Disorders, 152-154*, 491-497. https://doi.org/10.1016/j.jad.2013.08.034

Fydrich, T., Sommer, G., & Brähler, E. (2007). F-SozU, Fragebogen zur sozialen Unterstützung. Göttingen: Hogrefe.

Fydrich, T., Sommer, G., Menzel, U., & Höll, B. (1987). Social Support Questionnaire (short-form; SozU-K-22). *Zeitschrift für Klinische Psychologie und Psychotherapie, 16*, 434-436.

Fydrich, T., Sommer, G., Tydecks, S., & Brähler, E. (2009). Fragebogen zur sozialen Unterstützung (F-SozU): Normierung der Kurzform (K-14) [Social Support Questionnaire (F-SozU): Standardization of short form (K-14)]. *Zeitschrift für Medizinische Psychologie, 18*(1), 43-48.

Gaebel, W., & Wölwer, W. (1992). Facial expression and emotional face recognition in schizophrenia and depression. *European Archives of Psychiatry and Clinical Neuroscience, 242*(1), 46-52. doi:10.1007/BF02190342

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences, 2*(12), 493-501. https://doi.org/10.1016/S1364-6613(98)01262-5

Gallese, V., Keysers, C., & Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends in Cognitive Sciences, 8*(9), 396-403. https://doi.org/10.1016/j.tics.2004.07.002

Garnefski, N., & Kraaij, V. (2007). The cognitive emotion regulation questionnaire. *European Journal of Psychological Assessment, 23*(3), 141-149. https://doi.org/10.1027/1015-5759.23.3.141

Garrido-Vásquez, P., Jessen, S., & Kotz, S. A. (2011). Perception of emotion in psychiatric disorders: On the possible role of task, dynamics, and multimodality. *Social Neuroscience, 6*(5-6), 515-536. doi:10.1080/17470919.2011.620771

Getz, G. E., Shear, P. K., & Strakowski, S. M. (2003). Facial affect recognition deficits in bipolar disorder. *Journal of the International Neuropsychological Society, 9*(4), 623-632. doi:10.1017/S1355617703940021

Gitlin, M. J., Swendsen, J., Heller, T. L., & Hammen, C. (1995). Relapse and impairment in bipolar disorder. *The American Journal of Psychiatry, 152*(11), 1635-1640. doi:10.1176/ajp.152.11.1635

Godard, J., Baruch, P., Grondin, S., & Lafleur, M. F. (2012). Psychosocial and neurocognitive functioning in unipolar and bipolar depression: A 12-month prospective study. *Psychiatry Research, 196*(1), 145-153. https://doi.org/10.1016/j.psychres.2011.09.013

Golan, O., Baron-Cohen, S., & Hill, J. (2006). The Cambridge mindreading (CAM) face-voice battery: Testing complex emotion recognition in adults with and without Asperger syndrome. *Journal of Autism and Developmental Disorders, 36*(2), 169-183. doi:10.1007/s10803-005-0057-y

Golan, O., Baron-Cohen, S., Hill, J. J., & Rutherford, M. (2007). The 'Reading the Mind in the Voice'test-revised: a study of complex emotion recognition in adults with and without autism spectrum conditions. *Journal of Autism and Developmental Disorders, 37*(6), 1096-1106. doi:10.1007/s10803-006-0252-5

Gollan, J. K., McCloskey, M., Hoxha, D., & Coccaro, E. F. (2010). How do depressed and healthy adults interpret nuanced facial expressions? *Journal of Abnormal Psychology*, *119*(4), 804-810. doi:10.1037/a0020234

Gollan, J. K., Pane, H. T., McCloskey, M. S., & Coccaro, E. F. (2008). Identifying differences in biased affective information processing in major depression. *Psychiatry Research, 159*(1), 18-24. https://doi.org/10.1016/j.psychres.2007.06.011

Gotlib, I. H., & Joormann, J. (2010). Cognition and depression: current status and future directions. *Annual Review of Clinical Psychology, 6*(1), 285-312. doi:10.1146/annurev.clinpsy.121208.131305

Grant, D. A., & Berg, E. A. (1993). Wisconsin card sorting test. *Journal of Experimental Psychology*.
https://doi.org/10.1007/978-0-387-79948-3_1913

Gray, J., Venn, H., Montagne, B., Murray, L., Burt, M., Frigerio, E., . . . Young, A. H. (2006). Bipolar patients show mood-congruent biases in sensitivity to facial expressions of emotion when exhibiting depressed symptoms, but not when exhibiting manic symptoms. *Cognitive Neuropsychiatry, 11*(6), 505-520. doi:10.1080/13546800544000028

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

Green, M. F., Penn, D. L., Bentall, R., Carpenter, W. T., Gaebel, W., Gur, R. C., . . . Heinssen, R. (2008). Social cognition in schizophrenia: An NIMH workshop on definitions, assessment, and research opportunities. *Schizophrenia Bulletin, 34*(6), 1211-1220. doi:10.1093/schbul/sbm145

Greenberg, M. S., & Beck, A. T. (1989). Depression versus anxiety: A test of the content-specificity hypothesis. *Journal of Abnormal Psychology, 98*(1), 9-13. doi:10.1037/0021-843X.98.1.9

Gross, J. J. (2015). Emotion regulation: current status and future prospects. *Psychological Inquiry, 26*(1), 1-26. doi:10.1080/1047840X.2014.940781

Gruber, J. (2011). Can feeling too good be bad? Positive emotion persistence (PEP) in bipolar disorder. *Current Directions in Psychological Science, 20*(4), 217-221. https://doi.org/10.1177/0963721411414632

Gruber, S., Rathgeber, K., Bräunig, P., & Gauggel, S. (2007). Stability and course of neuropsychological deficits in manic and depressed bipolar patients compared to patients with major depression. *Journal of Affective Disorders, 104*(1), 61-71. https://doi.org/10.1016/j.jad.2007.02.011

Guhn, A., Steinacher, B., Merkl, A., Sterzer, P., & Köhler, S. (2019). Negative mood induction: Affective reactivity in recurrent, but not persistent depression. *PLoS ONE, 14*(1), e0208616. https://doi.org/10.1371/journal.pone.0213761

Haldane, M., & Frangou, S. (2006). Functional neuroimaging studies in mood disorders. *Acta Neuropsychiatrica, 18*(2), 88-99. doi:10.1111/j.1601-5215.2006.00131.x

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry, 23*(1), 56-62. doi:10.1136/jnnp.23.1.56

Hamilton, M. (1967). Development of a rating scale for primary depressive illness. *British Journal of Social and Clinical Psychology, 6*(4), 278-296. doi:10.1111/j.2044-8260.1967.tb00530.x

Hamilton, M. (1980). Rating depressive patients. *The Journal of Clinical Psychiatry, 41*(12, Sec 2), 21-24.

Hamilton, M. (2000). Hamilton rating scale for depression (Ham-D). *Handbook of psychiatric measures*, 526-528.

Hamshere, M. L., O'Donovan, M. C., Jones, I. R., Jones, L., Kirov, G., Green, E. K., . . . McQuillin, A. (2011). Polygenic dissection of the bipolar phenotype. *The British Journal of Psychiatry, 198*(4), 284-288. doi:10.1192/bjp.bp.110.087866

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. doi:10.1007/BF02172093

Harkness, K. L., Jacobson, J. A., Duong, D., & Sabbagh, M. A. (2010). Mental state decoding in past major depression: Effect of sad versus happy mood induction. *Cognition and Emotion, 24*(3), 497-513. https://doi.org/10.1080/02699930902750249

Harkness, K. L., Washburn, D., Theriault, J. E., Lee, L., & Sabbagh, M. A. (2011). Maternal history of depression is associated with enhanced theory of mind in depressed and nondepressed adult women. *Psychiatry Research, 189*(1), 91-96. https://doi.org/10.1016/j.psychres.2011.06.007

Harmer, C. J., Grayson, L., & Goodwin, G. M. (2002). Enhanced recognition of disgust in bipolar illness. *Biological Psychiatry, 51*(4), 298-304. doi:10.1016/S0006-3223(01)01249-5

Harris, P. L. (1989). *Children and emotion: The development of psychological understanding*. Basil Blackwell.

Hassan, E. (2006). Recall bias can be a threat to retrospective and prospective research designs. *The Internet Journal of Epidemiology, 3*(2), 339-412.

Healey, K. M., Bartholomeusz, C. F., & Penn, D. L. (2016). Deficits in social cognition in first episode psychosis: A review of the literature. *Clinical Psychology Review, 50*, 108-137. https://doi.org/10.1016/j.cpr.2016.10.001

Helmstaedter, C., & Durwen, H. F. (1990). VLMT: Verbaler Lern-und Merkfähigkeitstest: Ein praktikables und differenziertes Instrumentarium zur Prüfung der verbalen Gedächtnisleistungen [VLMT: A useful tool to assess and differentiate verbal memory performance]. *Schweizer Archiv für Neurologie, Neurochirurgie und Psychiatrie, 141*(1), 21-30.

Helmstaedter, C., Lendt, M., & Lux, S. (2001). *Verbaler Lern-und Merkfähigkeitstest: VLMT; Manual*. Beltz-Test.

