Review Article

TRAUMA, VIOLENCE, & ABUSE
1-19

DOI: 10.1177/15248380211013143
journals.sagepub.com/home/tva


Theory of Mind in Offending: A Systematic Review

Nilda Karoğlu[1], Heather J. Ferguson[1], and Caoilte Ó Ciardha[1]

Abstract

Theory of mind (ToM) impairment is associated with poor social functioning in some psychological disorders (e.g., autism and schizophrenia). ToM deficits have also been linked with offending behavior in the theoretical literature. However, no review has examined the empirical evidence for such a link. We carried out a systematic review to provide a critical overview of studies involving ToM ability in offenders. We included studies published in English that used an instrument to measure at least one aspect of ToM. Twenty-eight eligible studies were identified and coded. Our findings reveal a generally mixed literature. Taking study quality into account, our findings suggest that offenders and nonoffenders do not differ in their first-order ToM. For second-order ToM, findings are mixed, even when only the highest quality studies are examined. Studies exploring advanced ToM showed mixed results overall, though the highest quality research appeared to indicate that offenders have impairments in advanced ToM which means that they may have difficulty understanding various mental states such as pretense, white lies, irony, double bluffs, and sarcasm. We suggest that well-controlled future studies, which also measure other facets of ToM (e.g., distinguishing between cognitive and affective ToM or examining ToM content), are needed to fully understand the role of ToM in offending.

Keywords

theory of mind, theory of nasty mind, mind reading, offending, offender cognition

Theory of mind (ToM) is a term used to describe complex cognitive processes (Duval et al., 2011) that allow humans to understand their own mental states and those of others (Klin, 2000). This phenomenon appears cognate with the terms *mind-reading*, *mentalizing*, and *cognitive empathy*. ToM is argued to be a subcomponent of social cognition (Bellerose et al., 2011) and is seen as essential for interpreting people's behavior, regulating behavior, and interacting socially (Beer & Ochsner, 2006). ToM is a multidimensional construct that has cognitive and affective components (Brothers & Ring, 1992). While cognitive ToM refers to the ability to understand other people's intentions, beliefs, and knowledge, affective ToM refers to the ability to infer other people's emotions (Shamay-Tsoory et al., 2010).

ToM ability is often operationalized in terms of first-order ToM, second-order ToM, and advanced ToM. Success in first-order false belief tasks requires the ability to understand that another person is holding an incorrect belief (Shamay-Tsoory et al., 2005). Success in second-order false belief tasks entails the ability to comprehend that a person holds a false belief about another person's belief. Advanced ToM tasks involve insights into mental states such as jokes, sarcasm, double bluffs, and faux pas. Children typically show implicit awareness of others' perspectives from around 18 months old (e.g., Buttelmann et al., 2009; Kovács et al., 2010; Onishi & Baillargeon, 2005; Rubio-Fernández & Geurts, 2013; Senju et al., 2009) and are thought to develop the necessary skills to pass false belief ToM tasks between the ages of 2 and 7 years old (Wellman et al., 2001). Moreover, empirical findings suggest that ToM performance is affected by sociodemographic variables such as age (Brunsdon et al., 2019; Ferguson et al., 2018), socioeconomic status, and education (Li et al., 2013; Shatz et al., 2003), as well as individual difference variables such as intelligence (Charlton et al., 2009), and executive functioning (Cane et al., 2017; Devine & Hughes, 2014).

Absent or impaired functioning of ToM is thought to be associated with psychosocial difficulties in various types of psychopathology, both in children and adults (Brüne & Brüne-Cohrs, 2006), including, but not limited to, schizophrenia (Frith, 1992), autism (Baron-Cohen, 1995; Bradford et al., 2018), bipolar affective disorder (Kerr et al., 2003), and antisocial personality disorder (Richell et al., 2003). Studies also appear to show that individuals exhibiting violent, antisocial, and delinquent behavior have deficits in ToM (Abu-Akel & Abushua'leh, 2004; Fonagy & Levinson, 2004). The

[1] School of Psychology, University of Kent, Canterbury, United Kingdom

Corresponding Author:
Caoilte Ó Ciardha, School of Psychology, University of Kent, Canterbury CT2 7NP, United Kingdom.
Email: c.c.ociardha@kent.ac.uk

primary aim of this review is to consolidate, synthesize, and critically evaluate existing research on the ToM-offending link. We aim to establish whether there is sufficient evidence to substantiate this link and highlight areas for future research.

While there is no comprehensive theory that models the relationship between ToM and offending in general, theory relating to sexual offending provides a starting point for how such a model might work. Ward et al. (2000; see also Keenan & Ward, 2000) proposed that sexual offending is linked with ToM deficits. According to their model, individuals who commit sexual offenses may have failed to develop an adequate ToM, and this failure may lead those individuals to view or process information about their own or other people's mental states in a biased or distorted way. Alternatively, these individuals may have a ToM impairment specific to particular kinds of mental states in certain relationships—for example, having a theory that is underpinned by false assumptions about women or children. In a similar vein, Elsegood and Duff (2010) suggested that ToM impairment might contribute to offending by underpinning criminogenic needs, such as intimacy deficits. Since criminogenic needs are the focus of offender treatment programs (Andrews & Bonta, 2010; Serin et al., 2009), it is important for practitioners and policy makers to know whether or not deficits in ToM represent a criminogenic need that should be targeted in treatment.

Empathy is a multidimensional term that describes the affective/emotional response to another's mental state (e.g., Stotland, 1969) or the cognitive mechanisms that enable people to understand others' perspectives (Dymond, 1949). This cognitive conceptualization of empathy therefore overlaps with ToM (Ferguson et al., 2015; Wang & Wang, 2015) and has been studied widely in forensic populations (Jolliffe & Farrington, 2004; van Langen et al., 2014). Importantly, empathy has been a key component of intervention programs (Laws & Ward, 2011) for offenders who committed serious crimes, such as sexual and violent offending (Day et al., 2010). However, targeting empathy in interventions is controversial because evidence for the impact of treating empathy on later recidivism is mixed (Brown et al., 2012; Hanson & Morton-Bourgon, 2004, 2005). One factor that might have led to inconsistent results is that most treatment programs focus on generalized empathy deficit training and overlook the cognitive and affective components of empathy (Brown et al., 2012). Considering that cognitive empathy is closely related to ToM and that the two terms are even used interchangeably by some researchers, treatments targeting general empathy might actually be targeting aspects of ToM (or potentially missing important aspects of ToM). Therefore, it is important to fully understand any relationship between offending and ToM in order to inform clinical decision making and to underpin interventions for offending populations. To date, there have been no adequate reviews of the current state of the literature on ToM and offending.

Method

This systematic review was conducted in line with Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA; Moher et al., 2009). PRISMA is an empirical reporting process which uses a 27-item checklist and a four-phase flow diagram to organize the process of identifying relevant studies for systematic analysis.

Eligibility Criteria

Research articles published in English that included one or more instruments that measure at least one aspect of ToM (e.g., first-order ToM, affective ToM) in offenders were eligible. Studies had to compare the ToM of a minimum of two groups, including at least one offending group and a nonoffending control group. We excluded articles based on the following criteria: (1) articles that were not published in English; (2) articles not measuring an aspect of ToM; (3) articles solely measuring affective empathy (see Eisenberg et al., 2006); (4) articles measuring ToM with a basic facial emotion recognition task;[1] (5) studies assessing ToM with interview methods (which therefore measure people's evaluation of their own ToM rather than providing an objective measure of ToM; see Discussion section); (6) studies measuring ToM with questionnaires where participants rate their own ToM (since responses to the questionnaires may not reflect participants' true ToM abilities; see Discussion section); (7) articles involving case reports, literature reviews, book reviews, and commentaries; and (8) studies with fewer than 14 participants per group in a core analysis of interest (this reflects the minimum group size for a one-tailed t test with 80% power to detect a difference with a very large effect size [Cohen's d of 1]).

Search Strategy and Screening

There was no restriction on the year of publication of the study. Searches were conducted of relevant databases: PsycINFO, PsycARTICLES, Science Direct, Scopus, Criminal Justice Abstracts (from EBSCO), Open Access Theses and Dissertations, EBSCO host, and ProQuest Dissertations & Theses Global. Additionally, targeted searches in Google and in the references of identified studies were performed. The main search terms were "theory of mind," "mentalizing," "mentalising," "mentalization," "mentalisation," "mindreading, " "mind reading," "mind-reading," "mind perception," "social intelligence," "cognitive empathy," "false belief reasoning," "metacognition," and "social cognition," and were systematically paired with each of the following key words: "incarcerated," "offenders," "criminals," "offending," "prisoners," "inmates," "convicts," and "forensic." Subsequently, after removing duplicates, titles and abstracts of the remaining articles were evaluated to determine whether they were eligible for this review. Furthermore, we contacted all corresponding authors of eligible papers for whom we could find current email addresses to request

unpublished manuscripts that could be included in the systematic review.

Quality Assessment

After we identified eligible studies, we performed a quality assessment by adapting the *quality appraisal checklist-quantitative studies reporting correlations and associations checklist* developed by the National Institute for Health and Care Excellence (2012). We used this checklist as a template but modified it by excluding items that were not directly relevant to the assessment of quality in ToM studies. New items were added to the checklist to examine factors that might influence the outcome and quality of ToM studies. We assert that good-quality studies of ToM and offending should (1) use tasks that are age appropriate, valid, reliable, and ecologically valid; (2) have a control group that consists of neurotypical nonoffenders; (3) include a control task or control items such as non-ToM questions, attention, or memory check questions; and (4) control for potential confounding factors such as executive functioning (Devine & Hughes, 2014), age, intelligence (Charlton et al., 2009), socioeconomic status, and education (Li et al., 2013; Shatz, et al., 2003). For the full checklist, see Online Supplementary Material. Two researchers scored four studies on each quality aspect, obtaining excellent interrater agreement (ICC = 0.99). After scoring, we categorized studies as being either high, moderate, or low quality.

Results

Description of the Included Studies

A total of 6,294 articles were identified: 4,277 from Scopus, 1,515 from PsycINFO, 158 from PubMed, 162 from PsyARTICLES, 89 from Criminal Justice, 35 from Open Access Theses and Dissertations, 31 from EBSCO host, 26 from ProQuest Dissertations & Theses Global, and one from the references of identified studies. After duplicates were removed, 2,982 remaining studies were carefully checked and assessed to determine their eligibility for the review. From these, 2,889 were excluded (233 were non-English articles; 1,095 were literature reviews, meta-analyses, interview studies, case reports, conference presentations, or commentaries; and 1,561 did not measure ToM; see Figure 1).

