

No links between genetic variation and developing theory of mind: A preregistered replication attempt of candidate gene studies

Timm Opitz^{1,2} | Tobias Schuwerk³  | Markus Paulus³ | Daniela Kloos³ |
Christopher Osterhaus³  | Klaus-Peter Lesch^{4,5,6}  | Beate Sodian³

¹Max Planck Institute for Innovation and Competition, Munich, Germany

²Department of Economics, Ludwig-Maximilians-Universität, Munich, Germany

³Department of Psychology, Ludwig-Maximilians-Universität, Munich, Germany

⁴Division of Molecular Psychiatry, Center of Mental Health, University of Würzburg, Würzburg, Germany

⁵Laboratory of Psychiatric Neurobiology, Institute of Molecular Medicine, I. M. Sechenov First Moscow State Medical University, Moscow, Russia

⁶Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience (MHeNS), Maastricht University, Maastricht, The Netherlands

Correspondence

Tobias Schuwerk, Department of Psychology, Ludwig-Maximilians-Universität München, Leopoldstr. 13, 80802 Munich, Germany.
Email: tobias.schuwerk@psy.lmu.de

Funding information

Deutsche Forschungsgemeinschaft, Grant/Award Number: SO 213/27-1,2,3

Abstract

Genetic variability is being discussed as a source of inter-individual differences in Theory of Mind development. Previous studies documented an association between variations in DRD4 VNTR 48 bp, OXTR rs53576, COMT rs4680, and Theory of Mind task performance. As empirical evidence on these associations is sparse, we conducted a preregistered replication attempt of a study reporting a link between DRD4 VNTR 48 bp and false belief understanding in 50-month-old children [Lackner, C., Sabbagh, M. A., Hallinan, E., Liu, X., & Holden, J. J. (2012). *Developmental Science*, 15(2), 272–280.]. Additionally, we attempted a replication of studies on the role of OXTR rs53576 and COMT rs4680 in Theory of Mind. In both replication attempts, we did not find any evidence for associations between the sampled genetic markers and Theory of Mind ability in a series of analyses. Extending the replication attempt of Lackner et al., we employed longitudinal data from several tasks and measurement points, which allowed us to run follow-up robustness checks with more reliable scores. These extensive analyses corroborated our null finding. This comprehensive non-replication is important to balance current research on genetic markers of Theory of Mind. In a combined evaluation of our own and previous studies, we point to substantial methodological issues that research on the genetic basis of Theory of Mind development faces. We conclude that these limitations currently prevent firm conclusions on genetic influences on Theory of Mind development.

KEYWORDS

COMT, DRD4, false belief, OXTR, theory of mind

1 | INTRODUCTION

The ability to impute mental states to others and oneself, termed Theory of Mind, is crucial for social interactions. Understanding that

behavior can be guided based on a false belief about the world is considered a hallmark of Theory of Mind acquisition (Wimmer & Perner, 1993). Although the ability to attribute false beliefs to others at around 4 years of age constitutes a universal developmental

Timm Opitz and Tobias Schuwerk should be considered joint first author.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2021 The Authors. *Developmental Science* published by John Wiley & Sons Ltd.

milestone, there are substantial inter-individual differences both during development (Osterhaus et al., 2016; Wellman et al., 2001) and in mature Theory of Mind skills (e.g., Dodell-Feder et al., 2013). Studying the origins of these differences promises to advance theoretical models of Theory of Mind and its development.

As one source of inter-individual differences in Theory of Mind development, researchers discuss genetic variability (Hughes & Plomin, 2000). For instance, a twin study by Hughes and Cutting (1999) provided evidence for heritability of Theory of Mind abilities. Results from Theory of Mind task performance in identical and fraternal 42-month-old twins suggested that individual differences were attributable to genetic factors. By contrast, a later study with a substantially larger sample of 60-month-olds suggested only little genetic impact and pointed to a strong influence of environmental factors (Hughes et al., 2005), which is in line with theoretical accounts that emphasize the social origins of Theory of Mind (Carpendale & Lewis, 2004; Heyes, 2018). In sum, to date available evidence from twin studies is inconclusive and it remains unclear how genetic factors contribute to Theory of Mind development.