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. doi:10.1037/a0030677

Hezel, D. M., & McNally, R. J. (2014). Theory of mind impairments in social anxiety disorder. *Behavior Therapy, 45*(4), 530-540. https://doi.org/10.1016/j.beth.2014.02.010

Hillmann, K., Neukel, C., Krauch, M., Spohn, A., Schnell, K., Herpertz, S. C., & Bertsch, K. (2021). Cognitive and affective theory of mind in female patients with borderline personality disorder. *Journal of Personality Disorders, 35*(5), 672-690. doi:10.1521/pedi_2020_34_490

Hirschfeld, R. M. A., Montgomery, S. A., Keller, M. B., Kasper, S., Schatzberg, A. F., Möller, H.-J., . . . Bourgeois, M. (2000). Social functioning in depression: A review. *The Journal of Clinical Psychiatry, 61*(4), 268-275. doi:10.4088/JCP.v61n0405

Hoertnagl, C. M., Muehlbacher, M., Biedermann, F., Yalcin, N., Baumgartner, S., Schwitzer, G., . . . Hofer, A. (2011). Facial emotion recognition and its relationship to subjective and functional outcomes in remitted patients with bipolar I disorder. *Bipolar Disorders, 13*(5-6), 537-544. doi:10.1111/j.1399-5618.2011.00947.x

Hoertnagl, C. M., Oberheinricher, S., & Hofer, A. (2014). Soziale Kognition bei PatientInnen mit affektiven Störungen. Teil II: Bipolar affektive Störung. *Neuropsychiatrie, 28*(2), 84-91. doi:10.1007/s40211-013-0096-0

Hollander, E., Bartz, J., Chaplin, W., Phillips, A., Sumner, J., Soorya, L., . . . Wasserman, S. (2007). Oxytocin increases retention of social cognition in autism. *Biological Psychiatry, 61*(4), 498-503. https://doi.org/10.1016/j.biopsych.2006.05.030

Horning, S. M., Cornwell, R. E., & Davis, H. P. (2012). The recognition of facial expressions: An investigation of the influence of age and cognition. *Aging, Neuropsychology, and Cognition, 19*(6), 657-676. doi:10.1080/13825585.2011.645011

Horowitz, L., Strauß, B., & Kordy, H. (2000). Inventar zur erfassung Interpersoneller Probleme [Inventory of interpersonal problems]. Manual (2nd ed.). Göttingen: Beltz Test.

Hosmer, J., David, W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398). John Wiley & Sons.

Howlin, P., Baron-Cohen, S., & Hadwin, J. (1999). *Teaching children with autism to mind-read: A practical guide for teachers and parents*. John Wiley & Sons.

Hulvershorn, L. A., Karne, H., Gunn, A. D., Hartwick, S. L., Wang, Y., Hummer, T. A., & Anand, A. (2012). Neural activation during facial emotion processing in unmedicated bipolar depression, euthymia, and mania. *Biological Psychiatry, 71*(7), 603-610. https://doi.org/10.1016/j.biopsych.2011.10.038

Ibanez, A., Huepe, D., Gempp, R., Gutiérrez, V., Rivera-Rei, A., & Toledo, M. I. (2013). Empathy, sex and fluid intelligence as predictors of theory of mind. *Personality and Individual Differences, 54*(5), 616-621. https://doi.org/10.1016/j.paid.2012.11.022

Ibanez, A., Urquina, H., Petroni, A., Baez, S., Lopez, V., do Nascimento, M., . . . Blenkmann, A. (2012). Neural processing of emotional facial and semantic expressions in euthymic bipolar disorder (BD) and its association with theory of mind (ToM). *PLoS ONE, 7*(10), Article e46877. https://doi.org/10.1371/journal.pone.0046877

Inoue, Y., Tonooka, Y., Yamada, K., & Kanba, S. (2004). Deficiency of theory of mind in patients with remitted mood disorder. *Journal of Affective Disorders, 82*(3), 403-409. https://doi.org/10.1016/j.jad.2004.04.004

Inoue, Y., Yamada, K., & Kanba, S. (2006). Deficit in theory of mind is a risk for relapse of major depression. *Journal of Affective Disorders, 95*(1), 125-127. https://doi.org/10.1016/j.jad.2006.04.018

Ioannidi, N., Konstantakopoulos, G., Sakkas, D., & Oulis, P. (2015). The relationship of Theory of Mind with symptoms and cognitive impairment in bipolar disorder: a prospective study. *Psychiatrike = Psychiatriki, 26*(1), 17-27.

Isaksson, J., Van't Westeinde, A., Cauvet, É., Kuja-Halkola, R., Lundin, K., Neufeld, J., . . . Bölte, S. (2019). Social cognition in autism and other neurodevelopmental disorders: A co-twin control study. *Journal of Autism and Developmental Disorders, 49*(7), 2838-2848. doi:10.1007/s10803-019-04001-4

Johnson, L., Lundström, O., Åberg-Wistedt, A., & Mathé, A. A. (2003). Social support in bipolar disorder: its relevance to remission and relapse. *Bipolar Disorders, 5*(2), 129-137. doi:10.1034/j.1399-5618.2003.00021.x

Johnston, P. J., Stojanov, W., Devir, H., & Schall, U. (2005). Functional MRI of facial emotion recognition deficits in schizophrenia and their electrophysiological correlates. *European Journal of Neuroscience, 22*(5), 1221-1232. doi:10.1111/j.1460-9568.2005.04294.x

Joormann, J., & Gotlib, I. H. (2006). Is this happiness I see? Biases in the identification of emotional facial expressions in depression and social phobia. *Journal of Abnormal Psychology, 115*(4), 705-714. doi:10.1037/0021-843X.115.4.705

Joormann, J., & Gotlib, I. H. (2007). Selective attention to emotional faces following recovery from depression. *Journal of Abnormal Psychology, 116*(1), 80. doi:10.1037/0021-843X.116.1.80

Judd, L. L., Akiskal, H. S., Zeller, P. J., Paulus, M., Leon, A. C., Maser, J. D., . . . Keller, M. B. (2000). Psychosocial disability during the long-term course of unipolar major depressive disorder. *Archives of General Psychiatry, 57*(4), 375-380. doi:10.1001/archpsyc.57.4.375

Judd, L. L., Schettler, P. J., Solomon, D. A., Maser, J. D., Coryell, W., Endicott, J., & Akiskal, H. S. (2008). Psychosocial disability and work role function compared across the long-term course of bipolar I, bipolar II and unipolar major depressive disorders. *Journal of Affective Disorders, 108*(1), 49-58. https://doi.org/10.1016/j.jad.2007.06.014

Kan, Y., Kawamura, M., Hasegawa, Y., Mochizuki, S., & Nakamura, K. (2002). Recognition of emotion from facial, prosodic and written verbal stimuli in parkinson's disease. *Cortex, 38*(4), 623-630. https://doi.org/10.1016/S0010-9452(08)70026-1

Kan, Y., Mimura, M., Kamijima, K., & Kawamura, M. (2004). Recognition of emotion from moving facial and prosodic stimuli in depressed patients. *Journal of Neurology, Neurosurgery & Psychiatry, 75*(12), 1667-1671. doi:10.1136/jnnp.2004.036079

Kanade, T., Cohn, J. F., & Tian, Y. (2000). Comprehensive database for facial expression analysis. In *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition* (Cat. No. PR00580) (pp. 46-53). IEEE.

Kanske, P., Böckler, A., & Singer, T. (2015). Models, mechanisms and moderators dissociating empathy and theory of mind. *Social Behavior from Rodents to Humans*, 193-206. doi:10.1007/7854_2015_412

Kardum, I., & Daskijević, K. T. (2001). Absolute and relative accuracy in the retrospective estimate of positive and negative mood. *European Journal of Psychological Assessment, 17*(1), 69-77. doi:10.1027/1015-5759.17.1.69

Kasari, C., Chamberlain, B., & Bauminger, N. (2001). Social emotions and social relationships: Can children with autism compensate? In J. A. Burack, T. Charman, N. Yirmiya, & P. R. Zelazo (Eds.), *The development of autism: Perspectives from theory and research* (pp. 309-323). Lawrence Erlbaum Associates Publishers.