From the remaining 93 studies, a final sample of 28 studies spanning 16 years (2004–2019) was included in this review. Table 1 summarizes the characteristics of each study.

Description of Study Characteristics

All studies were cross-sectional. Ten studies were conducted in the United Kingdom. The remaining studies took place in one of the following countries: Canada ($N = 1$), Germany ($N = 6$), Hungary ($N = 1$), Israel ($N = 1$), Italy ($N = 2$), the Netherlands ($N = 3$), Norway ($N = 1$), Serbia ($N = 1$), Spain ($N = 1$), and Switzerland ($N = 1$). Data were gathered from a total of 2,845 participants (1,431 offenders; 99% male). Among the reviewed studies, 21 different tasks[2] were used to measure ToM, examining one or more of the following ToM levels: first-order ToM, second-order ToM, and advanced ToM. Six of the 28 studies measured multiple levels of ToM, whereas the remaining studies assessed a single level. We will report the results of each study, and briefly describe each task, in the category that corresponds to the ToM level.

Summary of First-Order ToM Methods

First-order ToM tasks assess whether people can infer another person's thoughts, feelings, beliefs, and intentions accurately. In this review, nine studies measured first-order ToM, with seven of them focusing on cognitive first-order ToM. Five of the studies (Castellino et al., 2011; Dolan & Fullam, 2004; Hammond & Beail, 2017; Majorek et al., 2009, Proctor & Beail, 2007) utilized different tasks such as the Smarties task (Hogrefe et al., 1986), the Sally–Anne test (Baron-Cohen et al., 1985; Wimmer & Perner, 1983), the Marble Story Task, and the Picture Sequencing Task (Brüne, 2003), each of which measured participants' ability to understand another person's false-belief about the content of an item or the location of an object. The latter task also included picture sequencings that assessed participants' ability to understand the intention (cooperative and deceptive) of cartoon characters (an example of the task can be found in Brüne, 2003). A study by Jones et al. (2007) measured ToM with an animation task (Abell et al., 2000) where participants were required to attribute mental states to triangles based on their interactions. Engelstad et al. (2019) used the Hinting Task (Corcoran et al., 1995) where participants were asked to explain the intention or the message in the protagonist's statement to another character. Shamay-Tsoory et al.'s (2010) study which measured both cognitive and affective first-order ToM used the Yoni task where participants had to infer the mental state of a cartoon character Yoni based on eye gaze (examples of the task can be found in Shamay-Tsoory et al. 2010). Robinson et al. (2007) measured affective first-order ToM using the empathy continuum (Strayer, 1993) where participants were required to infer the mental state of a protagonist presented in video sketches and express it in an interview.

First-Order ToM Results

Two of the nine first-order ToM studies found that offenders performed significantly worse than nonoffenders (Majorek et al., 2009; Robinson et al., 2007).[3] The remaining studies did not find performance differences in first-order ToM between offenders and nonoffenders (Castellino et al., 2011; Dolan & Fullam, 2004; Engelstad et al., 2019; Hammond & Beail, 2017; Jones et al., 2007; Proctor & Beail, 2007; Shamay-Tsoory et al., 2010).

Our quality assessment indicated that eight of the first-order studies were of low quality, with one moderate quality study. Dolan and Fullam's (2004) study, which found no first-order ToM deficits in offenders, had the highest quality rating among the reviewed first-order ToM studies because it controlled for

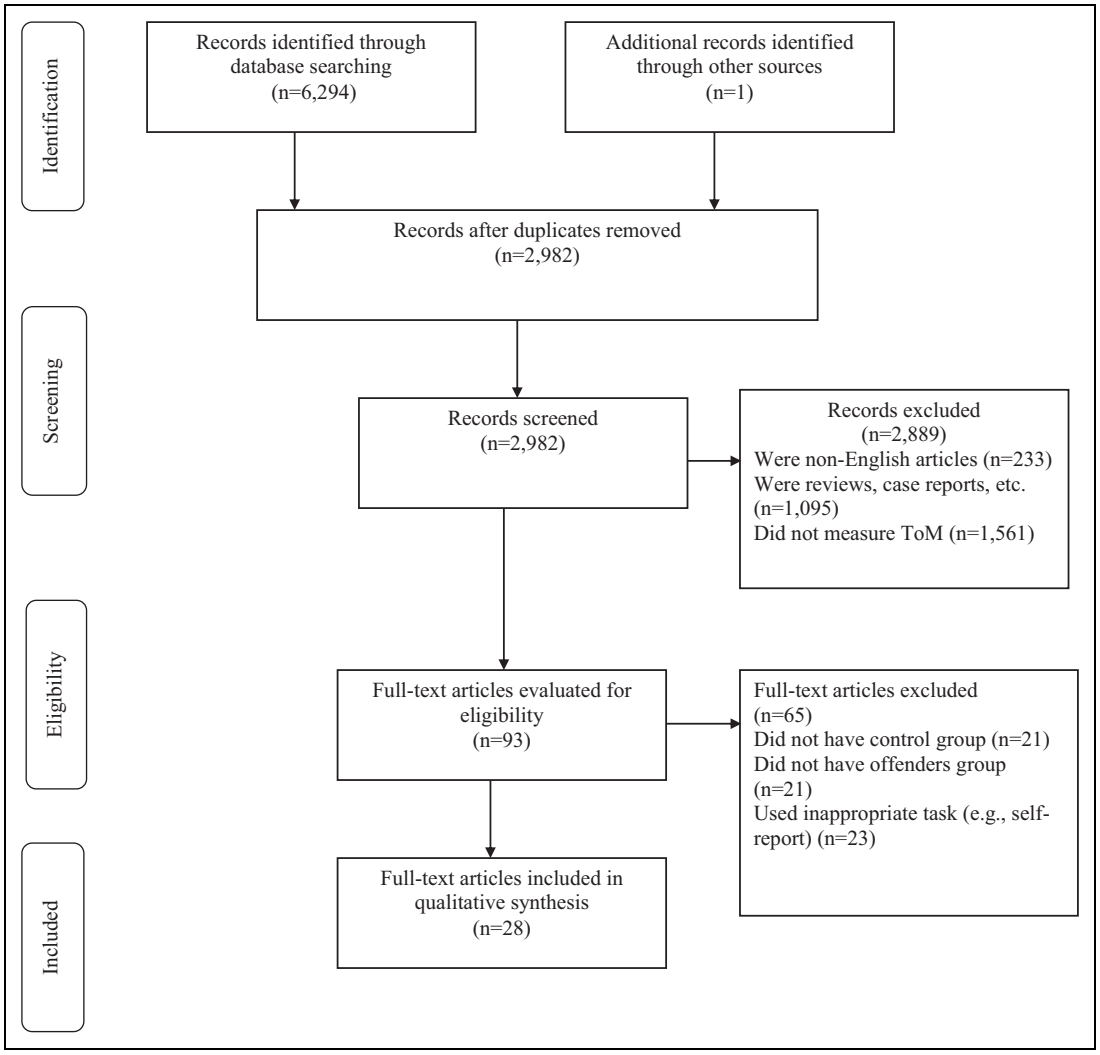

Figure 1. Flowchart of literature review.

several potential confounding variables, reduced memory load, and accounted for comprehension. Nevertheless, results from Dolan and Fullam's (2004) study should be interpreted with caution, because almost all participants passed the first-order ToM task, and the nonsignificant differences might have been caused by a ceiling effect as a result of using a task (the Sally–Anne test) that was too simple for an adult population. We identified that other studies using the same or similar tasks that are easy for adults (i.e., the Smarties, cartoon, or animation tasks) also found nonsignificant first-order ToM differences between offenders and nonoffenders. However, the studies which used tasks that were more appropriate for adults (e.g., video stimuli) found first-order ToM deficits in offenders with a medium to large effect size (Majorek et al., 2009; Robinson et al., 2007).

Summary of Second-Order ToM Methods

Second-order ToM tasks examine whether people can accurately understand a person's mental state about another person's mental state (what X thinks about Y's thoughts, feelings, intentions, or beliefs). Second-order ToM was examined by five studies and used tasks to evaluate cognitive second-order ToM (Castellino et al., 2011; Dolan & Fullam, 2004; Majorek et al., 2009; Proctor & Beail, 2007). Specifically, three studies used the Ice Cream Van story (Perner & Wimmer, 1985) alone, or together with the Burglar Story (Happé et al., 1999), where the protagonists in both tasks held false-beliefs about the thoughts and beliefs of another person. Majorek et al. (2009) used the Picture Sequencing Task (Brüne, 2003), where participants were required to understand the beliefs or thoughts of a character about another character's intentions or thoughts. A similar task, the Yoni Task, was used by Shamay-Tsoory et al. (2010), but this task evaluated both cognitive and affective second-order ToM.

Second-Order ToM Results

Results on second-order ToM tasks were particularly inconsistent between studies. While Dolan and Fullam (2004) found no

Table 1. Details of Studies Included in the Review.

Authors (Year; Country)	Sample	Mean Age	Measure of ToM	Variables Controlled in Design	Summary of Crime Types	Key Findings	Quality Rating
Castellino et al. (2011; Italy)	• 21 Os • 21 HNOs	44.3 44.4	• Smarties (cigarettes) test • Sally–Anne task • Burglar Story • Ice Cream Van story • Strange stories	Age, education, IQ, socioeconomic status, comprehension, memory, and attention	Sexual offenses	No significant first-order ToM differences. Offenders obtained significantly lower scores in second-order and advanced ToM than nonoffenders	First order = 11.5/low Second order = 12.5/moderate Advanced = 14.5/moderate
de Jong et al. (2018; the Netherlands)	• 23 Os with psychosis • 27 NOs with psychosis • 33 HNOs	37.3 35.4 38.6	• Faux pas	Age and education	Violent offenses	Offenders had lower ToM scores than the NO with psychosis, and both groups obtained lower scores than HNO. However, correlations between the group membership and the discriminant function revealed that two variables other than ToM loaded onto the first function	Advanced = 10/low
Dolan & Fullam (2004; United Kingdom)	• 89 ASPD Os with and without psychopathy • 20 HNOs	36.7 33.6	• Sally–Anne task • Ice Cream Van story • RMET • Faux pas	Age, education, IQ, memory, and attention	Violent offenses and unspecified offenses	No first-order or second-order ToM deficit in offenders. Advanced ToM: No significant group differences at detecting faux pas. However, offenders were worse than nonoffenders at understanding the mental state of the person who committed a faux pas and of the person to whom the faux pas was made Offenders without psychopathy performed worse in recognizing basic emotions than the offenders with psychopathy and nonoffenders in the RMET. There were no significant group differences in recognizing complex mental states in the RMET	First order = 12.5/moderate Second order = 12.5/moderate Advanced = 14/moderate