Recently, researchers started addressing the molecular basis of social cognition and social behavior by investigating the link between measured behavior in controlled environments and specific genetic markers that are supposed to relate to the tested behavior (Ebstein et al., 2010; Skuse & Gallagher, 2009). For example, Lackner et al. (2012) found that preschoolers with two shorter alleles of the dopamine D4 receptor gene (DRD4 VNTR 48 bp) performed better in a false belief task than preschoolers with one or two longer alleles, independent of their performance in an executive functions task. Such findings opened avenues to speculate about neurogenetic pathways or gene \times environment interactions explaining Theory of Mind development. Other studies suggested links between different allelic variants of COMT rs4680 (linked to the regulation of the level of synaptic dopamine) and OXTR rs53576 (an oxytocin receptor genotype) and the understanding of false beliefs (Wu & Su, 2015; Xia et al., 2012), as well as attributing mental states to facial expressions (Reading the Eyes in the Mind task; Lucht et al., 2013; Rodrigues et al. 2009).

This line of research is novel and evidence on the molecular basis of Theory of Mind is sparse (for quantitative review of the link between DRD4 and the broader category social/emotional development see Pappa et al., 2015). Furthermore, previous studies combining children's behavioral task performance with genotyping of candidate genetic markers face methodological issues. Due to its novelty, the field is lacking procedural and analytical standards which open up many degrees of freedom for selecting, preprocessing, and analyzing the data (cf., Simmons et al., 2011). Moreover, the sample sizes are small, typically ranging from 48 to 350 children (e.g., Lackner et al., 2012; Grossmann et al., 2013; Knafo et al., 2011; Wade et al., 2014; Wu & Su, 2015). This substantially limits the power to find associations between genetic variants and behavior as subgroups with the same genotype can become quite small.

One important way to address these issues is the replication of previously reported links between specific genetic markers and

Research Highlights

- We investigated genetic variability as one source of inter-individual differences in Theory of Mind development
- In our replication study, we found no evidence for associations between sampled genetic markers and Theory of Mind ability in a series of analyses
- We point out substantial methodological issues that research on the genetic basis of Theory of Mind development faces

performance in specific Theory of Mind tasks. Moreover, preregistering hypotheses and analyses before evaluating the outcome substantially reduces problematic methodological degrees of freedom in this new field of research (Nosek et al., 2018).

In the present preregistered study, we tested the replicability of the association between candidate genetic markers and Theory of Mind. We used an available longitudinal data set from a project on social cognitive development. More specifically, we were in the favorable position to test the influence of the dopamine system-related marker DRD4 VNTR 48 bp on Theory of Mind in preschoolers previously reported by Lackner et al. (2012), as we had data from a closely comparable battery of Theory of Mind and executive functions tasks in a sample of the same age. Thanks to this original study, we were able to substantially reduce degrees of freedom by concentrating on a specific prediction for a circumscribed set of variables. The inclusion of OXTR rs53576 was based on additional empirical evidence linking different allelic variants to Theory of Mind and related abilities (Lucht et al., 2013; Rodrigues et al. 2009; Wu & Su, 2015).

In our sample of 50- to 70-month-old children, we analyzed the association between allelic variations of DRD4 VNTR 48 bp, COMT rs4680, and OXTR rs53576 and the performance in false belief tasks from the Theory of Mind scale by Wellman and Liu (2004), controlling for the level of executive functioning assessed via a behavioral inhibition task (Strommen, 1973). According to Lackner et al. (2012), we predicted to find that children with two short alleles in DRD4 VNTR 48 bp (<6 repeats) outperform children with one or two long alleles (≥ 6 repeats) in the false belief tasks. Lackner et al. (2012) did not report differences based on the allelic distribution of COMT rs4680 (which is in line with the findings in an adult sample by Xia et al., 2012). We tested the corresponding null hypothesis.

For OXTR rs53576, previous literature reported evidence on the association between G alleles and empathy (e.g., Gong et al., 2017), the attribution of mental states to facial expressions (e.g. Lucht et al. 2013; Rodrigues et al., 2009; Weisman et al., 2015) as well as Theory of Mind (Wu & Su, 2015). We therefore expected to find that carriers and non-carriers of G alleles in OXTR rs53576 differ in their Theory of Mind performance. In our preregistered analyses, we assessed the effect of the genetic markers on Theory of Mind development around the age of 50 months.