Keller, M. B., Lavori, P. W., Friedman, B., Nielsen, E., Endicott, J., McDonald-Scott, P., & Andreasen, N. C. (1987). The longitudinal interval follow-up evaluation: a comprehensive method for assessing outcome in prospective longitudinal studies. *Archives of General Psychiatry, 44*(6), 540-548. doi:10.1001/archpsyc.1987.01800180050009

Kempf, L., Hussain, N., & Potash, J. B. (2005). Mood disorder with psychotic features, schizoaffective disorder, and schizophrenia with mood features: trouble at the borders. *International Review of Psychiatry, 17*(1), 9-19. doi:10.1080/09540260500064959

Kerr, N., Dunbar, R. I. M., & Bentall, R. P. (2003). Theory of mind deficits in bipolar affective disorder. *Journal of Affective Disorders, 73*(3), 253-259. doi: 10.1016/S0165-0327(02)00008-3

Kessels, R. P., Montagne, B., Hendriks, A. W., Perrett, D. I., & de Haan, E. H. (2014). Assessment of perception of morphed facial expressions using the emotion recognition task: Normative data from healthy participants aged 8–75. *Journal of Neuropsychology, 8*(1), 75-93. doi:10.1111/jnp.12009

Kettle, J. W. L., O'Brien-Simpson, L., & Allen, N. B. (2008). Impaired theory of mind in first-episode schizophrenia: comparison with community, university and depressed controls. *Schizophrenia Research, 99*(1), 96-102. https://doi.org/10.1016/j.schres.2007.11.011

Keysers, C., & Gazzola, V. (2010). Social neuroscience: mirror neurons recorded in humans. *Current Biology, 20*(8), R353-R354. https://doi.org/10.1016/j.cub.2010.03.013

Kilts, C. D., Egan, G., Gideon, D. A., Ely, T. D., & Hoffman, J. M. (2003). Dissociable neural pathways are involved in the recognition of emotion in static and dynamic facial expressions. *NeuroImage, 18*(1), 156-168. https://doi.org/10.1006/nimg.2002.1323

Kinderman, P., Dunbar, R., & Bentall, R. P. (1998). Theory-of-mind deficits and causal attributions. *British Journal of Psychology, 89*(2), 191-204. https://doi.org/10.1111/j.2044-8295.1998.tb02680.x

Kobak, K., Williams, J., & Rush, A. (2007). *Strukturierte Interview-Richtlinien für die Hamilton Depression Scale (SIGH-D) und das Inventar depressiver Symptome (IDS_C)(SIGHD-IDS).*

Krüger, S., Bräunig, P., & Shugar, G. (1997). *Manie-Selbstbeurteilungsskala: MSS; Manual*. Beltz Test.

Kuehner, C., Huffziger, S., & Liebsch, K. (2009). Rumination, distraction and mindful self-focus: Effects on mood, dysfunctional attitudes and cortisol stress response. *Psychological Medicine, 39*(2), 219-228. doi:10.1017/S0033291708003553

Kupferberg, A., Bicks, L., & Hasler, G. (2016). Social functioning in major depressive disorder. *Neuroscience & Biobehavioral Reviews, 69*, 313-332. doi:10.1016/j.neubiorev.2016.07.002

Kurtz, M. M., & Gerraty, R. T. (2009). A meta-analytic investigation of neurocognitive deficits in bipolar illness: Profile and effects of clinical state. *Neuropsychology, 23*(5), 551-562. doi:10.1037/a0016277

Ladegaard, N., Larsen, E. R., Videbech, P., & Lysaker, P. H. (2014). Higher-order social cognition in first-episode major depression. *Psychiatry Research, 216*(1), 37-43. https://doi.org/10.1016/j.psychres.2013.12.010

Ladegaard, N., Videbech, P., Lysaker, P. H., & Larsen, E. R. (2016). The course of social cognitive and metacognitive ability in depression: Deficit are only partially normalized after full remission of first episode major depression. *British Journal of Clinical Psychology, 55*(3), 269-286. doi:10.1111/bjc.12097

Lahera, G., Benito, A., Montes, J. M., Fernández-Liria, A., Olbert, C. M., & Penn, D. L. (2013). Social cognition and interaction training (SCIT) for outpatients with bipolar disorder. *Journal of Affective Disorders, 146*(1), 132-136. https://doi.org/10.1016/j.jad.2012.06.032

Lahera, G., Herrera, S., Reinares, M., Benito, A., Rullas, M., González-Cases, J., & Vieta, E. (2015). Hostile attributions in bipolar disorder and schizophrenia contribute to poor social functioning. *Acta Psychiatrica Scandinavica, 131*(6), 472-482. doi: 10.1111/acps.12399

Lahera, G., Montes, J. M., Benito, A., Valdivia, M., Medina, E., Mirapeix, I., & Sáiz-Ruiz, J. (2008). Theory of mind deficit in bipolar disorder: Is it related to aprevious history of psychotic symptoms? *Psychiatry Research, 161*(3), 309-317. doi: 10.1016/j.psychres.2007.08.009

Lahera, G., Ruiz-Murugarren, S., Iglesias, P., Ruiz-Bennasar, C., Herrería, E., Montes, J. M., & Fernández-Liria, A. (2012). Social cognition and global functioning in bipolar disorder. *The Journal of Nervous and Mental Disease, 200*(2), 135-141. doi:10.1097/NMD.0b013e3182438eae

Langenecker, S. A., Bieliauskas, L. A., Rapport, L. J., Zubieta, J.-K., Wilde, E. A., & Berent, S. (2005). Face emotion perception and executive functioning deficits in depression. *Journal of Clinical and Experimental Neuropsychology, 27*(3), 320-333. doi:10.1080/13803390490490515720

Lawlor-Savage, L., Sponheim, S. R., & Goghari, V. M. (2014). Impaired recognition of happy facial expressions in bipolar disorder. *Acta Neuropsychiatrica, 26*(4), 253-259. doi:10.1017/neu.2014.6

Lazarus, R. S. (1968). Emotions and adaptation: conceptual and empirical relations. *Nebraska Symposium on Motivation, 16*, 175-266.

Lazarus, R. S. (1991). *Emotion and adaptation*. Oxford University Press.

Lee, J., Altshuler, L., Glahn, D. C., Miklowitz, D. J., Ochsner, K., & Green, M. F. (2013). Social and nonsocial cognition in bipolar disorder and schizophrenia: relative levels of impairment. *American Journal of Psychiatry, 170*(3), 334-341. doi:10.1176/appi.ajp.2012.12040490

Lee, L., Harkness, K. L., Sabbagh, M. A., & Jacobson, J. A. (2005). Mental state decoding abilities in clinical depression. *Journal of Affective Disorders, 86*(2), 247-258. https://doi.org/10.1016/j.jad.2005.02.007

Lehrl, S. (2005). Mehrfachwahl-Wortschatz-Intelligenztest MWT-B. Balingen: Spitta Verlag.

Leiberg, S., & Anders, S. (2006). The multiple facets of empathy: a survey of theory and evidence. In S. Anders, G. Ende, M. Junghofer, J. Kissler, & D. Wildgruber (Eds.), *Progress in Brain Research* (Vol. 156, pp. 419-440). Elsevier.

Lembke, A., & Ketter, T. A. (2002). Impaired recognition of facial emotion in mania. *American Journal of Psychiatry, 159*(2), 302-304. doi:10.1176/appi.ajp.159.2.302

LeMoult, J., Joormann, J., Sherdell, L., Wright, Y., & Gotlib, I. H. (2009). Identification of emotional facial expressions following recovery from depression. *Journal of Abnormal Psychology, 118*(4), 828-833. https://doi.org/10.1037/a0016944

Leppänen, J. M., Milders, M., Bell, J. S., Terriere, E., & Hietanen, J. K. (2004). Depression biases the recognition of emotionally neutral faces. *Psychiatry Research, 128*(2), 123-133. https://doi.org/10.1016/j.psychres.2004.05.020

Lezak, M. D. (1976). *Neuropsychological assessment.* Oxford University Press, New York, NY.

Lezak, M. D. (1983). *Neuropsychology assessment(2nd edn.).* Oxford University Press, New York, NY.

Lezak, M. D. (1995). *Neuropsychological assessment (3rd edn.).* Oxford University Press, New York, NY.

Lichtenstein, P., Yip, B. H., Björk, C., Pawitan, Y., Cannon, T. D., Sullivan, P. F., & Hultman, C. M. (2009). Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *The Lancet, 373*(9659), 234-239. doi:10.1016/S0140-6736(09)60072-6

Likowski, K. U., Weyers, P., Seibt, B., Stöhr, C., Pauli, P., & Mühlberger, A. (2011). Sad and lonely? Sad mood suppresses facial mimicry. *Journal of Nonverbal Behavior, 35*(2), 101-117. https://doi.org/10.1007/s10919-011-0107-4

Liu, W., Fan, J., Gan, J., Lei, H., Niu, C., Chan, R. C. K., & Zhu, X. (2017). Disassociation of cognitive and affective aspects of theory of mind in obsessive-compulsive disorder. *Psychiatry Research, 255*, 367-372. https://doi.org/10.1016/j.psychres.2017.06.058

López-Jaramillo, C., Lopera-Vásquez, J., Gallo, A., Ospina-Duque, J., Bell, V., Torrent, C., . . . Vieta, E. (2010). Effects of recurrence on the cognitive performance of patients with bipolar I disorder: implications for relapse prevention and treatment adherence. *Bipolar Disorders, 12*(5), 557-567. doi:10.1111/j.1399-5618.2010.00835.x

Ludyga, S., Schilling, R., Colledge, F., Brand, S., Pühse, U., & Gerber, M. (2020). Association between cardiorespiratory fitness and social cognition in healthy adults. *Scandinavian Journal of Medicine & Science in Sports, 30*(9), 1722-1728. doi:10.1111/sms.13730.