(continued)

Table 1. (continued)

Authors (Year; Country)	Sample	Mean Age	Measure of ToM	Variables Controlled in Design	Summary of Crime Types	Key Findings	Quality Rating
Domes et al. (2013; Germany)	• 90 Os with psychopathy • 28 HNOs	48.5 52.5	• RMET-R • MET	Age, education, and IQ	Homicide, sexual offenses, and violent offenses	No significant differences between offenders and nonoffenders in the RMET-R Offenders had significantly lower scores than nonoffenders in the MET	Advanced = 13.5/ moderate
Elsegood & Duff (2010; United Kingdom)	• 46 Os • 46 HNOs	44.9 43.2	• RMET-R • MCET	Age, socioeconomic status, and verbal knowledge	Sexual offenses against children	Offenders scored significantly lower than nonoffenders in understanding mental states of adults. No significant differences between the scores of offenders and nonoffenders in understanding mental states of children	Advanced = 15.5/ moderate
Engelstad et al. (2019; Norway)	• 26 Homicide Os with schizophrenia (HOS) • 28 Non-HOS • 71 HNOs	38.2 36.7 29.3	• Hinting task • MASC	IQ	Homicide	No first-order ToM differences between non-HOS and HOS participants. HOS participants performed significantly worse than both HNO and non-HOS participants on the advanced ToM measure	First order = 9/low Advanced = 14/moderate
Hammond & Beail (2017; United Kingdom) [a]	• 21 Os with ID • 21 NOs with ID	33.7 45.9	• The Marble story • Animation task • Faux pas	IQ and verbal knowledge	Violent offenses and sexual offenses	No significant first-order ToM difference between offenders and nonoffenders	First order = 9.5/low
Jones et al. (2007; United Kingdom)	• 15 Os • 22 HNOs	16.8 17.3		Age	Not specified	No significant first-order ToM performance differences between offenders and nonoffenders	First order = 6.5/low
Kristof et al. (2018; Hungary)	• 22 Violent Os with schizophrenia • 22 Patients with schizophrenia • 22 Violent Os without schizophrenia • 22 HNOs	36.2 37.4 33.0 28		Age and education	Violent offenses and unspecified offenses	Violent offenders performed significantly better than offenders with schizophrenia and nonoffenders with schizophrenia. No significant ToM differences between HNO and the remaining three groups	Advanced = 10/low
Majorek et al. (2009; Germany)	• 33 Os with schizophrenia • 38 NOs with schizophrenia • 29 HNOs	31.8 35.4 36.7	• Six picture stories	Age, IQ, and executive functioning	Theft, robbery, arson, sexual offenses, violent offenses, manslaughter, and other offenses	Offenders obtained lower ToM scores than nonoffenders	First/Second order = 8.5/ low

(continued)

Table 1. (continued)

Authors (Year; Country)	Sample	Mean Age	Measure of ToM	Variables Controlled in Design	Summary of Crime Types	Key Findings	Quality Rating
Mariano et al. (2017; Italy)	• 74 Os • 65 HNOs	35.7 33.7	• RMET-R • Emotion attribution task	Age and education	Violent offenses, property offenses, and sexual offenses	Offenders had lower scores than non-offenders in the RMET-R. Compared to nonoffenders, offenders were worse at recognizing sadness but were better at recognizing fear in the emotion attribution task	Advanced = 10/low
Mayer et al. (2018; Germany)	• 42 Violent Os • 32 HNOs	32.8 28.8	• Hinting task • MASC	IQ, age, and education	Violent offenses and sexual offenses	No advanced ToM differences between offenders and nonoffenders	Advanced = 16/moderate
Milojević & Dimitrijević (2014; Serbia)	• 43 Os • 47 HNOs	16.4 15.2	• RMET (child and adolescent versions)	Age, education, socioeconomic status, and IQ	Not specified	Offenders had lower scores than nonoffenders	Advanced = 13/moderate
Morosan et al. (2017; Switzerland)	• 22 Os • 23 HNOs	16.5 16.4	• GERT • Director task	IQ and age	Violent offenses, property offenses, drug offenses, and other offenses	Offenders had significantly lower scores on recognition of interest, anxiety, and amusement, and in than nonoffenders. There were no differences between the two groups in the recognition of the other emotions. Offenders performed significantly worse than nonoffenders on the Director task	Advanced = 15/moderate
Mundy (2004; England)	• 20 Os • 38 HNOs	Between 14 and 16	• RMET (child and adolescent versions)	Attachment	Not specified	Offenders had significantly lower scores than nonoffenders. However, when attachment was entered as a covariant, ToM differences between the groups disappeared	Advanced = 12/moderate
Nentjes, Bernstein, Arntz, Slaats, & Hannemann, (2015; the Netherlands)	• 42 Os with psychopathy • 40 Os without psychopathy • 26 HNOs	39.1 38.8 35.6	• Strange stories	Age, IQ, and working memory	Sexual offenses, homicide, violent offenses, property offenses, and drug offenses	No significant differences in ToM task performance between offenders and nonoffenders	Advanced = 15/moderate

(continued)

Table 1. (continued)

Authors (Year; Country)	Sample	Mean Age	Measure of ToM	Variables Controlled in Design	Summary of Crime Types	Key Findings	Quality Rating
Nentjes, Bernstein, Arntz, van Breukelen, & Slaats (2015; the Netherlands)	• 39 Os with psychopathy • 37 Os without psychopathy • 26 HNOs	38.6 39.3 35.6	• RMET-R	Age, IQ, and working memory	Homicide, violent offenses, property offenses, sexual offenses arson, and drug offenses	Offenders and nonoffenders performed equally well on the advanced ToM task Offenders with psychopathy were better than nonoffenders at recognizing hostile eyes when there was no time limit on stimulus viewing No significant differences between offenders with and without psychopathy and nonoffenders when hostility of eyes was low in the short viewing condition Offenders' incorrect answers on both long and short trials were not associated with higher level of hostility. Thus, offenders had intact advanced ToM and did not display a hostile attribution bias	
Proctor & Beail (2007; United Kingdom)	• 25 Os with ID 25 NOs with ID	31 41	• Sally–Anne task • Ice Cream Van story	IQ	Not specified	No first-order TOM deficit in offenders Offenders had better second-order ToM scores than nonoffenders	First order = 10.5/low Second order = 10/low
Robinson et al. (2007; Canada)	• 64 Os • 60 HNOs	16.3 16.6	• Empathy continuum	Age, education, and reading speed	Homicide, violent offenses, property, and nonviolent offenses	Offenders performed significantly worse than nonoffenders	First order = 11.5/low
Romero-Martínez et al. (2013; Spain)	• 19 Os • 21 HNOs	38.0 35.8	• RMET-R	Age and executive functioning	Intimate partner violence	Offenders obtained lower scores than nonoffenders	Advanced = 14/moderate
Schiffer et al. (2017; Germany)	• 16 Os with schizophrenia without CD/ASPD • 18 Os with CD/ASPD • 18 NOs with schizophrenia without CD/ASPD • 18 HNOs	38.4 35.3 37.8 36.3	• RMET-R	Age, education, and verbal IQ	Homicide, violent offenses, and sexual offenses	Offenders and nonoffenders with schizophrenia performed worse than HNO. Offenders who had CD/ASPD without schizophrenia made similar numbers of errors as HNO	Advanced = 10.5/low

(continued)

Table 1. (continued)

Authors (Year; Country)	Sample	Mean Age	Measure of ToM	Variables Controlled in Design	Summary of Crime Types	Key Findings	Quality Rating
Schuler et al. (2019; Germany)	• 85 Pedophilic Os • 72 Pedophilic NO • 128 Teleiophilic NO	39.4 33.6 33.3	• MET	Age and IQ	Sexual offenses against children	Pedophilic offenders had significantly worse ToM scores than pedophilic nonoffenders	Advanced = 17/moderate
Shamay-Tsoory et al. (2010; Israel)	• 17 Os with ASPD • 20 HNOs • 27 PNOs	29.8 27.7 38.4	• Yoni Task	Age and IQ	Not specified	No significant group differences in the first-order ToM ability Offenders had deficits in second-order affective ToM, but their second-order cognitive ToM was intact	First order = 10.5/low Second order = 10.5/low
Spenser (2017; United Kingdom)	• 100 Male Os • 100 Female Os • 100 Male HNOs • 100 Female HNOs	27.4 30.2 27.9 31.5	• The faux pas (the SSQ) and RMET-R	Age and IQ	Not specified	Offenders performed significantly worse than nonoffenders	Advanced = 14/moderate
Spenser et al. (2015; United Kingdom)	• 46 Os • 46 HNOs	21.2 19.0	• Faux pas (the SSQ)	Age, comprehension, and memory	Not specified	Offenders had significantly lower scores than nonoffenders	Advanced = 15/moderate
Winter et al. (2017; Germany)	• 29 Os • 34 HNOs	32.2 31.7	• EmpaToM	age and covaried verbal IQ and education	Violent offenses	No significant differences between offender and nonoffender groups	Advanced = 14/moderate
Woodbury-Smith et al. (2005; United Kingdom)	• 21 Os with ASD • 23 NOs with ASD • 23 HNOs	35.4 29.7 38.2	• RMET-R	IQ, memory, and executive functioning	Manslaughter, violent offenses, and arson	No ToM differences between ASD offenders and healthy nonoffenders	Advanced = 15/moderate

Note. Os = offenders; NOs = nonoffenders; HNOs = healthy nonoffenders; PNOs = patient nonoffenders; ID = intellectual disability; ASD = autistic spectrum disorder; PD = personality disorder; CD = conduct disorder; ASPD = antisocial personality disorder; GERT = Geneva Emotion Recognition Task; MASC = Movie for the Assessment of Social Cognition; MET = Multifaceted Empathy Test; RMET(-R) = Reading the Mind in the Eyes Test–Revised; MCET= The Mind in a Child's Eyes Task; SSQ= The Social Stories Questionnaire; ToM = theory of mind.

[a]We excluded the second order task in this study due to low numbers achieving second-order ToM.

difference between the second-order ToM ability of offenders and nonoffenders, Castellino et al. (2011) demonstrated second-order ToM deficits in offenders. On the other hand, Proctor and Beail (2007) found that second-order ToM among offenders was significantly better than second-order ToM among nonoffenders. Shamay-Tsoory et al. (2010) found that offenders' second-order cognitive ToM was intact, but their second-order affective ToM was impaired. In addition to these four studies, Majorek et al. (2009) used tasks that combined aspects of first- and second-order ToM (see Note 3). The results of this study suggested that ToM task performance among offenders was worse than nonoffenders.