TABLE 1 Descriptive statistics for false belief tasks. Proportions of children passing the respective task across measurement points

Task	Outcome	Passing		
		50 months	60 months	70 months
Content FB	{0,1}	0.39	0.81	0.84
Location FB	{0,1}	0.42	0.76	0.88

Our rich longitudinal data set allowed us to run additional analyses with aggregated task outcomes over several measurement points to get more reliable measures for our main outcome variables. This enabled us to test for the influence of genetic variation on Theory of Mind development more extensively than the original studies did. In addition, we were able to conduct several follow-up analyses to corroborate the robustness of our confirmatory findings by exploiting additional data to compute scaled Theory of Mind scores based on a larger Theory of Mind battery.

2 | METHODS

The study was preregistered at the Open Science Framework (OSF). The preregistration and data of this study are available at <https://osf.io/vujgs>. Following best practice recommendations, we preregistered all methods used in the confirmatory analysis. This includes the determination of our sample size, the calculation of the dependent and independent variables, all data exclusions and manipulations as well as the statistical methods we applied. We specified which behavioral tasks we included at which age, how we coded the task outcomes, how we aggregated the scores, and how we grouped children based on variants of the selected genetic markers. We highlight all methods and analyses that go beyond the preregistration as being exploratory.

2.1 | Participants

The final sample size for this study was $N = 80$ children (34 girls and 46 boys). This sample came from a larger longitudinal study on social cognitive development from infancy to middle childhood. Overall, 155 monolingual children (68 girls and 87 boys) from the lower to upper middle class in an urban area of Germany participated in this study. At the last measurement point, we obtained the reported genetic markers with the goal to replicate Lackner et al. (2012) and other relevant studies. The present sample size is thus determined by the sample size of the larger longitudinal study and is a result of non-participation in the respective tasks over the course of the longitudinal study and/or missing parental consent for genotyping. The relevant tasks were administered at 50 months ($M = 50.58$, $SD = 0.85$), 60 months ($M = 60.69$, $SD = 0.69$), and 70 months of age ($M = 70.34$, $SD = 0.51$). We decided to exclude data from one child and restrict the analyses to a mono-ethnic European sample to avoid problems of varying allelic frequencies and functionalities across ethnic groups (Lichter et al., 1993). Our sample size is

comparable with that of the original study by Lackner et al. (2012) with 73 participants (in their sample, two children with a different ethnicity remained included). The children were recruited via birth records and received gifts for their participation. Their caregivers gave informed written consent. The families received monetary compensation for participating. The ethics committee of the Department of Psychology and Education of LMU Munich approved this study. All procedures performed in the reported experiments were in accordance with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

2.2 | Measures and materials

2.2.1 | Behavioral tasks

Corresponding to Lackner et al. (2012), we measured false belief understanding around the age of 50 months and controlled for executive functions. Although the *content false belief* task is identical, we used a slightly different paradigm to assess *location false belief* and did not collect data on the *deceptive pointing* task that Lackner et al. employed in addition. A further difference to the original study is the measurement of executive functioning. The Theory of Mind tasks were part of the Theory of Mind scale by Wellman and Liu (2004; German version by Kristen et al., 2006).

In the *content false belief* task, the child was asked to judge another person's false belief about what was in a container (Smarties box) when the child knew what was in it (a piglet). The child was shown a Smarties box and was asked what she or he thought was inside. Then the box was opened, revealing a piglet inside. After closing the box and assuring that the child knew the content, a Playmobil® figure was introduced by telling the child that the figure had never looked inside the box. After that, the test question ("What does Lukas think is inside the box?") was asked, followed by a control question ("Has he ever looked inside?") to ensure understanding of the situation.

The *location false belief* task required the child to judge whether someone would search for an object based on a false belief. The child was told that a Playmobil® figure was looking for his mittens that could be either in the closet or in the backpack (which were presented on colored drawings to the child). The child was informed that the mittens were in the backpack, but that the Playmobil® figure (Paul) believed that they were in the closet. Following this, the test question ("Where will Paul look for his mittens?") was asked and understanding of the situation ensured by a control question ("Where are they really?").