Lyon, H. M., Startup, M., & Bentall, R. P. (1999). Social cognition and the manic defense: Attributions, selective attention, and self-schema in bipolar affective disorder. *Journal of Abnormal Psychology, 108*(2), 273-282. doi:10.1037/0021-843X.108.2.273

MacQueen, G. M., & Memedovich, K. A. (2017). Cognitive dysfunction in major depression and bipolar disorder: Assessment and treatment options. *Psychiatry and Clinical Neurosciences, 71*(1), 18-27. https://doi.org/10.1111/pcn.12463

MacQueen, G. M., Young, L. T., & Joffe, R. T. (2001). A review of psychosocial outcome in patients with bipolar disorder. *Acta Psychiatrica Scandinavica, 103*(3), 163-170. doi:10.1034/j.1600-0447.2001.00059.x

MacQueen, G. M., Young, L. T., Robb, J. C., Marriott, M., Cooke, R. G., & Joffe, R. T. (2000). Effect of number of episodes on wellbeing and functioning of patients with bipolar disorder. *Acta Psychiatrica Scandinavica, 101*(5), 374-381. doi:10.1034/j.1600-0447.2000.101005374.x

Malhi, G. S., Lagopoulos, J., Das, P., Moss, K., Berk, M., & Coulston, C. M. (2008). A functional MRI study of theory of mind in euthymic bipolar disorder patients. *Bipolar Disorders, 10*(8), 943-956. doi:10.1111/j.1399-5618.2008.00643.x

Mann-Wrobel, M. C., Carreno, J. T., & Dickinson, D. (2011). Meta-analysis of neuropsychological functioning in euthymic bipolar disorder: an update and investigation of moderator variables. *Bipolar Disorders, 13*(4), 334-342. doi:10.1111/j.1399-5618.2011.00935.x

Mansell, W., Colom, F., & Scott, J. (2005). The nature and treatment of depression in bipolar disorder: A review and implications for future psychological investigation. *Clinical Psychology Review, 25*(8), 1076-1100. https://doi.org/10.1016/j.cpr.2005.06.007

Marazziti, D., Consoli, G., Picchetti, M., Carlini, M., & Faravelli, L. (2010). Cognitive impairment in major depression. *European Journal of Pharmacology, 626*(1), 83-86. https://doi.org/10.1016/j.ejphar.2009.08.046

Martinez-Aran, A., Vieta, E., Torrent, C., Sanchez-Moreno, J., Goikolea, J., Salamero, M., . . . Alvarez-Grandi, S. (2007). Functional outcome in bipolar disorder: the role of clinical and cognitive factors. *Bipolar Disorders, 9*(1-2), 103-113. doi:10.1111/j.1399-5618.2007.00327.x

Martinez, G., Alexandre, C., Mam-Lam-Fook, C., Bendjemaa, N., Gaillard, R., Garel, P., . . . Krebs, M.-O. (2017). Phenotypic continuum between autism and schizophrenia: Evidence from the Movie for the Assessment of Social Cognition (MASC). *Schizophrenia Research, 185*, 161-166. https://doi.org/10.1016/j.schres.2017.01.012

Martino, D. J., Strejilevich, S. A., Fassi, G., Marengo, E., & Igoa, A. (2011). Theory of mind and facial emotion recognition in euthymic bipolar I and bipolar II disorders. *Psychiatry Research, 189*(3), 379-384. https://doi.org/10.1016/j.psychres.2011.04.033

Martino, D. J., Strejilevich, S. A., Scápola, M., Igoa, A., Marengo, E., Ais, E. D., & Perinot, L. (2008). Heterogeneity in cognitive functioning among patients with bipolar disorder. *Journal of Affective Disorders, 109*(1), 149-156. https://doi.org/10.1016/j.jad.2007.12.232

Marton, P., Connolly, J., Kutcher, S., & Korenblum, M. (1993). Cognitive social skills and social self-appraisal in depressed adolescents. *Journal of the American Academy of Child & Adolescent Psychiatry, 32*(4), 739-744. doi:https://doi.org/10.1097/00004583-199307000-00006

Matarazzo, J. D., Wiens, A. N., Matarazzo, R. G., & Goldstein, S. G. (1974). Psychometric and clinical test-retest reliability of the Halstead impairment index in a sample of healthy, young, normal men. *Journal of Nervous and Mental Disease, 158*(1), 37–49. https://doi.org/10.1097/00005053-197401000-00006

Mathews, S., Williams, B., & Nedeljkovic, M. (2020). Temporal self-appraisal in depression and anxiety: A comparison between a clinical and non-clinical sample. *Australian Psychologist, 55*(4), 389-396. doi:10.1111/ap.12442

Mattern, M., Walter, H., Hentze, C., Schramm, E., Drost, S., Schoepf, D., . . . Schnell, K. (2015). Behavioral evidence for an impairment of affective theory of mind capabilities in chronic depression. *Psychopathology, 48*(4), 240-250. doi:10.1159/000430450

Mauss, I. B., Levenson, R. W., McCarter, L., Wilhelm, F. H., & Gross, J. J. (2005). The tie that binds? Coherence among emotion experience, behavior, and physiology. *Emotion, 5*(2), 175-190. doi:10.1037/1528-3542.5.2.175

Mayes, A. K., Pipingas, A., Silberstein, R. B., & Johnston, P. (2009). Steady state visually evoked potential correlates of static and dynamic emotional face processing. *Brain Topography, 22*(3), 145-157. doi:10.1007/s10548-009-0106-5

McClintock, S. M., Husain, M. M., Greer, T. L., & Cullum, C. M. (2010). Association between depression severity and neurocognitive function in major depressive disorder: A review and synthesis. *Neuropsychology, 24*(1), 9-34. doi:10.1037/a0017336

McCullough Jr, J. P. (2003). Treatment for chronic depression: Cognitive behavioral analysis system of psychotherapy (CBASP). *Journal of Psychotherapy Integration, 13*(3-4), 241-263. doi:10.1037/1053-0479.13.3-4.241

McDonald, S. (2002). *Awareness of social inference test: manual*. Thames valley test company.

McKinnon, M. C., Cusi, A. M., & MacQueen, G. M. (2010). Impaired theory of mind performance in patients with recurrent bipolar disorder: Moderating effect of cognitive load. *Psychiatry Research, 177*(1), 261-262. https://doi.org/10.1016/j.psychres.2010.02.004

McKinnon, M. C., Cusi, A. M., & MacQueen, G. M. (2013). Psychological factors that may confer risk for bipolar disorder. *Cognitive Neuropsychiatry, 18*(1-2), 115-128. https://doi.org/10.1080/13546805.2012.702505

Meltzoff, A. N., & Brooks, R. (2001). "Like me" as a building block for understanding other minds: Bodily acts, attention, and intention. *Intentions and intentionality: Foundations of social cognition* (pp. 171–191). The MIT Press.

Mezulis, A. H., Abramson, L. Y., Hyde, J. S., & Hankin, B. L. (2004). Is there a universal positivity bias in attributions? A meta-analytic review of individual, developmental, and cultural differences in the self-serving attributional bias. *Psychological Bulletin, 130*(5), 711-747. doi:10.1037/0033-2909.130.5.711

Miklowitz, D. J., Simoneau, T. L., Sachs-Ericsson, N., Warner, R., & et al. (1996). Family risk indicators in the course of bipolar affective disorder. In C. Mundt, & H. L. Freeman (Eds.), *Interpersonal factors in the origin and course of affective disorders* (pp. 204-217). Gaskell/Royal College of Psychiatrists.

Milders, M., Bell, S., Platt, J., Serrano, R., & Runcie, O. (2010). Stable expression recognition abnormalities in unipolar depression. *Psychiatry Research, 179*(1), 38-42. https://doi.org/10.1016/j.psychres.2009.05.015

Mill, A., Allik, J., Realo, A., & Valk, R. (2009). Age-related differences in emotion recognition ability: A cross-sectional study. *Emotion, 9*(5), 619-630. doi:10.1037/a0016562

Montag, C., Dziobek, I., Richter, I. S., Neuhaus, K., Lehmann, A., Sylla, R., . . . Gallinat, J. (2011). Different aspects of theory of mind in paranoid schizophrenia: Evidence from a video-based assessment. *Psychiatry Research, 186*(2), 203-209. doi: 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), 264-269. https://doi.org/10.1016/j.jad.2009.08.017

Montagne, B., Kessels, R. P. C., Frigerio, E., de Haan, E. H. F., & Perrett, D. I. (2005). Sex differences in the perception of affective facial expressions: Do men really lack emotional sensitivity? *Cognitive Processing, 6*(2), 136-141. doi:10.1007/s10339-005-0050-6

Moore, M. T., & Fresco, D. M. (2012). Depressive realism: A meta-analytic review. *Clinical Psychology Review, 32*(6), 496-509. doi:10.1016/j.cpr.2012.05.004

Muehlbacher, M., Egger, C., Kaplan, P., Simhandl, C., Grunze, H., Geretsegger, C., . . . Stuppaeck, C. (2011). Relibilität und Übereinstimmungsvalidität der deutschen Version der Young Mania Rating Scale (YMRS-D). *Neuropsychiatrie, 25*(1), 1-10.

Mur, M., Portella, M. J., Martinez-Aran, A., Pifarré, J., & Vieta, E. (2008). Long-term stability of cognitive impairment in bipolar disorder: a 2-year follow-up study of lithium-treated euthymic bipolar patients. *Journal of Clinical Psychiatry, 69*(5), 712-719.

Murray, L. K. (2000). *Facial expression perception in different psychiatric groups* [Doctoral dissertation, University of St Andrews].