Our quality assessment showed that the quality of the second-order studies ranged between low (three studies) and moderate (two studies). Dolan and Fullam (2004), who found no second-order ToM deficits in offenders, and Castellino et al. (2011), who found second-order ToM deficits in offenders, had the highest quality ratings. Caution is needed when interpreting and generalizing the results of both studies because, although they successfully controlled several ToM-related variables, the validity and reliability of their second-order ToM tasks for adults are yet to be established.

Summary of Advanced ToM Methods

Advanced ToM tasks investigate whether people can understand another person's higher functioning mental states, such as sarcasm, jokes, double bluffs, accusing, and preoccupation. Advanced ToM in offenders was examined by 22 studies, of which 15 utilized a single measure of advanced ToM, and seven used multiple advanced ToM measures. Advanced ToM tasks in these studies fall into three groups: tasks that measure cognitive ToM, tasks that look at affective ToM, and tasks that simultaneously evaluate cognitive and affective ToM in the same task without reporting the results separately (we refer to these as testing *cognitive–affective ToM*).

Cognitive advanced ToM. Cognitive ToM was assessed in six studies by using three different tasks. Two studies (Castellino et al., 2011; Nentjes, Bernstein, Arntz, Slaats, & Hannemann, 2015) employed the Strange Stories Task (Happé, 1994), which examines participants' understanding of various mental states involving jokes, pretense, white lies, irony, double bluffs, and sarcasm. Two studies (de Jong et al., 2018; Kristof et al., 2018) used the Faux-pas task (Baron-Cohen et al., 1999; Varga et al., 2008, respectively) in which one person tells the second person something inappropriate, hurtful, or insulting, without realizing that it should not have been said. Two studies (Morosan et al., 2017; Newbury-Helps et al., 2017) employed the perspective taking task known as the director task (Keysar et al., 2000) where participants are required to mentally position themselves where a director is standing in a scene and, when instructed by the director to move an object from some shelves, move the correct object, taking into account whether the object can be seen by the director or not.

Affective advanced ToM. Two studies examined advanced affective ToM. Mariano et al. (2017) used the emotion attribution task (Blair & Cipolotti, 2000), which measures the ability to understand the emotions of other people through stories. Morosan et al. (2017) used the Geneva Emotion Recognition Test (GERT; Schlegel et al., 2014), where emotional states were presented through videos.

Cognitive–affective advanced ToM. Seventeen studies assessed advanced cognitive–affective ToM, and five of them used two cognitive–affective tasks. Eleven studies (Domes et al., 2013; Elsegood & Duff, 2010; Mariano et al., 2017; Milojević &Dimitrijević, 2014; Mundy, 2004; Nentjes, Bernstein, Arntz, van Breukelen, & Slaats, 2015; Newbury-Helps et al., 2017; Romero-Martínez et al., 2013; Schiffer et al., 2017; Spenser, 2017; Woodbury-Smith et al., 2005) utilized the Reading the Mind in the Eyes Test–Revised (RMET-R; Baron-Cohen et al., 2001), in which a wide range of mental states are presented through pictures of eyes. Dolan and Fullam (2004) used an earlier version of the RMET-R, while another study (Elsegood & Duff, 2010) used a version of the test that presents the images of children's eyes (The Mind in a Child's Eyes Task; Duff & Schulte-Mecklenbeck, 2010). Additionally, one study (Winter et al., 2017) used a similar task to the RMET-R, the EmpaToM (Kanske et al., 2015) where different emotional states were depicted in videos. Three studies (Engelstad et al., 2019; Mayer et al., 2018; Newbury-Helps et al., 2017) used the Movie for the Assessment of Social Cognition task (Dziobek et al., 2006), which required participants to watch video clips and answer questions involving intentions, feelings, and thoughts of the characters in the videos. Two studies (Spenser, 2017; Spenser et al., 2015) used the Social Stories Questionnaire (Lawson et al., 2004), which is similar to the faux-pas task that was described above, but this task also included an affective component of ToM. Additionally, a study by Dolan and Fullam (2004) used a cognitive–affective faux-pas task, and two studies (Domes et al., 2013; Schuler et al., 2019) measured advanced ToM with the Multifaceted Empathy Test (MET; Dziobek et al., 2008), which contains pictures of people in emotionally charged situations from everyday life.

Advanced ToM results. Among the 22 studies, seven did not find differences in advanced ToM between the offenders and nonoffenders (Kristof et al., 2018; Mayer et al., 2018; Mundy, 2004; Nentjes, Bernstein, Arntz, Slaats, & Hannemann, 2015; Nentjes, Bernstein, Arntz, van Breukelen, & Slaats, 2015; Winter et al., 2017; Woodbury-Smith et al., 2005), whereas nine studies found deficits in advanced ToM in offenders (Castellino et al., 2011; Engelstad et al., 2019; Mariano et al., 2017; Milojević & Dimitrijević, 2014; Newbury-Helps et al., 2017; Romero-Martínez et al., 2013; Schuler et al., 2019; Spenser, 2017; Spenser et al., 2015). Additionally, six studies reported inconsistent patterns of impairment, depending on the task that was assessed. For example, Domes et al. (2013) found deficits in advanced ToM among offenders using the MET (Dziobek et al., 2008), but no difference between groups using the

RMET-R (Baron-Cohen et al., 2001). Further, De Jong et al. (2018) found that violent offenders who had psychotic disorder scored lower on advanced ToM than both nonviolent participants who had psychotic disorder and healthy control participants. However, discriminant function analyses indicated that between-group differences were better explained by impairments in metacognition and neurocognitive function than advanced ToM (as measured by faux pas). Morosan et al. (2017) found a deficit in advanced ToM among offenders using the director task and a partial impairment in advanced ToM using the GERT, where offenders scored lower than nonoffenders on recognition of interest, anxiety, and amusement. Additionally, Schiffer et al. (2017) found that violent offenders who had schizophrenia and nonoffenders with schizophrenia had lower scores on tests of advanced ToM than healthy controls. However, violent offenders with conduct disorder or antisocial personality disorder[4] without schizophrenia had similar advanced ToM scores to healthy nonoffenders.

The remaining two advanced ToM studies demonstrated a selective impairment in advanced ToM among offenders. That is, Dolan and Fullam (2004) found that offenders and nonoffenders did not differ in understanding complex emotions from pictures, detecting faux pas, or identifying the person who committed the faux pas. However, offenders were worse than nonoffenders at recognizing basic emotions from pictures, understanding the mental state of the person who committed a faux pas or the person to whom the faux pas was made. Additionally, Elsegood and Duff (2010) reported that individuals who committed sexual offenses against children showed impairments in advanced ToM when inferring the mental states of adults but intact advanced ToM when understanding the mental states of children (i.e., the age-group consistent with their victims).

The quality of the reviewed studies that tested advanced ToM ranged between low and high. Four of the 22 studies were of low quality, 17 were of moderate quality, and one was of high quality. The highest quality study (Newbury-Helps et al., 2017) reported that offenders had lower advanced ToM scores than nonoffenders on all advanced ToM measures. This study included a control group who did not have criminal records; selected participants who did not have a learning disability or head injury; controlled for potentially confounding variables such as age, education, and verbal intelligence; and additionally assessed participants' memory, attention, and comprehension capacity in control questions.

Offense Type and ToM

In this review, we also explored the relationship between ToM and crime type. Among studies examining first-order ToM, three studies did not report crime type. Studies that focused on individuals who committed sexual offenses (Castellino et al., 2011; Hammond & Beail, 2017) and violent offenses (Dolan & Fullam, 2004; Engelstad et al., 2019; Hammond & Beail, 2017) found no first-order ToM deficits in these offender groups. Studies that found first-order ToM deficits in offenders (Majorek et al., 2009; Robinson et al., 2007) had recruited mixed offender groups reflecting six or more different criminal offense types, including but not limited to, sexual and violent crimes. It is possible that presence or absence of first-order ToM deficits in offenders relates to the crime type they committed. However, the current evidence does not allow us to draw firm conclusions.

In two of the five studies examining second-order ToM studies, researchers did not report a breakdown of the crime types in the sample. The study by Majorek et al. (2009), which employed an offender group with mixed crime types, found a second-order ToM impairment in this group. Castellino et al. (2011) also reported second-order ToM deficits in individuals who committed sexual offenses against children and adults. Additionally, the study by Dolan and Fullam (2004), which consisted of individuals who committed violent crimes, found no second-order ToM deficits in this offending group. As with the findings for first-order ToM, the heterogeneity of findings limits the conclusions we were able to draw regarding second-order ToM and crime type, though it appears that violent crime, at least, is not strongly associated with second-order deficits.

Again, we examined whether crime type[5] relates to the patterns of findings in studies examining advanced ToM. Four studies that included participants who had exclusively committed sexual offenses found that those individuals had global or selective impairments in advanced ToM (Castellino et al., 2011; Elsegood & Duff, 2010; Schuler et al., 2019). A study by Romero-Martínez et al. (2013) found that individuals who perpetrated intimate partner violence had advanced ToM deficits. Studies of individuals convicted of violent crimes yielded mixed findings, with different measures/studies indicating impairment (Engelstad et al., 2019; Newbury-Helps et al., 2017), no impairment (de Jong et al., 2018; Winter et al., 2017), or selective impairment (Dolan & Fullam, 2004) in advanced ToM. Studies that included offender groups consisting of five or more different criminal offense types yielded mixed results. Some of these studies showed no deficits in advanced ToM among offenders (Mayer et al., 2018; Nentjes, Bernstein, Arntz, Slaats, & Hannemann, 2015; Nentjes, Bernstein, Arntz, van Breukelen, & Slaats, 2015; Woodbury-Smith et al., 2005). However, findings were mixed in other studies showing impairment and no impairment in advanced ToM depending on the type of ToM tasks that were used or participant psychopathology (Domes et al., 2013; Kristof et al., 2018; Mariano et al., 2017; Morosan et al., 2017; Schiffer et al., 2017).

Discussion

This systematic review examined ToM in offenders by reviewing 28 published studies. Overall, our review revealed inconsistent and sometimes conflicting results for first-order, second-order, and advanced ToM among offenders. There are many potential reasons for these discrepancies, most notably the fact that so little research has been conducted on ToM in offenders (recall that we used a broad range of search terms and

Table 2. Summary of Critical Findings.