We assessed whether the child answered the respective task and the control question correctly or not. Thus, we had a binary outcome for each task. Children participated in the *content* and the *location false belief* task at the age of 50, 60, and 70 months. We summarize the task outcomes in Table 1 and display the cross-correlations between the task outcomes in Section S6 of the supplemental material.

To control for executive functions, we included the Simon Says task (Kloo & Sodian, 2017; Strommen, 1973) in which children are asked to imitate the experimenter's actions (but only if the phrase "Simon says" precedes the command). Children performed the task for 20 trials (each coded from 0 to 2 according to whether full, partial, or no movements were performed on a non-Simon trial) at the age of 60 months. Notably, this differs from the procedure of Lackner et al. (2012) because they more extensively administered executive functions using a battery of several tasks. Further, Lackner et al. assessed executive functions at around 50 months of age.

In an exploratory analysis, we investigated whether there was an association between genetic markers and three scaled Theory of Mind scores. The scaled scores (at 50, 60, and 70 months) were based on a larger Theory of Mind battery, and hence gave a more reliable estimate of individual differences. At 50 months, the Theory of Mind battery comprised six tasks from the Theory of Mind scale (*diverse desires*, *diverse beliefs*, *knowledge access*, *content false belief*, *location false belief*, and *real-apparent emotion*; Wellman & Liu, 2004), two false belief tasks in a morally relevant context (Killen et al., 2011), a task on children's understanding of mental verbs (Moore et al., 1989), and a task to assess second-order false belief competence (Coull et al., 2006). At 60 and 70 months, the three least difficult tasks from the Theory of Mind scale were omitted (*diverse desires*, *diverse beliefs*, and *knowledge access*) and instead two Strange Stories (Happé, 1994) were added (60 months, Lie and Joke; 70 months, Lie and Metaphor). Additionally, the scaled Theory of Mind score at 70 months included an assessment of children's metacognitive understanding of their own ignorance (Rohwer et al., 2012). As longitudinal stability was found between the three scaled Theory of Mind scores (Osterhaus et al., 2019), we computed a single composite score that was based on the three assessment points.

2.2.2 | Genotyping

The testing was limited to three genetic markers and taken through a saliva sample. We standardized testing and processing of the sample. Genomic DNA was isolated from the saliva sample. Section S1 of the supplemental materials outlines the procedure in more detail. The two SNPs COMT rs4680 and OXTR rs53576 are autosomal (A/A, A/G, G/G). By contrast, DRD4 VNTR 48 bp can have up to 20 genetic variants. Therefore, we dichotomized it following the procedure of Lackner et al. (2012; see also Pappa et al., 2015).¹ We differentiated between carriers of at least one long allele (≥ 6 repeats) and the rest of the population. We treated all genetic markers as

TABLE 2 Distribution of allele variations and gender

Gene	Variant	Total N	Males	Females
COMT_rs4680	A/A	18	8	10
	A/G	50	32	18
	G/G	12	6	6
OXTR_rs53576	A/G	3	2	1
	A/G	42	25	17
	G/G	35	19	16
DRD4	all ≤ 5 repeat	57	32	25
	≥ 6 repeat	23	14	9

categorical variables. In Table 2, we display the sample distribution of the allelic variants.

2.3 | Procedure

The behavioral data came from a larger longitudinal study on social cognitive development from infancy to middle childhood (6–99 months), carried out at LMU Munich. For all behavioral tasks, 30% of the data was re-coded by a second coder revealing high interrater reliabilities (for details see e.g., Sodian & Kristen-Antonow, 2015). The genetic markers were sampled at the penultimate measurement point.