Naranjo, C., Kornreich, C., Campanella, S., Noël, X., Vandriette, Y., Gillain, B., . . . Constant, E. (2011). Major depression is associated with impaired processing of emotion in music as well as in facial and vocal stimuli. *Journal of Affective Disorders, 128*(3), 243-251. https://doi.org/10.1016/j.jad.2010.06.039

Nejati, V., Zabihzadeh, A., Maleki, G., & Tehranchi, A. (2012). Mind reading and mindfulness deficits in patients with major depression disorder. *Procedia-Social and Behavioral Sciences, 32*, 431-437. https://doi.org/10.1016/j.sbspro.2012.01.065

Németh, N., Mátrai, P., Hegyi, P., Czéh, B., Czopf, L., Hussain, A., . . . 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

Neves, M. d. C. L., Albuquerque, M. R., Malloy-Diniz, L., Nicolato, R., Silva Neves, F., de Souza-Duran, F. L., . . . Corrêa, H. (2015). A voxel-based morphometry study of gray matter correlates of facial emotion recognition in bipolar disorder. *Psychiatry Research: Neuroimaging, 233*(2), 158-164. https://doi.org/10.1016/j.pscychresns.2015.05.009

Newman, C. F., Leahy, R. L., Beck, A. T., Reilly-Harrington, N. A., & Gyulai, L. (2002). *Bipolar disorder: A cognitive therapy approach*. American Psychological Association: Washington, DC.

Newman, L. S., Duff, K. J., & Baumeister, R. F. (1997). A new look at defensive projection: Thought suppression, accessibility, and biased person perception. *Journal of Personality and Social Psychology, 72*(5), 980-1001. https://doi.org/10.1037/0022-3514.72.5.980

Nickerson, R. S. (1999). How we know—and sometimes misjudge—what others know: Imputing one's own knowledge to others. *Psychological Bulletin, 125*(6), 737-759. https://doi.org/10.1037/0033-2909.125.6.737

Olejnik, S., & Algina, J. (2003). Generalized eta and omega squared statistics: measures of effect size for some common research designs. *Psychological Methods, 8*(4), 434-447. doi:10.1037/1082-989X.8.4.434

Olley, A. L., Malhi, G. S., Bachelor, J., Cahill, C. M., Mitchell, P. B., & Berk, M. (2005). Executive functioning and theory of mind in euthymic bipolar disorder. *Bipolar Disorders, 7*, 43-52. doi:10.1111/j.1399-5618.2005.00254.x

Ortony, A., Clore, G. L., & Collins, A. (1988). *The cognitive structure of emotions.* Cambridge University Press. https://doi.org/10.1017/CBO9780511571299

Owen, M. J., Craddock, N., & Jablensky, A. (2007). The genetic deconstruction of psychosis. *Schizophrenia Bulletin, 33*(4), 905-911. doi:10.1093/schbul/sbm053

Ozel-Kizil, E., Baskak, B., Uran, P., Cihan, B., Zivrali, E., Ates, E., & Cangoz, B. (2012). Recognition of faux pas dysfunction in patients with schizophrenia, bipolar disorder, their unaffected relatives and healthy controls. *European Neuropsychopharmacology*(22), 306.

Pearlin, L. I. (1999). Stress and mental health: A conceptual overview. In A. V. Horwitz, & T. L. Scheid (Eds.), *A handbook for the study of mental health: Social contexts, theories, and systems* (pp. 161-175). Cambridge University Press.

Pearlson, G. D. (2015). Etiologic, phenomenologic, and endophenotypic overlap of schizophrenia and bipolar disorder. *Annual Review of Clinical Psychology, 11*, 251-281. doi: 10.1146/annurev-clinpsy-032814-112915

Perlis, R. H., Ostacher, M. J., Patel, J. K., Marangell, L. B., Zhang, H., Wisniewski, S. R., . . . Gyulai, L. (2006). Predictors of recurrence in bipolar disorder: primary outcomes from the Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD). *American Journal of Psychiatry, 163*(2), 217-224.

Pessoa, L., Padmala, S., & Morland, T. (2005). Fate of unattended fearful faces in the amygdala is determined by both attentional resources and cognitive modulation. *NeuroImage, 28*(1), 249-255. https://doi.org/10.1016/j.neuroimage.2005.05.048

Phillips, L. H., MacLean, R. D. J., & Allen, R. (2002). Age and the understanding of emotions: neuropsychological and sociocognitive perspectives. *The Journals of Gerontology: Series B, 57*(6), 526-530. doi:10.1093/geronb/57.6.P526

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

Platt, J. J., & Spivack, G. (1975). *Manual for the Means-Ends Problem-Solving Procedure (MEPS): A measure of interpersonal cognitive problem-solving skill.* Philadelphia: Hahnemann Medical College and Hospital.

Porcelli, S., Van Der Wee, N., van der Werff, S., Aghajani, M., Glennon, J. C., van Heukelum, S., . . . Lobo, E. (2019). Social brain, social dysfunction and social withdrawal. *Neuroscience & Biobehavioral Reviews, 97*, 10-33. https://doi.org/10.1016/j.neubiorev.2018.09.012

Porter, R. J., Robinson, L. J., Malhi, G. S., & Gallagher, P. (2015). The neurocognitive profile of mood disorders – a review of the evidence and methodological issues. *Bipolar Disorders, 17*(S2), 21-40. https://doi.org/10.1111/bdi.12342

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1*(4), 515-526. doi:10.1017/S0140525X00076512

Prkachin, G. C. (2003). The effects of orientation on detection and identification of facial expressions of emotion. *British Journal of Psychology, 94*(1), 45-62.

Purcell, A. L., Phillips, M., & Gruber, J. (2013). In your eyes: Does theory of mind predict impaired life functioning in bipolar disorder? *Journal of Affective Disorders, 151*(3), 1113-1119. https://doi.org/10.1016/j.jad.2013.06.051

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Ratcliffe, M. (2006). Folk psychology' is not folk psychology. *Phenomenology and the Cognitive Sciences, 5*(1), 31-52. doi:10.1007/s11097-005-9010-y

Rehman, U. S., Ginting, J., Karimiha, G., & Goodnight, J. A. (2010). Revisiting the relationship between depressive symptoms and marital communication using an experimental paradigm: The

moderating effect of acute sad mood. *Behaviour Research and Therapy, 48*(2), 97-105. https://doi.org/10.1016/j.brat.2009.09.013

Reitan, R. (1992). TMT, Trail Making Test A & B. South Tucson, AR: Reitan Neuropsychology Laboratory.

Rey, A. (1964). *L'examen clinique en psychologie.* [The clinical examination in psychology]. Presses Universitaries De France.

Rich, B. A., Grimley, M. E., Schmajuk, M., Blair, K. S., Blair, R. J. R., & Leibenluft, E. (2008). Face emotion labeling deficits in children with bipolar disorder and severe mood dysregulation. *Development and Psychopathology, 20*(2), 529-546. doi:10.1017/S0954579408000266

Richter, D., Dietzel, C., & Kunzmann, U. (2011). Age differences in emotion recognition: The task matters. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 66*(1), 48-55. doi:10.1093/geronb/gbq068

Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Neurosciences, 21*(5), 188-194. https://doi.org/10.1016/S0166-2236(98)01260-0

Roberts, D. L., Penn, D. L., & Combs, D. R. (2006). *Social cognition and interaction training.* Unpublished treatment manual. University of North Carolina, Chapel Hill, NC.

Robinson, J. L., Monkul, E. S., Tordesillas-Gutiérrez, D., Franklin, C., Bearden, C. E., Fox, P. T., & Glahn, D. C. (2008). Fronto-limbic circuitry in euthymic bipolar disorder: Evidence for prefrontal hyperactivation. *Psychiatry Research: Neuroimaging, 164*(2), 106-113. https://doi.org/10.1016/j.pscychresns.2007.12.004

Robinson, L. J. (2010). *Neuropsychological performance, emotion processing and psychosocial function in bipolar disorder* [Doctoral dissertation, Newcastle University].