- Studies examining theory of mind (ToM) in individuals who have committed offenses have been characterized by limitations impacting the quality of the findings
- Typical limitations include ToM measures that may not be suitable for the population being tested, lack of control for potential confounding variables, and use of small heterogeneous samples
- Our review found that the literature is characterized by mixed findings
- On balance, there is little clear evidence for deficits in first- and second-order ToM among individuals who have committed offenses
- There is greater evidence for impairments in advanced ToM among people who have committed offenses

no publication date restrictions to maximize our selection of studies). Another important reason for the inconsistent findings is that the studies reviewed here employed a range of different ToM tasks to measure the same domain, meaning that comparison across studies was difficult. This finding fits with recent observations of wide variability in ToM performance among children, adolescents, and adults, as well as minimal correlations between ToM tasks (Warnell & Redcay, 2019). In fact, our review showed that even when identical ToM measures were employed, this did not always produce the same outcome across studies, even in similar offending populations (e.g., convicted individuals with antisocial personality disorder; Newbury-Helps et al., 2017; Schiffer et al., 2017). As such, ToM should be considered a multidimensional process that relies on input from a number of other abilities, which is likely to have led to the inconsistencies seen here. Relevant factors might include differences in the samples, such as sample size (sample size of the offenders ranged from 15 to 200), cognitive abilities, offending history, variance in early socialization of offenders, levels of neuroticism (Dolan & Fullam, 2004), and offenders' differential relatedness, or closeness, to their victims (Möller et al., 2014). These factors were rarely measured or controlled in the studies we identified.

Contradictory results might also have stemmed from limitations in the studies themselves, as reflected in the quality ratings (see Table 2 for critical findings). For example, the majority of studies included in the review failed to control for important confounding differences in cognitive abilities (e.g., intelligence, knowledge of vocabulary, executive functioning, and working memory capacity). Thus, ToM performance might have been influenced by confounding factors, and it is not clear whether the findings from these studies truly reflect the relationship between offending and ToM ability.

The type of crime committed by offender groups employed in each study might also be a factor that contributed to contradictory results. Although the number of available studies was small, our review suggests that different crime types may produce different ToM outcomes for first-order, second-order, and advanced ToM. Specifically, our review suggests that sexual offending may not be related to first-order ToM deficits but that it may be associated with second-order and advanced ToM deficits. Violent offending appears unrelated to first- and second-order ToM and is inconsistent on measures of advanced ToM. Given the dearth of studies looking at ToM in sexual and violent offending, any possible relationships between these types of offending and ToM deficits should be viewed with caution. We believe that understanding the link between ToM and crime types is important to determine whether certain offending groups need a treatment program that includes a ToM component. We suggest that there is a need to conduct rigorous ToM studies that compare distinct categories of offenders rather than combining individuals with mixed offense types.

The choice of tasks used to assess ToM in offenders is another important factor that is likely to contribute to the inconsistent results. First, we note that the tasks were simple response-based tasks, many of which were originally developed for child or clinical samples, and therefore have the potential for ceiling performance in adults. The wider research on ToM in healthy adults has developed sophisticated tasks that examine real-time inferences about others' mental states and are therefore more sensitive to subtle processing differences between individuals. These tasks provide insights not only into whether a person's ToM is impaired or not but also the mechanisms and timings with which these inferences are made (e.g., Bradford et al., 2015; Ferguson & Breheny, 2012; Kovács et al., 2010; Samson et al., 2010). Future research on ToM and offending should therefore adopt some of these more complex tasks to identify the specific nature of any difficulties in ToM.

Second, our quality assessment identified concerns about the validity and reliability of some of the ToM tasks employed by the studies we reviewed. Most did not report the validity and reliability of their ToM tasks. While many of the basic ToM tasks featured in these studies have been used frequently in the literature, there is insufficient evidence to show that they are valid, reliable, and are suitable to demonstrate individual differences in adults. For example, the mental state items in the faux-pas task have good test–retest reliability (Zhu et al., 2007) and excellent internal consistency, but the control items have low internal consistency, skewed distribution, and ceiling effects (Söderstrand & Almkvist, 2012).

The widely used RMET-R has been criticized for its association with verbal intelligence (Baker et al., 2014) and concerns that it reflects emotion recognition rather than ToM (Oakley et al., 2016). However, there is now evidence that ToM has cognitive and affective components (Shamay-Tsoory et al., 2010), and affective ToM refers to understanding others' emotions (Gabriel et al., 2019). Further, there is a claim that the RMET-R assesses mental states more comprehensively than the earlier version of the RMET, including basic and complex emotions, cognitive mental states (e.g., thinking and scheming), and relational mental states such as flirting (Warrier et al., 2017). Considering the current state of the RMET-R and its wide use throughout offending research, we decided to include the studies that used the RMET-R for completeness. Nevertheless, we note that it is important to clarify the

controversy around what the RMET-R measures to help researchers assess the usability of the RMET-R.

Another aspect that is overlooked by the majority of reviewed studies is the importance of measuring ToM as a construct that has two distinct, cognitive and affective, components. While accumulating evidence from empirical studies supports this distinction (Hynes et al., 2006; Shamay-Tsoory et al., 2002, 2005; Vollm et al., 2006), the reviewed studies—with the exception of Shamay-Tsoory et al. (2010)—either treated ToM as a single construct or did not take this distinction into account adequately when analyzing or reporting their results. Moreover, the majority of studies reported here employed only a single task to measure one aspect of ToM rather than a multimodal approach that uses a battery of ToM tasks to examine the broad spectrum of ToM skills. Failure to assess ToM as a multifaceted construct leads to an ambiguity about the source of potential ToM deficit, because it does not enable us to identify whether a ToM deficit is cognitive or affective, and within these subdivisions the specific mechanisms that are impaired/intact.

We suggest that ambiguities on which specific components of ToM each task measures partly stem from a more general problem in defining cognitive and affective ToM. For example, cognitive ToM has been defined as "our ability to make inference regarding other people's beliefs," whereas affective ToM was described as an "inference one makes regarding others' emotions" (Shamay-Tsoory et al., 2010, p. 669). These definitions are not explicit enough to prevent confusion over the task selection to evaluate cognitive and affective ToM. We suggest that explicit definitions for cognitive and affective ToM, and detailed information about what qualities they measure, are essential. The definition must clearly state whether cognitive ToM just detects thoughts, beliefs, and intentions or whether it also detects emotions. The definition of affective ToM must express whether it simply detects emotions and feelings of others or whether it also includes understanding the detected emotions and feelings of others. Without clearly defining these aspects and identifying the tasks that measure each dimension, research in this area might unintentionally be misleading and misinforming researchers, as well as practitioners and policy makers.

The studies that met our inclusion criteria assessed ToM only quantitatively. However, a small number of studies in the ToM literature examining the content of ToM suggest that although children with problem behaviors (e.g., antisocial behavior, conduct disorder) have intact ToM, the content of their ToM is problematic. They may, therefore, have a *Theory of Nasty Mind* (Happé & Frith, 1996). Research that has examined ToM in children who have behavioral problems, or who have been rejected by peers, suggests that these phenomena may be associated with deficits in the content of their ToM (Badenes et al., 2000; Hughes et al., 1998). Although none of the studies we reviewed assessed Theory of Nasty Mind directly, Nentjes, Bernstein, Arntz, van Breukelen, & Slaats (2015) examined hostile attribution bias in offenders and nonoffenders. Hostile attribution bias has been conceptualized as either Theory of Nasty Mind (Blair, 2003) or a cognitive process that is caused by ToM deficits (Kinderman et al., 1998). Nentjes, Bernstein, Arntz, van Breukelen, & Slaats (2015) presented participants with pictures from the RMET-R (Baron-Cohen et al., 2001) but altered the response options by adding some hostile words (e.g., "aggressive" and "attacking"). The results showed that when the pictures were presented briefly (1,000 ms), there was no difference between the scores of offenders and nonoffenders. In addition, when the relative hostility of eyes was low, offenders did not make significantly more mistakes than nonoffenders. However, when stimuli were presented with no time restriction, offenders were better at identifying hostility than nonoffenders. This finding suggests that, under certain conditions, offenders may have a potentially adaptive Theory of Nasty Mind that informs their emotion recognition. Future studies should therefore investigate the content of ToM, to shed light on its relationship with offending and indeed to examine how it may interact with the types of ToM deficits examined in this review.

The current systematic review also has its own limitations. For example, we included studies that used small sample sizes (Castellino et al., 2011; Jones et al., 2007; Morosan et al., 2017; Proctor & Beail, 2007; Romero-Martínez et al., 2013; Shamay-Tsoory et al., 2010; Woodbury-Smith et al., 2005). Two of our included studies were with participants who had intellectual disability (i.e., Hammond & Beail, 2017; Proctor & Beail, 2007), meaning that their conclusions may not generalize to other populations. While the studies contributing to the systematic review spanned 11 countries, all bar two were based in Europe. As a result, our conclusions from these studies—tentative as they are—may not generalize to other jurisdictions or cultural contexts. Importantly, the studies included in the review overwhelmingly focused on male offending and ToM. Only around 7% of offending participants were reported as female; thus, any conclusions may not apply to females. Furthermore, we included only studies that had a nonoffender control group resulting in the omission of a study (Richell et al., 2003) that did not have a control group but instead compared the mean score of their experimental group with the mean score of nonoffenders who participated in another study.

We also excluded studies that used interview methods such as the Reflective Functioning Task which focuses on participants' attachment experiences with their parents during childhood (Fonagy et al., 1997). We had two main reasons for these exclusions. First, in tasks such as these, the accuracy of participant inferences about others' mental states is unknown to the researchers who score participant interviews. Therefore, researchers can rate whether or not participants articulate certain mental states but cannot know whether their inference of these states is accurate. For example, a participant who states, "I thought my mother felt resentful of us, but I'm not really sure if she felt that way herself," would get a point for mental inference in the Reflective Functioning Task, but the researcher would not know whether the mother felt resentful or not. In contrast, the types of ToM measure included in our review, in which mental inferences are presented through pictures, videos,

or stories, provide researchers with certainty about the accuracy of the participants' mental state inferences of others' minds.

The second reason for exclusion is that the studies using reflective functioning as a measure of ToM is limited by a focus on mental inferences specific to attachment figures. Fonagy et al. (1998) suggested that reflective capacity in the attachment context may not generalize to other domains. We also excluded studies that used questionnaires asking participants to self-report their ability to theorize about other people's minds (e.g., I find it easy to put myself in somebody else's shoes) because scores obtained from questionnaire may not reflect the true ToM abilities of participants and may not predict actual ToM abilities in everyday situations (Queirós et al., 2018). While our focus on studies that measured ToM using performance-based tasks limited the scope of our systematic review, we believe that doing so eliminated a number of potential confounding or contaminating factors.