2.4 | Data analysis

For the confirmatory analysis of genetic influences on Theory of Mind, we used one (multiple) linear model for each of the two Theory of Mind tasks and a composite false belief score. As independent variables, we included the three genetic markers. In a second step, we added executive functioning to the model. To be consistent with the ANCOVA analysis by Lackner et al. (2012), we resorted to multiple linear regressions. We included COMT rs4680 as a control variable but refrain from interpreting effects causally in the confirmatory analysis due to deviations from Hardy-Weinberg equilibrium. The rationale for using an aggregated score was as follows: The problem especially developmental psychologists encounter is that performance in a single task at one point in time can be influenced by several unspecific factors (e.g., fuzziness), which results in measurement error or missing data. To minimize this potential source of error, we used the aggregated Theory of Mind scores built from several tasks and several measurements. Here, this approach is particularly suitable because Osterhaus et al. (2019) showed that individual differences in overall Theory of Mind performance were stable across the three measurement time points. In the exploratory analysis, we re-estimated these models with different dependent variables. The analyses were performed using Stata, Version 15.1.

3 | RESULTS

3.1 | Confirmatory analysis

In our preregistered analysis, we regressed average individual Theory of Mind scores across the three measurement points at 50, 60, and 70 months on the three genetic markers. Additionally, we controlled for executive functioning at the age of 60 months. We assessed the influence on Theory of Mind independently for the *content false belief* and the *location false belief* tasks. Additionally, we aggregated both measures to a composite false belief score and re-ran the analysis (with and without controlling for executive functioning). With this procedure, we followed the rationale presented by Lackner et al. (2012). They reported a low internal consistency of Cronbach's alpha of 0.095 of the three employed false belief tasks (*content false belief*, *location false belief*, *deceptive pointing task*). Lackner et al. excluded the *content false belief* task, which led to a Cronbach's alpha of 0.43. In the current study, we found a Cronbach's alpha of 0.60 of the two false belief tasks. Although this internal consistency is still considerably low, it was higher than that of the original study after excluding the *content false belief* task. Therefore, as preregistered, we ran the additional analysis using this compound score which we calculated by taking the mean score (instead of first calculating z-scores like Lackner et al., 2012). Figure 1 and Table 3 show the results of these analyses (in S2 of the supplemental material, we outline the construction of the scores in detail and display summary statistics in S3).

Our confirmatory analyses provided no evidence for genetic influences on Theory of Mind performance (Table 3). We neither detected effects on performance in the *content*, nor in the *location false belief* tasks. The point estimates were close to zero. These results remained stable when controlling for executive functioning. While executive functioning itself was strongly associated with Theory of Mind performance, the coefficients of the genetic markers remained statistically insignificant and close to zero. The patterns for the single domains also

held when aggregating them to a composite score. In the Supplemental Material, we report the detailed correlations between performance in the executive function task and false belief task performance.

In sum, our main analyses did not provide any evidence for an effect of the genetic markers on Theory of Mind performance. Yet, due to the limited sample size, we were unable to rule out even modest effects. While our (95%) confidence interval ruled out effect sizes above $\eta^2 = 0.09$ and is thus inconsistent with the larger effect size in Lackner et al. (2012), it is still consistent with Theory of Mind abilities being up to 11.8 (23.1) percent lower (higher) for carriers of long alleles in DRD4 VNTR 48 bp repeat. Therefore, in addition to the preregistered main specification, we ran additional exploratory analyses to check for the robustness of our null findings.

3.2 | Exploratory analyses

3.2.1 | Robustness checks

In a first step, we ran analyses even closer to those of Lackner et al. (2012) to allow for a better comparability of the results. In a second step, we exploited additional available data from the longitudinal data set to check for the robustness of our results. Our confirmatory analyses relied on Theory of Mind scores averaged over tasks administered longitudinally at 50, 60, and 70 months of age.² While this resulted in a more reliable measure of Theory of Mind abilities, we also slightly deviated from the original study by Lackner et al. (2012) with a mean age of 47.25 months ($SD = 3.46$). Therefore, we re-ran our main analysis and restricted the analysis to the measurement point at 50 months ($M = 50.58$, $SD = 0.85$). As detailed in Table 4, this analysis corroborated our previous null-finding at the pre-specified significance level of $p \leq 0.05$.³ We detected one marginally significant ($p = 0.086$) association between performance in the *location false belief* task and two variants of COMT rs4680 when controlling for executive function.