Roiser, J., Farmer, A., Lam, D., Burke, A., O'Neill, N., Keating, S., . . . McGuffin, P. (2009). The effect of positive mood induction on emotional processing in euthymic individuals with bipolar disorder and controls. *Psychological Medicine, 39*(5), 785-791. doi:10.1017/S0033291708004200

Rojas, R., Geissner, E., & Hautzinger, M. (2014). Kognitive Reaktivität und Stressbelastung als Prädiktoren eines Rezidivs bei remittiert depressiven Personen [Cognitive reactivity and stressful life events as predictors of relapse in remitted depressed persons]. *Zeitschrift für Klinische Psychologie und Psychotherapie: Forschung und Praxis, 43*(1), 17–26. https://doi.org/10.1026/1616-3443/a000251

Rosa, A. R., Sánchez-Moreno, J., Martínez-Aran, A., Salamero, M., Torrent, C., Reinares, M., . . . Ayuso-Mateos, J. L. (2007). Validity and reliability of the Functioning Assessment Short Test (FAST) in bipolar disorder. *Clinical Practice and Epidemiology in Mental Health, 3*(1), 1-8. doi:10.1186/1745-0179-3-5

Rossell, S. L., & Van Rheenen, T. E. (2013). Theory of mind performance using a story comprehension task in bipolar mania compared to schizophrenia and healthy controls. *Cognitive Neuropsychiatry, 18*(5), 409-421. doi:10.1080/13546805.2012.725820

Rowland, J., Hamilton, M., Vella, N., Lino, B., Mitchell, P., & Green, M. (2013). Adaptive associations between social cognition and emotion regulation are absent in schizophrenia and bipolar disorder. *Frontiers in Psychology, 3*(607). doi:10.3389/fpsyg.2012.00607

Rubinow, D. R., & Post, R. M. (1992). Impaired recognition of affect in facial expression in depressed patients. *Biological Psychiatry, 31*(9), 947-953. doi:10.1016/0006-3223(92)90120-O

Ruffman, T., Henry, J. D., Livingstone, V., & Phillips, L. H. (2008). A meta-analytic review of emotion recognition and aging: Implications for neuropsychological models of aging. *Neuroscience & Biobehavioral Reviews, 32*(4), 863-881. https://doi.org/10.1016/j.neubiorev.2008.01.001

Rush, A. J., Gullion, C. M., Basco, M. R., Jarrett, R. B., & Trivedi, M. H. (1996). The Inventory of Depressive Symptomatology (IDS): psychometric properties. *Psychological Medicine, 26*(3), 477-486. doi:10.1017/S0033291700035558

Rush, A. J., Trivedi, M. H., Ibrahim, H. M., Carmody, T. J., Arnow, B., Klein, D. N., . . . Manber, R. (2003). The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): A psychometric evaluation in patients with chronic major depression. *Biological Psychiatry, 54*(5), 573-583. https://doi.org/10.1016/S0006-3223(02)01866-8

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*(6), 1161-1178. doi:10.1037/h0077714

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

Sakarya, A., & Ozgüven, H. (2012). *Association of ToM deficits with insight and other cognitive functions among remitted schizophrenia and bipolar disorder patients* [Doctoral dissertation, Ankara University].

Samamé, C. (2013). Social cognition throughout the three phases of bipolar disorder: A state-of-the-art overview. *Psychiatry Research, 210*(3), 1275-1286. doi: 10.1016/j.psychres.2013.08.012

Sánchez-Cubillo, I., Periáñez, J. A., Adrover-Roig, D., Rodríguez-Sánchez, J. M., Ríos-Lago, M., Tirapu, J. E. E. A., & Barceló, F. (2009). Construct validity of the Trail Making Test: role of task-switching, working memory, inhibition/interference control, and visuomotor abilities. *Journal of the International Neuropsychological Society, 15*(3), 438-450. doi:10.1017/S1355617709090626

Santos, J. M., Pousa, E., Soto, E., Comes, A., Roura, P., Arrufat, F. X., & Obiols, J. E. (2017). Theory of mind in euthymic bipolar patients and first-degree relatives. *The Journal of Nervous and Mental Disease, 205*(3), 207-212. doi:10.1097/nmd.0000000000000595

Sarfati, Y., Hardy-Baylé, M.-C., Brunet, E., & Widlöcher, D. (1999). Investigating theory of mind in schizophrenia: influence of verbalization in disorganized and non-disorganized patients. *Schizophrenia Research, 37*(2), 183-190. https://doi.org/10.1016/S0920-9964(98)00154-6

Saß, H., Wittchen, H.-U., & Zaudig, M. (1996). Diagnostisches und statistisches Manual psychischer Störungen-DSM-IV. Deutsche Bearbeitung u. Einleitung von H. Sass, H.-U. Wittchen, M. Zaudig.

Schaefer, K. L., Baumann, J., Rich, B. A., Luckenbaugh, D. A., & Zarate, C. A. (2010). Perception of facial emotion in adults with bipolar or unipolar depression and controls. *Journal of Psychiatric Research, 44*(16), 1229-1235. https://doi.org/10.1016/j.jpsychires.2010.04.024

Scherer, K. R. (1982). Emotion as a process: Function, origin and regulation. *Social Science Information, 21*(4-5), 555-570. doi:10.1177/053901882021004004

Scherer, K. R. (1984a). Emotion as a multicomponent process: A model and some cross-cultural data. *Review of Personality & Social Psychology, 5*, 37-63.

Scherer, K. R. (1984b). On the nature and function of emotion: A component process approach. *Approaches to Emotion, 2293*(317), 31.

Scherer, K. R. (1993). Neuroscience projections to current debates in emotion psychology. *Cognition and Emotion, 7*(1), 1-41. doi:10.1080/02699939308409174

Scherer, K. R. (2000). Psychological models of emotion. *The Neuropsychology of Emotion, 137*(3), 137-162.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences, 36*(4), 393-414. doi:10.1017/S0140525X12000660

Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences, 10*(1), 24-30. https://doi.org/10.1016/j.tics.2005.11.009

Schlegel, K., Palese, T., Mast, M. S., Rammsayer, T. H., Hall, J. A., & Murphy, N. A. (2020). A meta-analysis of the relationship between emotion recognition ability and intelligence. *Cognition and Emotion, 34*(2), 329-351. doi:10.1080/02699931.2019.1632801

Schmid, P. C., & Schmid Mast, M. (2010). Mood effects on emotion recognition. *Motivation and Emotion, 34*(3), 288-292. doi:10.1007/s11031-010-9170-0

Schrader, G., Davis, A., Stefanovic, S., & Christie, P. (1990). The recollection of affect. *Psychological Medicine, 20*(1), 105-109. doi:10.1017/S0033291700013271

Schreiter, S., Pijnenborg, G. H. M., & aan het Rot, M. (2013). Empathy in adults with clinical or subclinical depressive symptoms. *Journal of Affective Disorders, 150*(1), 1-16. https://doi.org/10.1016/j.jad.2013.03.009

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: a meta-analysis of functional brain imaging studies. *Neuroscience & Biobehavioral Reviews, 42*, 9-34. doi:10.1016/j.neubiorev.2014.01.009

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. doi:10.1177/1362361314526004

Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology, 45*(3), 513-523. doi:10.1037/0022-3514.45.3.513

Schwert, C., Stohrer, M., Aschenbrenner, S., Weisbrod, M., & Schröder, A. (2018). Biased neurocognitive self-perception in depressive and in healthy persons. *Journal of Affective Disorders, 232*, 96-102. https://doi.org/10.1016/j.jad.2018.02.031

Scott, J. (2001). Cognitive therapy as an adjunct to medication in bipolar disorder. *British Journal of Psychiatry, 178*(S41), 164-168. doi:10.1192/bjp.178.41.s164

Segal, Z. V., Gemar, M., & Williams, S. (1999). Differential cognitive response to a mood challenge following successful cognitive therapy or pharmacotherapy for unipolar depression. *Journal of Abnormal Psychology, 108*(1), 3-10. doi:10.1037/0021-843X.108.1.3

Segal, Z. V., Kennedy, S., Gemar, M., Hood, K., Pedersen, R., & Buis, T. (2006). Cognitive reactivity to sad mood provocation and the prediction of depressive relapse. *Archives of General Psychiatry, 63*(7), 749-755. doi:10.1001/archpsyc.63.7.749

Seidel, E.-M., Habel, U., Finkelmeyer, A., Schneider, F., Gur, R. C., & Derntl, B. (2010). Implicit and explicit behavioral tendencies in male and female depression. *Psychiatry Research, 177*(1), 124-130. doi:https://doi.org/10.1016/j.psychres.2010.02.001

Shamay-Tsoory, S., Aharon-Peretz, J., & Perry, D. (2009). Two systems for empathy: a double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. *Brain, 132*(3), 617-627. doi:10.1093/brain/awn279

Shamay-Tsoory, S., Harari, H., Szepsenwol, O., & Levkovitz, Y. (2009). Neuropsychological evidence of impaired cognitive empathy in euthymic bipolar disorder. *The Journal of Neuropsychiatry and Clinical Neurosciences, 21*(1), 59-67. doi:10.1176/jnp.2009.21.1.59

Simon, M., Herold, R., Hajnal, A., Varga, E., Fekete, S., & Tenyi, T. (2013). Theory of mind deficit in bipolar patients with subsyndromal illness. *European Neuropsychopharmacology*(23), 295-296.

Skuse, D. H., & Gallagher, L. (2011). Genetic influences on social cognition. *Pediatric Research, 69*(8), 85-91. doi:10.1203/PDR.0b013e318212f562

Smeets, T., Dziobek, I., & Wolf, O. T. (2009). Social cognition under stress: Differential effects of stress-induced cortisol elevations in healthy young men and women. *Hormones and Behavior, 55*(4), 507-513. https://doi.org/10.1016/j.yhbeh.2009.01.011

Smith, E. M., Reynolds, S., Orchard, F., Whalley, H. C., & Chan, S. W. Y. (2018). Cognitive biases predict symptoms of depression, anxiety and wellbeing above and beyond neuroticism in adolescence. *Journal of Affective Disorders, 241*, 446-453. https://doi.org/10.1016/j.jad.2018.08.051

Sodian, B. (2011). Theory of mind in infancy. *Child Development Perspectives, 5*(1), 39-43. https://doi.org/10.1111/j.1750-8606.2010.00152.x

Soeiro-de-Souza, M. G., Bio, D. S., David, D. P., Rodrigues dos Santos, D., Kerr, D. S., Gattaz, W. F., . . . Moreno, R. A. (2012). COMT Met (158) modulates facial emotion recognition in bipolar I disorder mood episodes. *Journal of Affective Disorders, 136*(3), 370-376. https://doi.org/10.1016/j.jad.2011.11.021

Soeiro-de-Souza, M. G., Otaduy, M. C. G., Dias, C. Z., Bio, D. S., Machado-Vieira, R., & Moreno, R. A. (2012). The impact of the CACNA1C risk allele on limbic structures and facial emotions recognition in bipolar disorder subjects and healthy controls. *Journal of Affective Disorders, 141*(1), 94-101. https://doi.org/10.1016/j.jad.2012.03.014

Sommer, G., & Fydrich, T. (1991). Entwicklung und Überprüfung eines Fragebogens zur sozialen Unterstützung (F-SOZU). *Diagnostica, 37*(2), 160–178.