Our quality assessment also has limitations. Evaluating task difficulty and the ecological validity of the tasks required subjective judgments. In addition, studies that scored highest on quality showed that there were no first-order ToM differences between offenders and nonoffenders and produced inconsistent results for second-order ToM. However, these studies were still only of moderate quality, and, like the relatively lower ranking studies, they had shortcomings of their own (e.g., using an age-inappropriate task for adults, failing to control some potential confounding variables, treating ToM as a single construct, and recruiting offenders who had various personality disorders). These factors may have an influence on ToM task performance, directly or indirectly, because studies indicate that they are significant moderators or predictors of ToM success (Brock et al., 2018; Spenser et al., 2019). Therefore, despite their higher quality scores, caution is recommended when drawing conclusions from the studies, and future research should aim to overcome these shortcomings.

In conclusion, the current review adds to a growing body of literature on ToM in offending populations in several aspects (see "Implications of the Review" section). The vast majority of the studies in the review indicated that offenders had intact first-order ToM. On the other hand, results regarding second-order and advanced ToM were more mixed. Some studies found that offenders had intact second-order and advanced ToM, whereas others found that both were impaired. Even more curiously, a number of studies found superior ToM among offenders or reported selective impairment in their second-order and advanced ToM. However, we note that the vast majority of studies used a single response-based measure of ToM to assess ToM as a single construct rather than considering its cognitive and affective aspects independently. Consequently, these studies do not clearly distinguish whether offenders had intact/impaired cognitive or affective ToM, or both, or identify the specific mechanisms that are impaired. This review demonstrates that the relationship between ToM and offending is complex and influenced by multiple factors. We suggest that the only way to have an accurate understanding about the relationship between ToM and offending is to establish clear definitions and distinctions for ToM components, use valid and reliable ToM measures, and conduct well-designed studies. Finally, it remains an open question whether ToM impairment may be criminogenic and whether it may form a treatment need within offender rehabilitation.

Implications of the Review

Research

- Our review shows clear gaps in the research on ToM and offending.
- Researchers should ensure precision in the definition and operationalization of ToM constructs.
- ToM measures should be validated for use with the population under investigation.
- Confounds should be controlled where feasible.
- Research designs should allow for ToM data to be examined across offense types.
- Researchers should use our quality checklist to help guide study design.

Policy

- There is currently no clear evidence to suggest that work on first-order, second-order, or advanced ToM should be routinely incorporated into treatment programs for individuals who have offended.
- Funders should prioritize rigorous and generalizable research on ToM and offending.

Practice

- Practitioners should carefully examine the weight of evidence for ToM deficits in their client group, paying close attention to the quality of studies and the limitations of the evidence base.
- Practitioners should use case formulation to explore whether a facet of ToM represents a treatment need for individual clients.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Caoilte Ó Ciardha https://orcid.org/0000-0001-5383-8403

Supplemental Material

The supplemental material for this article is available online.

Notes

1. Baron-Cohen et al. (2001) suggested that basic emotions are recognized universally, even by very young children without needing to infer the mental states of the other individual.

2. We excluded the Th.o.m.a.s task (Bosco et al., 2009) from Castellino et al.'s (2011) study because this task did not meet our inclusion criteria (used both interview and questionnaire measures). It should be noted that this decision did not affect the review outcome as the result from the task showed offenders had theory of mind (ToM) impairment, which is congruent with the result obtained from the other task included in this review.
3. The task that was used by Majorek et al. (2009) measured both first-order and second-order ToM. However, the authors did not define the tasks as being first- or second-order ToM and did not report the results separately. We were not able to disentangle the results for each aspect of ToM and therefore included their results in both first-order and second-order ToM sections.
4. Schiffer et al. (2017) also included a sample of violent offenders with schizophrenia and conduct disorder or antisocial personality disorder. The sample size was below our minimum for inclusion in the systematic review, though we note that they reported no difference between the ToM scores of this subsample and healthy controls.
5. Studies by Milojević and Dimitrijević (2014), Mundy (2004), Spencer (2017), and Spenser et al. (2015) were not included in crime type analysis because the offense type was not reported in their articles.

References

References marked with an asterisk indicate studies included in the systematic review.

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

Abu-Akel, A., & Abushua'leh, K. (2004). "Theory of mind" in violent and nonviolent patients with paranoid schizophrenia. *Schizophrenia Research*, *69*(1), 45–53. https://doi.org/10.1016/S0920-9964(03)00049-5

Andrews, D. A., & Bonta, J. (2010). *The psychology of criminal conduct* (5th ed.). Anderson.

Badenes, L. V., Estevan, R. C., & Bacete, F. G. (2000). Theory of mind and peer rejection at school. *Social Development*, *9*(3), 271–283. https://doi.org/10.1111/1467-9507.00125

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

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. The MIT Press.

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

Baron-Cohen, S., O'Riordan, M., Stone, V., Jones, R., & Plaisted, K. (1999). Recognition of faux pas by normally developing children with Asperger syndrome or high-functioning autism. *Journal of Autism and Developmental Disorders*, *29*(5), 407–418. https://doi.org/10.1023/A:1023035012436

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

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

Bellerose, J. S., Beauchamp, M. H., & Lassonde, M. (2011). New insights into neurocognition provided by brain mapping: Social cognition and theory of mind. In D. Hugues (Ed.), *Brain mapping: From neural basis of cognition to surgical applications* (pp. 181–190). Springer Vienna.

Blair, R. J. R. (2003). Did Cain fail to represent the thoughts of Abel before he killed him? The relationship between theory of mind and aggression. In B. Repacholi & V. Slaughter (Eds.), *Individual differences in theory of mind: Implications for typical and atypical development* (pp. 143–170). Psychology Press.

Blair, R. J. R., & Cipolotti, L. (2000). Impaired social response reversal: A case of "acquired sociopathy." *Brain: A Journal of Neurology*, *123*(6), 1122–1141. https://doi.org/10.1093/brain/123.6.1122

Bosco, F. M., Colle, L., De Fazio, S., Bono, A., Ruberti, S., & Tirassa, M. (2009). Th.o.m.a.s.: An exploratory assessment of theory of mind in schizophrenic subjects. *Consciousness and Cognition: An International Journal*, *18*(1), 306–319. https://doi.org/10.1016/j.concog.2008.06.006

Bradford, E. E. F., Hukker, V., Smith, L., & Ferguson, H. J. (2018). Belief-attribution in adults with and without autistic spectrum disorders. *Autism Research*, *11*, 1542–1553.

Bradford, E. E. F., Jentzsch, I., & Gomez, J. (2015). From self to social cognition: Theory of mind mechanisms and their relation to executive functioning. *Cognition*, *138*, 21–34. https://doi.org/10.1016/j.cognition.2015.02.001

Brock, L. L., Kim, H., Gutshall, C. C., & Grissmer, D. W. (2018). The development of theory of mind: Predictors and moderators of improvement in kindergarten. *Early Child Development and Care*, *189*(12), 1914–1924. https://doi.org/10.1080/03004430.2017.1423481

Brothers, L., & Ring, B. (1992). A neuroethological framework for the representation of minds. *Journal of Cognitive Neuroscience*, *4*(2), 107–118. https://doi.org/10.1162/jocn.1992.4.2.107

Brown, S., Harkins, L., & Beech, A. R. (2012). General and victim-specific empathy: Associations with actuarial risk, treatment outcome, and sexual recidivism. *Sexual Abuse: Journal of Research and Treatment*, *24*(5), 411–430. https://doi.org/10.1177/1079063211423944

Brunsdon, V., Bradford, E., & Ferguson, H. J. (2019). Sensorimotor mu rhythm during action observation changes across the lifespan independently from social cognitive processes. *Developmental Cognitive Neuroscience*, *38*, 100659.

Brüne, M. (2003). Social cognition and behaviour in schizophrenia. In M. Brüne, H. Ribbert, & W. Schiefenhövel (Eds.), *The social brain: Evolution and pathology* (pp. 277–313). Wiley & Sons.

Brüne, M., & Brüne-Cohrs, U. (2006). Theory of mind—Evolution, ontogeny, brain mechanisms and psychopathology. *Neuroscience and Biobehavioral Reviews*, *30*(4), 437–455. https://doi.org/10.1016/j.neubiorev.2005.08.001

Buttelmann, D., Carpenter, M., & Tomasello, M. (2009). Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition*, *112*(2), 337–342. https://doi.org/10.1016/j.cognition.2009.05.006

Cane, J. E., Ferguson, H. J., & Apperly, I. (2017). Using perspective to resolve reference: The impact of cognitive load and motivation. *Journal of Experimental Psychology: Language, Memory & Cognition*, *43*, 591–610.

*Castellino, N., Bosco, F. M., Marshall, W. L., Marshall, L. E., & Veglia, F. (2011). Mindreading abilities in sexual offenders: An analysis of theory of mind processes. *Consciousness and Cognition: An International Journal*, *20*(4), 1612–1624. https://doi.org/10.1016/j.concog.2011.08.011

Charlton, R. A., Barrick, T. R., Markus, H. S., & Morris, R. G. (2009). Theory of mind associations with other cognitive functions and brain imaging in normal aging. *Psychology and Aging*, *24*(2), 338–348. https://doi.org/10.1037/a0015225

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

Day, A., Casey, S., & Gerace, A. (2010). Interventions to improve empathy awareness in sexual and violent offenders: Conceptual, empirical, and clinical issues. *Aggression and Violent Behavior*, *15*(3), 201–208. https://doi.org/10.1016/j.avb.2009.12.003

*de Jong, S., van Donkersgoed, R., Renard, S., Carter, S., Bokern, H., Lysaker, P., Gaag, M. V. D., Aleman, A., & Pijnenborg, G. H. M. (2018). Social-cognitive risk factors for violence in psychosis: A discriminant function analysis. *Psychiatry Research*, *265*, 93–99. https://doi.org/10.1016/j.psychres.2018.04.048

Devine, R. T., & Hughes, C. (2014). Relations between false belief understanding and executive function in early childhood: A meta-analysis. *Child Development*, *85*(5), 1777–1794.

*Dolan, M., & Fullam, R. (2004). Theory of mind and mentalizing ability in antisocial personality disorders with and without psychopathy. *Psychological Medicine*, *34*(6), 1093–1102. https://doi.org/10.1017/S0033291704002028

Duff, S. C., & Schulte-Mecklenbeck, M. (*2010*). *The mind in a child's eyes: A test of theory of mind skills with child stimuli*. Unpublished manuscript.