Sprengelmeyer, R., Steele, J. D., Mwangi, B., Kumar, P., Christmas, D., Milders, M., & Matthews, K. (2011). The insular cortex and the neuroanatomy of major depression. *Journal of Affective Disorders, 133*(1), 120-127. https://doi.org/10.1016/j.jad.2011.04.004

Stephan, A. (2013). Enactive emotion and impaired agency in depression. *Journal of Consciousness Studies, 20*(7-8), 33-55.

Stoddart, S. D. R., Craddock, N. J., & Jones, L. A. (2007). Differentiation of executive and attention impairments in affective illness. *Psychological Medicine, 37*(11), 1613-1623. doi:10.1017/S0033291707000712

Stone, V. E., Baron-Cohen, S., & Knight, R. T. (1998). Frontal lobe contributions to theory of mind. *Journal of Cognitive Neuroscience, 10*(5), 640-656.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(6), 643-662. https://doi.org/10.1037/h0054651

Strunk, D. R., Lopez, H., & DeRubeis, R. J. (2006). Depressive symptoms are associated with unrealistic negative predictions of future life events. *Behaviour Research and Therapy, 44*(6), 861-882. https://doi.org/10.1016/j.brat.2005.07.001

Sullivan, S., Ruffman, T., & Hutton, S. B. (2007). Age differences in emotion recognition skills and the visual scanning of emotion faces. *The Journals of Gerontology: Series B, 62*(1), 53-60. doi:10.1093/geronb/62.1.P53

Summers, M., Papadopoulou, K., Bruno, S., Cipolotti, L., & Ron, M. A. (2006). Bipolar I and bipolar II disorder: cognition and emotion processing. *Psychological Medicine, 36*(12), 1799-1809. doi:10.1017/S0033291706008804

Surcinelli, P., Codispoti, M., Montebarocci, O., Rossi, N., & Baldaro, B. (2006). Facial emotion recognition in trait anxiety. *Journal of Anxiety Disorders, 20*(1), 110-117. https://doi.org/10.1016/j.janxdis.2004.11.010

Sweeney, P. D., Anderson, K., & Bailey, S. (1986). Attributional style in depression: a meta-analytic review. *Journal of Personality and Social Psychology, 50*(5), 974-991. doi:10.1037/0022-3514.50.5.974

Szanto, K., Dombrovski, A. Y., Sahakian, B. J., Mulsant, B. H., Houck, P. R., Reynolds, C. F., & Clark, L. (2012). Social emotion recognition, social functioning, and attempted suicide in late-life depression. *The American Journal of Geriatric Psychiatry, 20*(3), 257-265. https://doi.org/10.1097/JGP.0b013e31820eea0c

Szily, E., & Kéri, S. (2009). Anomalous subjective experience and psychosis risk in young depressed patients. *Psychopathology, 42*(4), 229-235. doi:10.1159/000218520

Takahashi, H., Yahata, N., Koeda, M., Matsuda, T., Asai, K., & Okubo, Y. (2004). Brain activation associated with evaluative processes of guilt and embarrassment: an fMRI study. *NeuroImage, 23*(3), 967-974. https://doi.org/10.1016/j.neuroimage.2004.07.054

Tapajóz Pereira de Sampaio, F., Soneira, S., Aulicino, A., & Allegri, R. F. (2013). Theory of mind in eating disorders and their relationship to clinical profile. *European Eating Disorders Review, 21*(6), 479-487. https://doi.org/10.1002/erv.2247

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin, 103*(2), 193-210. doi:10.1037/0033-2909.103.2.193

Teng, E. L. (1990). The 3RT Test: Three reaction time tasks for IBM PC computers. *Behavior Research Methods, Instruments, & Computers, 22*(4), 389-392. https://doi.org/10.3758/BF03203180

Thaler, N. S., Allen, D. N., Sutton, G. P., Vertinski, M., & Ringdahl, E. N. (2013). Differential impairment of social cognition factors in bipolar disorder with and without psychotic features and schizophrenia. *Journal of Psychiatric Research, 47*(12), 2004-2010. https://doi.org/10.1016/j.jpsychires.2013.09.010

Thayer, J., & Johnsen, B. H. (2000). Sex differences in judgement of facial affect: A multivariate analysis of recognition errors. *Scandinavian Journal of Psychology, 41*(3), 243-246. doi:10.1111/1467-9450.00193

Thoma, P., Schmidt, T., Juckel, G., Norra, C., & Suchan, B. (2015). Nice or effective? Social problem solving strategies in patients with major depressive disorder. *Psychiatry Research, 228*(3), 835-842. https://doi.org/10.1016/j.psychres.2015.05.015

Tombaugh, T. N. (2004). Trail Making Test A and B: normative data stratified by age and education. *Archives of Clinical Neuropsychology, 19*(2), 203-214. doi:10.1016/S0887-6177(03)00039-8

Torrent, C., Martínez-Arán, A., Daban, C., Sánchez-Moreno, J., Comes, M., & Goikolea, J. M. (2006). Cognitive impairment in bipolar II disorder. *The British Journal of Psychiatry, 189*(3), 254-259. doi:10.1192/bjp.bp.105.017269

Torres, I., Boudreau, V., & Yatham, L. (2007). Neuropsychological functioning in euthymic bipolar disorder: a meta-analysis. *Acta Psychiatrica Scandinavica, 116*, 17-26.

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., . . . Nelson, C. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research, 168*(3), 242-249. https://doi.org/10.1016/j.psychres.2008.05.006

Treuer, T., & Tohen, M. (2010). Predicting the course and outcome of bipolar disorder: a review. *European Psychiatry, 25*(6), 328-333. doi:10.1016/j.eurpsy.2009.11.012

Turkstra, L. S., Mutlu, B., Ryan, C. W., Despins Stafslien, E. H., Richmond, E. K., Hosokawa, E., & Duff, M. C. (2020). Sex and gender differences in emotion recognition and theory of mind after

TBI: a narrative review and directions for future research. *Frontiers in Neurology, 11*(59). doi:10.3389/fneur.2020.00059

Uekermann, J., Channon, S., LehmkÄMper, C., Abdel-Hamid, M., Vollmoeller, W., & Daum, I. (2008). Executive function, mentalizing and humor in major depression. *Journal of the International Neuropsychological Society, 14*(1), 55-62. doi:10.1017/S1355617708080016

Ullman, T. D., & Tenenbaum, J. B. (2020). Bayesian models of conceptual development: learning as building models of the world. *Annual Review of Developmental Psychology, 2*(1), 533-558. doi:10.1146/annurev-devpsych-121318-084833

Vakil, E., & Blachstein, H. (1993). Rey Auditory-Verbal Learning Test: Structure analysis. *Journal of Clinical Psychology, 49*(6), 883-890. https://doi.org/10.1002/1097-4679(199311)49:6<883::AID-JCLP2270490616>3.0.CO;2-6

van der Schalk, J., Hawk, S. T., Fischer, A. H., & Doosje, B. (2011). Moving faces, looking places: validation of the Amsterdam Dynamic Facial Expression Set (ADFES). *Emotion, 11*(4), 907-920. https://doi.org/10.1037/a0023853

Van Haren, N. E. M., Van Dam, D. S., & Stellato, R. K. (2019). Change in IQ in schizophrenia patients and their siblings: a controlled longitudinal study. *Psychological Medicine, 49*(15), 2573-2581.