Duval, C., Piolino, P., Bejanin, A., Eustache, F., & Desgranges, B. (2011). Age effects on different components of theory of mind. *Consciousness and Cognition: An International Journal*, *20*(3), 627–642. https://doi.org/10.1016/j.concog.2010.10.025

*Domes, G., Hollerbach, P., Vohs, K., Mokros, A., & Habermeyer, E. (2013). Emotional empathy and psychopathy in offenders: An experimental study. *Journal of Personality Disorders*, *27*(1), 67–84. https://doi.org/10.1521/pedi.2013.27.1.67

Dymond, R. F. (1949). A scale for the measurement of empathic ability. *Journal of Consulting Psychology*, *13*(2), 127–133.

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

Dziobek, I., Rogers, K., Fleck, S., Bahnemann, M., Heekeren, H. R., Wolf, O. T., & Convit, A. (2008). Dissociation of cognitive and emotional empathy in adults with Asperger syndrome using the Multifaceted Empathy Test (MET). *Journal of Autism and Developmental Disorders*, *38*(3), 464–473. https://doi.org/10.1007/s10803-007-0486-x

Eisenberg, N., Fabes, R. A., & Spinrad, T. L. (2006). Prosocial development. In N. Eisenberg, W. Damon, R. M. Lerner, N. Eisenberg, W. Damon, & R. M. Lerner (Eds.), *Handbook of child psychology: Vol. 3, Social, emotional, and personality development* (6th ed.) (pp. 646–718). John Wiley & Sons.

*Elsegood, K. J., & Duff, S. C. (2010). Theory of mind in men who have sexually offended against children: A U.K. Comparison study between child sex offenders and nonoffender controls. *Sexual Abuse: Journal of Research and Treatment*, *22*(1), 112–131. https://doi.org/10.1177/1079063209359926

*Engelstad, K. N., Rund, B. R., Torgalsbøen, A.-K., Lau, B., Ueland, T., & Vaskinn, A. (2019). Large social cognitive impairments characterize homicide offenders with schizophrenia. *Psychiatry Research*, *272*, 209–215. https://doi.org/10.1016/j.psychres.2018.12.087

Ferguson, H. J., & Breheny, R. (2012). Listeners' eyes reveal spontaneous sensitivity to others' perspectives. *Journal of Experimental Social Psychology*, *48*(1), 257–263. https://doi.org/10.1016/j.jesp.2011.08.007

Ferguson, H. J., Brunsdon, V., & Bradford, E. (2018). Age of avatar modulates the altercentric bias in a visual perspective-taking task: ERP and behavioural evidence. *Cognitive, Affective, & Behavioral Neuroscience*, *18*, 1298–1319.

Ferguson, H. J., Cane, J. E., Douchkov, M., & Wright, D. (2015). Empathy predicts false belief reasoning ability: Evidence from the N400. *Social Cognitive and Affective Neuroscience*, *10*(6), 848–855.

Frith, C. D. (1992). *The cognitive neuropsychology of schizophrenia*. Lawrence Erlbaum Associates.

Fonagy, P., & Levinson, A. (2004). Offending and attachment: The relationship between interpersonal awareness and offending in a prison population with psychiatric disorder. *Canadian Journal of Psychoanalysis*, *12*(2), 225–251. https://doi.org/10.1002/j.2167-4086.2009.tb00406.x

Fonagy, P., Steele, M., Steele, H., & Target, M. (1997). *Reflective-functioning manual, version 4.1, for application to adult attachment interviews*. University College London.

Fonagy, P., Target, M., Steele, H., & Steele, M. (1998). *Reflective functioning manual. Version 5 for application to adult attachment interviews*. University College London.

Gabriel, E. T., Oberger, R., Schmoeger, M., Deckert, M., Vockh, S., Auff, E., & Willinger, U. (2019). Cognitive and affective theory of mind in adolescence: Developmental aspects and associated neuropsychological variables. *Psychological Research*. https://doi.org/10.1007/s00426-019-01263-6

*Hammond, S., & Beail, N. (2017). Social-moral awareness and theory of mind in adult offenders who have intellectual disabilities. *Journal of Intellectual Disabilities and Offending Behaviour*, *8*(3), 111–121, https://doi.org/10.1108/JIDOB-12-2016-0024

Hanson, R. K., & Morton-Bourgon, K. (2004). *Predictors of sexual offender recidivism: An updated meta-analysis* (Corrections Research User Report No. 2004–02). Public Safety and Emergency Preparedness Canada.

Hanson, R. K., & Morton-Bourgon, K. E. (2005). The characteristics of persistent sexual offenders: A meta-analysis of recidivism studies. *Journal of Consulting and Clinical Psychology*, *73*(6), 1154–1163. https://doi.org/10.1037/0022-006X.73.6.1154

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

Happé, F. E., Brownell, H., & Winner, E. (1999). Acquired "theory of mind" impairments following stroke. *Cognition*, *70*(3), 211–240. https://doi.org/10.1016/S0010-0277(99)00005-0

Happé, F. E., & Frith, U. (1996). Theory of mind and social impairment in children with conduct disorder. *British Journal of Developmental Psychology*, *14*(4), 385–398. https://doi.org/10.1111/j.2044-835X.1996.tb00713.x

Hogrefe, G.-J., Wimmer, H., & Perner, J. (1986). Ignorance versus false belief: A developmental lag in attribution of epistemic states. *Child Development*, *57*(3), 567–582. https://doi.org/10.2307/1130337

Hughes, C., Dunn, J., & White, A. (1998). Trick or treat? Uneven understanding of mind and emotion and executive dysfunction in "hard-to-manage" preschoolers. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *39*(7), 981–994. http://doi.org/9804031

Hynes, C. A., Baird, A. A., & Grafton, S. T. (2006). Differential role of the orbital frontal lobe in emotional versus cognitive perspective-taking. *Neuropsychologia*, *44*(3), 374–383. https://doi.org/10.1016/j.neuropsychologia.2005.06.011

Jolliffe, D., & Farrington, D. P. (2004). Empathy and offending: A systematic review and meta-analysis. *Aggression and Violent Behavior*, *9*(5), 441–476. https://doi.org/10.1016/j.avb.2003.03.001

*Jones, A. P., Forster, A. S., & Skuse, D. (2007). What do you think you're looking at? Investigating social cognition in young offenders. *Criminal Behaviour and Mental Health*, *17*(2), 101–106. https://doi.org/10.1002/cbm.641

Kanske, P., Böckler, A., Trautwein, F.-M., & Singer, T. (2015). Dissecting the social brain: Introducing the EmpaToM to reveal distinct neural networks and brain–behavior relations for empathy and theory of mind. *NeuroImage*, *122*, 6–19. https://doi.org/10.1016/j.neuroimage.2015.07.082

Keenan, T., & Ward, T. (2000). A theory of mind perspective on cognitive, affective, and intimacy deficits in child sexual offenders. *Sexual Abuse: Journal of Research and Treatment*, *12*(1), 49–60. https://doi.org/10.1177/107906320001200106

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

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, *11*(1), 32–38.

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

Klin, A. (2000). Attributing social meaning to ambiguous visual stimuli in higher-functioning autism and Asperger syndrome: The social attribution task. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *41*(7), 831–846. http://doi.org/10.1111/1469-7610.00671

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science*, *330* (6012), 1830–1834.

*Kristof, Z., Kresznerits, S., Olah, M., Gyollai, A., Lukacs-Miszler, K. , Halmai, T., Fountoulakis, K N., Tenyi, T., Dome, P., & Gonda, X. (2018). Mentalization and empathy as predictors of violence in schizophrenic patients: Comparison with nonviolent schizophrenic patients, violent controls and nonviolent controls. *Psychiatry Research*, *268*, 198–205. https://doi.org/10.1016/j.psychres.2018.07.021

Laws, D. R., & Ward, T. (2011). *Desistance from sex offending: Alternatives to throwing away the keys*. Guilford Press.

Lawson, J., Baron-Cohen, S., & Wheelwright, S. (2004). Empathising and systemising in adults with and without Asperger syndrome. *Journal of Autism and Developmental Disorders*, *34*(3), 301–310. https://doi.org/10.1023/B:JADD.0000029552.42724.1b

Li, X., Wang, K., Wang, F., Tao, Q., Xie, Y., & Cheng, Q. (2013). Aging of theory of mind: The influence of educational level and cognitive processing. *International Journal of Psychology*, *48*(4), 715–727. https://doi.org/10.1080/00207594.2012.673724

*Majorek, K., Wolfkühler, W., Küper, C., Saimeh, N., Juckel, G., & Brüne, M. (2009). "Theory of mind" and executive functioning in forensic patients with schizophrenia. *Journal of Forensic Sciences*, *54*(2), 469–473. https://doi.org/10.1111/j.1556-4029.2008.00966.x

*Mariano, M., Pino, M. C., Peretti, S., Valenti, M., & Mazza, M. (2017). Understanding criminal behavior: Empathic impairment in criminal offenders. *Social Neuroscience*, *12*(4), 379–385. https://doi.org/10.1080/17470919.2016.1179670

*Mayer, S. V., Jusyte, A., Klimecki-Lenz, O. M., & Schönenberg, M. (2018). Empathy and altruistic behavior in antisocial violent offenders with psychopathic traits. *Psychiatry Research*, *269*, 625–632. https://doi.org/10.1016/j.psychres.2018.08.035

*Milojević, S., & Dimitrijević, A. (2014). Empathic capacity of delinquent convicted minors. *PSIHOLOGIJA*, *47*(1), 65–79. https://doi.org/10.2298/PSI1401065M

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine*, *151*, 264–26. https://doi.org/10.7326/0003-4819-151-4-200908180-00135

Möller, C., Falkenström, F., Holmqvist Larsson, M., & Holmqvist, R. (2014). Mentalizing in young offenders. *Psychoanalytic Psychology*, *31*(1), 84–99. https://doi.org/10.1037/a0035555

*Morosan, L., Badoud, D., Zaharia, A., Brosch, T., Eliez, S., Bateman, A., Heller, P, & Debbané, M. (2017). Emotion recognition and perspective taking: A comparison between typical and incarcerated

male adolescents. *PLoS One*, *12*(1). https://doi.org/10.1371/journal.pone.0170646

*Mundy, S. (2004). *Attachment and theory of mind in young offenders* [Unpublished doctoral dissertation]. University College London.