Van Neerven, T., Bos, D. J., & Van Haren, N. E. M. (2021). Deficiencies in Theory of Mind in patients with schizophrenia, bipolar disorder, and major depressive disorder: A systematic review of secondary literature. *Neuroscience & Biobehavioral Reviews, 120*, 249-261. doi:10.1016/j.neubiorev.2020.11.011

Van Rheenen, T. E., & Rossell, S. L. (2013). Picture sequencing task performance indicates theory of mind deficit in bipolar disorder. *Journal of Affective Disorders, 151*(3), 1132-1134. https://doi.org/10.1016/j.jad.2013.07.009

Vaskinn, A., Sundet, K., Friis, S., Simonsen, C., Birkenæs, A. B., Engh, J. A., . . . Andreassen, O. A. (2007). The effect of gender on emotion perception in schizophrenia and bipolar disorder. *Acta Psychiatrica Scandinavica, 116*(4), 263-270. doi:10.1111/j.1600-0447.2007.00991.x

Västfjäll, D. (2001). Emotion induction through music: A review of the musical mood induction procedure. *Musicae Scientiae, 5*(S1), 173-211. doi:10.1177/10298649020050s107

Vázquez, G. H., Holtzman, J. N., Lolich, M., Ketter, T. A., & Baldessarini, R. J. (2015). Recurrence rates in bipolar disorder: Systematic comparison of long-term prospective, naturalistic studies versus randomized controlled trials. *European Neuropsychopharmacology, 25*(10), 1501-1512. https://doi.org/10.1016/j.euroneuro.2015.07.013

Vederman, A. C., Weisenbach, S. L., Rapport, L. J., Leon, H. M., Haase, B. D., Franti, L. M., . . . Zubieta, J.-K. (2012). Modality-specific alterations in the perception of emotional stimuli in bipolar disorder compared to healthy controls and major depressive disorder. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior, 48*(8), 1027-1034. https://doi.org/10.1016/j.cortex.2011.03.017

Velthorst, E., Fett, A.-K. J., Reichenberg, A., Perlman, G., van Os, J., Bromet, E. J., & Kotov, R. (2017). The 20-year longitudinal trajectories of social functioning in individuals with psychotic disorders. *American Journal of Psychiatry, 174*(11), 1075-1085. doi:10.1176/appi.ajp.2016.15111419

Venn, H. R., Gray, J. M., Montagne, B., Murray, L. K., Michael Burt, D., Frigerio, E., . . . Young, A. H. (2004). Perception of facial expressions of emotion in bipolar disorder. *Bipolar Disorders, 6*(4), 286-293. doi:10.1111/j.1399-5618.2004.00121.x

Volkert, J., Schiele, M. A., Kazmaier, J., Glaser, F., Zierhut, K. C., Kopf, J., . . . Reif, A. (2016). Cognitive deficits in bipolar disorder: from acute episode to remission. *European Archives of Psychiatry and Clinical Neuroscience, 266*(3), 225-237. doi:10.1007/s00406-015-0657-2

Vreeker, A., Boks, M. P. M., Abramovic, L., Verkooijen, S., van Bergen, A. H., Hillegers, M. H. J., . . . Riemersma-Van der Lek, R. F. (2016). High educational performance is a distinctive feature of bipolar disorder: a study on cognition in bipolar disorder, schizophrenia patients, relatives and controls. *Psychological Medicine, 46*(4), 807-818. https://doi.org/10.1017/S0033291715002299

Wang, Y.-g., Wang, Y.-q., Chen, S.-l., Zhu, C.-y., & Wang, K. (2008). Theory of mind disability in major depression with or without psychotic symptoms: A componential view. *Psychiatry Research, 161*(2), 153-161. https://doi.org/10.1016/j.psychres.2007.07.018

Warshaw, M. G., Dyck, I., Allsworth, J., Stout, R. L., & Keller, M. B. (2001). Maintaining reliability in a long-term psychiatric study: an ongoing inter-rater reliability monitoring program using the longitudinal interval follow-up evaluation. *Journal of Psychiatric Research, 35*(5), 297-305. https://doi.org/10.1016/S0022-3956(01)00030-9

Warshaw, M. G., Keller, M. B., & Stout, R. L. (1994). Reliability and validity of the longitudinal interval follow-up evaluation for assessing outcome of anxiety disorders. *Journal of Psychiatric Research, 28*(6), 531-545. https://doi.org/10.1016/0022-3956(94)90043-4

Wechsler, D. (1999). *Wechsler abbreviated scale of intelligence (WASI)* [Database record]. APA PsycTests. https://doi.org/10.1037/t15170-000

Wegbreit, E., Weissman, A. B., Cushman, G. K., Puzia, M. E., Kim, K. L., Leibenluft, E., & Dickstein, D. P. (2015). Facial emotion recognition in childhood-onset bipolar I disorder: an evaluation of developmental differences between youths and adults. *Bipolar Disorders, 17*(5), 471-485. doi:10.1111/bdi.12312

Weightman, M. J., Air, T. M., & Baune, B. T. (2014). A Review of the Role of Social Cognition in Major Depressive Disorder. *Frontiers in Psychiatry, 5*(179). doi:10.3389/fpsyt.2014.00179

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. doi: 10.1111/1467-8624.00304

West, J. T., Horning, S. M., Klebe, K. J., Foster, S. M., Cornwell, R. E., Perrett, D., . . . Davis, H. P. (2012). Age effects on emotion recognition in facial displays: from 20 to 89 years of age. *Experimental Aging Research, 38*(2), 146-168. doi:10.1080/0361073X.2012.659997

Westby, C. E. (2014). Social neuroscience and theory of mind. *Folia Phoniatrica et Logopaedica, 66*(1-2), 7-17.

Weyers, P., Mühlberger, A., Hefele, C., & Pauli, P. (2006). Electromyographic responses to static and dynamic avatar emotional facial expressions. *Psychophysiology, 43*(5), 450-453. doi:10.1111/j.1469-8986.2006.00451.x

Wiener, D., Andrzejewska, M., Bodnar, A., & Rybakowski, J. (2011). Disturbances of the theory of mind and empathy in schizophrenia and bipolar affective illness. *Neuropsychiatria i Neuropsychologia, 6*(2), 85.

Wilbertz, G., Brakemeier, E.-L., Zobel, I., Härter, M., & Schramm, E. (2010). Exploring preoperational features in chronic depression. *Journal of Affective Disorders, 124*(3), 262-269. https://doi.org/10.1016/j.jad.2009.11.021

Williams, J. M. G., Teasdale, J. D., Segal, Z. V., & Soulsby, J. (2000). Mindfulness-based cognitive therapy reduces overgeneral autobiographical memory in formerly depressed patients. *Journal of Abnormal Psychology, 109*(1), 150-155. doi: 10.1037//0021-843x.109.1.150

Williams, L. M., Mathersul, D., Palmer, D. M., Gur, R. C., Gur, R. E., & Gordon, E. (2009). Explicit identification and implicit recognition of facial emotions: I. Age effects in males and females across 10 decades. *Journal of Clinical and Experimental Neuropsychology, 31*(3), 257-277. doi:10.1080/13803390802255635

Wirtz, D., Kruger, J., Scollon, C. N., & Diener, E. (2003). What to do on spring break? Predicting future choice based on online versus recalled affect. *Psychological Science, 14*(5), 520-524. https://doi.org/10.1111/1467-9280.03455

Wittchen, H.-U., Zaudig, M., & Fydrich, T. (1997). *SKID. Strukturiertes klinisches Interview für DSM-IV. Achse I und II. Handanweisung.* Göttingen: Hogrefe.

Wolf, F., Brüne, M., & Assion, H. J. (2010). Theory of mind and neurocognitive functioning in patients with bipolar disorder. *Bipolar Disorders, 12*(6), 657-666. doi:10.1111/j.1399-5618.2010.00854.x

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

Yamada, K., Inoue, Y., & Kanba, S. (2015). Theory of mind ability predicts prognosis of outpatients with major depressive disorder. *Psychiatry Research, 230*(2), 604-608. https://doi.org/10.1016/j.psychres.2015.10.011

Yoshikawa, S., & Sato, W. (2006). Enhanced perceptual, emotional, and motor processing in response to dynamic facial expressions of emotion. *Japanese Psychological Research, 48*(3), 213-222.

Young, A. W., Perrett, D., Calder, A., Sprengelmeyer, R., & Ekman, P. (2002). *Facial expressions of emotion: Stimuli and tests (FEEST).* Bury St. Edmunds: Thames Valley Test Company.

Young, A. W., Rowland, D., Calder, A. J., Etcoff, N. L., Seth, A., & Perrett, D. I. (1997). Facial expression megamix: Tests of dimensional and category accounts of emotion recognition. *Cognition, 63*(3), 271-313. https://doi.org/10.1016/S0010-0277(97)00003-6

Young, J. E., Klosko, J. S., Weishaar, M. E., & Kierdorf, T. (2005). *Schematherapie. Ein praxisorientiertes Handbuch.* Paderborn: Junfermann.

Yurgelun-Todd, D. A., Gruber, S. A., Kanayama, G., Killgore, W. D., Baird, A. A., & Young, A. D. (2000). fMRI during affect discrimination in bipolar affective disorder. *Bipolar Disorders, 2*(3p2), 237-248. https://doi.org/10.1034/j.1399-5618.2000.20304.x

Zahn, R., Lythe, K. E., Gethin, J. A., Green, S., Deakin, J. F. W., Young, A. H., & Moll, J. (2015). The role of self-blame and worthlessness in the psychopathology of major depressive disorder. *Journal of Affective Disorders, 186*, 337-341. https://doi.org/10.1016/j.jad.2015.08.001

Zobel, I., Werden, D., Linster, H., Dykierek, P., Drieling, T., Berger, M., & Schramm, E. (2010). Theory of mind deficits in chronically depressed patients. *Depression and Anxiety, 27*(9), 821-828. https://doi.org/10.1002/da.20713

Zwick, J. C., & Wolkenstein, L. (2017). Facial emotion recognition, theory of mind and the role of facial mimicry in depression. *Journal of Affective Disorders, 210*, 90-99. https://doi.org/10.1016/j.jad.2016.12.022