National Institute for Health and Care Excellence. (2012). *Methods for the development of NICE public health guidance* (3rd ed.). National Institute for Health and Care Excellence. https://www.nice.org.uk/article/pmg4/chapter/appendix-g-quality-appraisal-checklist-quantitative-studies-reporting-correlations-and-associations

*Nentjes, L., Bernstein, D. P., Arntz, A., Slaats, M. E., & Hannemann, T. (2015). Theory of mind, social desirability, and unlikely symptom reporting in offenders with and without psychopathy. *Journal of Nervous and Mental Disease*, *203*(8), 596–603. https://doi.org/10.1097/NMD.0000000000000335

*Nentjes, L., Bernstein, D. P., Arntz, A., van Breukelen, G., & Slaats, M. (2015). Examining the influence of psychopathy, hostility biases, and automatic processing on criminal offenders' theory of mind. *International Journal of Law and Psychiatry*, *38*, 92–99. https://doi.org/10.1016/j.ijlp.2015.01.012

*Newbury-Helps, J., Feigenbaum, J., & Fonagy, P. (2017). Offenders with antisocial personality disorder display more impairments in mentalizing. *Journal of Personality Disorders*, *31*(2), 232–255. https://doi.org/10.1521/pedipass:[_]2016_30_246

Oakley, B. M., Brewer, R., Bird, G., & Catmur, C. (2016). Theory of mind is not theory of emotion: A cautionary note on the Reading the Mind in the Eyes Test. *Journal of Abnormal Psychology*, *125*(6), 818–823. https://doi.org/10.1037/abn0000182

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, *308*(5719), 255–258. https://doi.org/10.1126/science.1107621

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

*Proctor, T., & Beail, N. (2007). Empathy and theory of mind in offenders with intellectual disability. *Journal of Intellectual and Developmental Disability*, *32*(2), 82–93. https://doi.org/10.1080/13668250701373331

Queirós, A., Fernandes, E., Reniers, R., Sampaio, A., Coutinho, J., & Seara-Cardoso, A. (2018). Psychometric properties of the questionnaire of cognitive and affective empathy in a Portuguese sample. *PLoS One*, *13*(6). https://doi.org/10.1371/journal.pone.0197755

Richell, R. A., Mitchell, D. V., Newman, C., Leonard, A., Baron-Cohen, S., & Blair, R. R. (2003). Theory of mind and psychopathy: Can psychopathic individuals read the "language of the eyes?" *Neuropsychologia*, *41*(5), 523–526. https://doi.org/10.1016/S0028-3932(02)00175-6

*Robinson, R., Roberts, W. L., Strayer, J., & Koopman, R. (2007). Empathy and emotional responsiveness in delinquent and non-delinquent adolescents. *Social Development*, *16*(3), 555–579. https://doi.org/10.1111/j.1467-9507.2007.00396.x

*Romero-Martínez, Á., Lila, M., Sariñana-González, P., González-Bono, E., & Moya-Albiol, L. (2013). High testosterone levels and sensitivity to acute stress in perpetrators of domestic violence with low cognitive flexibility and impairments in their emotional decoding process: A preliminary study. *Aggressive Behavior*, *39*(5), 355–369. https://doi.org/10.1002/ab.21490

Rubio-Fernández, P., & Geurts, B. (2013). How to pass the false-belief task before your fourth birthday. *Psychological Science*, *24*(1), 27–33. https://doi.org/10.1177/0956797612447819

Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(5), 1255–1266. https://doi.org/10.1037/a0018729

*Schiffer, B., Pawliczek, C., Müller, B. W., Wiltfang, J., Brüne, M., Forsting, M., Gizewski, E. R., Leygraf, N., & Hodgins, S. (2017). Neural mechanisms underlying affective theory of mind in violent antisocial personality disorder and/or schizophrenia. *Schizophrenia Bulletin*, *43*(6), 1229–1239. https://doi.org/10.1093/schbul/sbx012

Schlegel, K., Grandjean, D., & Scherer, K. R. (2014). Introducing the Geneva Emotion Recognition Test: An example of Rasch-based test development. *Psychological Assessment*, *26*(2), 666–672. https://doi.org/10.1037/a0035246.supp (Supplemental)

*Schuler, M., Mohnke, S., Amelung, T., Dziobek, I., Lemme, B., Borchardt, V., Gerwinn, H., Kärgel, C., Kneer, J., Massau, C., Pohl, A., Tenbergen, G., Weiß, S., Wittfoth, M., Waller, L., Beier, K. M., Walter, M., Ponseti, J. . . . Walter, H. (2019). Empathy in pedophilia and sexual offending against children: A multifaceted approach. *Journal of Abnormal Psychology*, *128*(5), 453–464. https://doi.org/10.1037/abn0000412

Senju, A., Southgate, V., White, S., & Frit, U. (2009). Mindblind eyes: An absence of spontaneous theory of mind in Asperger syndrome. *Science*, *325*(5942), 883–885. https://doi.org/10.1126/science.1176170

Serin, R. C., Gobeil, R., & Preston, D. L. (2009). Evaluation of the persistently violent offender treatment program. *International Journal of Offender Therapy and Comparative Criminology*, *53*(1), 57–73. https://doi.org/10.1177/0306624X07313985

*Shamay-Tsoory, S. G., Harari, H., Aharon-Peretz, J., & Levkovitz, Y. (2010). The role of the orbitofrontal cortex in affective theory of mind deficits in criminal offenders with psychopathic tendencies. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*, *46*(5), 668–677. https://doi.org/10.1016/j.cortex.2009.04.008

Shamay-Tsoory, S. G., Tomer, R., Berger, B. D., Goldsher, D., & Aharon-Peretz, J. (2005). Impaired "affective theory of mind" is associated with right ventromedial prefrontal damage. *Cognitive and Behavioral Neurology*, *18*(1), 55–67. https://doi.org/10.1097/01.wnn.0000152228.90129.99

Shamay-Tsoory, S. G., Tomer, R., Yaniv, S., & Aharon-Peretz, J. (2002). Empathy deficits in Asperger syndrome: A cognitive profile. *Neurocase*, *8*(3), 245–252. https://doi.org/10.1093/neucas/8.3.245

Shatz, M., Diesendruck, G., Martinez-Beck, I., & Akar, D. (2003). The influence of language and socioeconomic status on children's understanding of false belief. *Developmental Psychology*, *39*(4), 717–729. https://doi.org/10.1037/0012-1649.39.4.717

Söderstrand, P., & Almkvist, O. (2012). Psychometric data on the eyes test, the faux pas test, and the Dewey Social Stories Test in a population-based Swedish adult sample. *Nordic Psychology*, *64*(1), 30–43.

*Spenser, K. A. (2017). *Socio-cognitive determinants of offending behaviour* [Unpublished doctoral dissertation]. Nottingham Trent University.

*Spenser, K. A., Betts, L. R., & Das Gupta, M. (2015). Deficits in theory of mind, empathic understanding and moral reasoning: A comparison between young offenders and non-offenders. *Psychology, Crime & Law*, *21*(7), 632–647. https://doi.org/10.1080/1068316X.2015.1028542

Spenser, K. A., Bull, R., Betts, L., & Winder, B. (2019). Executive functioning as a predictive measure of offending behavior. *Journal of Criminal Psychology*. https://doi.org/10.1108/jcp-07-2018-0032

Stotland, E. (1969). Exploratory investigations of empathy. *Advances in Experimental Social Psychology*, *4*, 271–314.

Strayer, J. (1993). Children's concordant emotions and cognitions in response to observed emotions. *Child Development*, *64*(1), 188–201. https://doi.org/10.2307/1131445

van Langen, M. M., Wissink, I. B., van Vugt, E. S., Van der Stouwe, T., & Stams, G. M. (2014). The relation between empathy and offending: A meta-analysis. *Aggression and Violent Behavior*, *19*(2), 179–189. https://doi.org/10.1016/j.avb.2014.02.003

Varga, E., Tenyi, T., Fekete, S., & Herold, R. (2008). The evaluation of mentalization deficit by the faux pas test in schizophrenia. *Neuropsychopharmacology Hungarica*, *10* (2), 75–80.

Vollm, B. A., Taylor, A. N., Richardson, P., Corcoran, R., Stirling, J., McKie, S., Deakin, J. F. W., & Elliott, R. (2006). Neuronal correlates of theory of mind and empathy: A functional magnetic resonance imaging study in a nonverbal task. *Neuroimage*, *29*(1), 90–98.

Wang, Z., & Wang, L. (2015). The mind and heart of the social child: Developing the empathy and theory of mind scale. *Child Development Research*, 1–8. https://doi.org/10.1155/2015/171304. Retrieved from https://www.hindawi.com/journals/cdr/2015/171304/

Ward, T., Keenan, T., & Hudson, S. M. (2000). Understanding cognitive, affective, and intimacy deficits in sexual offenders: A developmental perspective. *Aggression and Violent Behavior*, *5*(1), 41–62. https://doi.org/10.1016/S1359-1789(98)00025-1

Warnell, K. R., & Redcay, E. (2019). Minimal coherence among varied theory of mind measures in childhood and adulthood. *Cognition*, *191*, 103997.

Warrier, V., Bethlehem, R.A., & Baron-Cohen, S. (2017). The "Reading the Mind in the Eyes" Test (RMET). In V. Zeigler-Hill & T. Shackelford (Eds.), *Encyclopedia of personality and individual differences*. Springer, Cham. http://doi-org-443.webvpn.fjmu.edu.cn/10.1007/978-3-319-28099-8_549-1

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.

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

*Winter, K., Spengler, S., Bermpohl, F., Singer, T., & Kanske, P. (2017). Social cognition in aggressive offenders: Impaired empathy, but intact theory of mind. *Scientific Reports*, *7*(670), 1–10. https://doi.org/10.1038/s41598-017-00745-0.

*Woodbury-Smith, M. R., Clare, I. H., Holland, A. J., Kearns, A., Staufenberg, E., & Watson, P. (2005). A case-control study of offenders with high functioning autistic spectrum disorders. *Journal of Forensic Psychiatry & Psychology*, *16*(4), 747–763. https://doi.org/10.1080/14789940500302554

Zhu, C., Lee, T. C., Li, X., Jing, S., Wang, Y., & Wang, K. (2007). Impairments of social cues recognition and social functioning in Chinese people with schizophrenia. *Psychiatry and Clinical Neurosciences*, *61*(2), 149–158. https://doi.org/10.1111/j.1440-1819.2007.016

Author Biographies

Nilda Karoğlu, MSc, is a PhD candidate in forensic psychology at the School of Psychology, University of Kent in the United Kingdom.

Heather J. Ferguson, PhD, is a professor of psychology at the School of Psychology, University of Kent in the United Kingdom.

Caoilte Ó Ciardha, PhD, is a senior lecturer in forensic psychology at the School of Psychology, University of Kent in the United Kingdom.