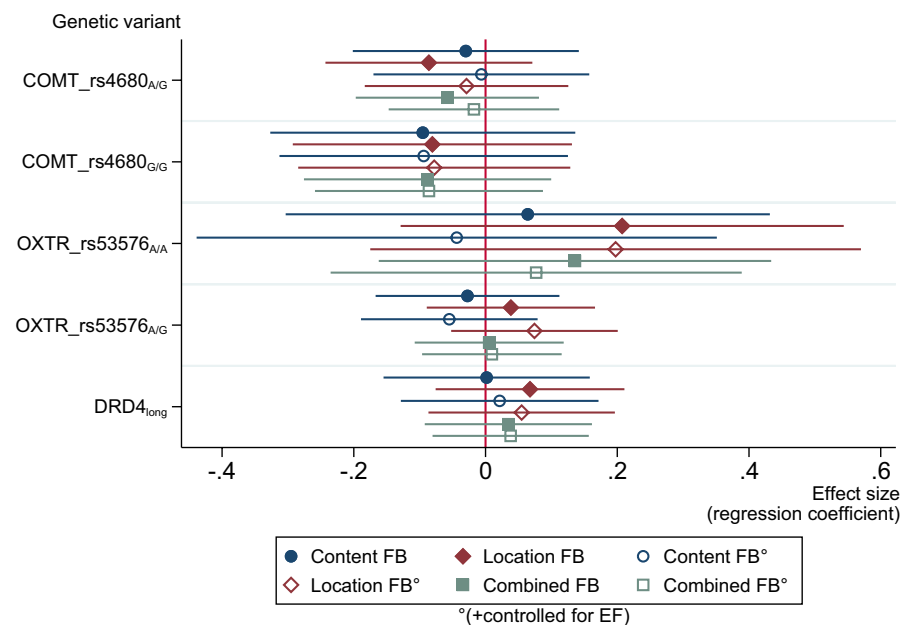


FIGURE 1 Regression coefficient plot (95% confidence intervals) for genetic markers at 50–70 months. The regressions with hollow markers (also indicated with the symbol °) include executive functions as a control variable. Notes: FB, false belief; EF, executive functions

TABLE 3 Regression analysis of genetic markers on Theory of Mind abilities at 50–70 months

	(1)	(2)	(3)	(4)	(5)	(6)
	Content FB	Location FB	Content FB	Location FB	Combined FB	Combined FB
COMT_rs4680 _{A/G}	-0.030 (0.086)	-0.086 (0.079)	-0.006 (0.082)	-0.029 (0.077)	-0.058 (0.070)	-0.018 (0.065)
COMT_rs4680 _{G/G}	-0.095 (0.116)	-0.081 (0.106)	-0.094 (0.110)	-0.078 (0.103)	-0.088 (0.094)	-0.086 (0.087)
OXTR_rs53576 _{A/A}	0.064 (0.184)	0.207 (0.169)	-0.044 (0.198)	0.197 (0.186)	0.136 (0.150)	0.077 (0.156)
OXTR_rs53576 _{A/G}	-0.027 (0.070)	0.038 (0.064)	-0.055 (0.067)	0.074 (0.063)	0.006 (0.057)	0.010 (0.053)
DRD4 _{long}	0.002 (0.079)	0.068 (0.072)	0.021 (0.075)	0.055 (0.071)	0.035 (0.064)	0.038 (0.059)
EF			0.017*** (0.005)	0.012** (0.005)		0.014*** (0.004)
Constant	0.724*** (0.083)	0.710*** (0.076)	0.620*** (0.087)	0.584*** (0.082)	0.717*** (0.067)	0.602*** (0.069)
Observations	80	80	71	71	80	71
R-squared	0.018	0.055	0.162	0.118	0.033	0.172

Notes: Standard errors in parentheses.

The reference category of OXTR_rs53576 is adjusted due to a low cell frequency in OXTR_rs53576_{A/A}.

Columns (1), (2), and (5) only include the genetic markers as explanatory variables; columns (3), (4), and (6) also control for executive functions. Executive functions are measured by *Simon Says* at the age of 60 months.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

This result was neither stable across specifications nor measurements. Again, due to deviations from the Hardy–Weinberg equilibrium, we refrained from interpreting the estimate causally. The point estimates for DRD4 VNTR 48 bp remained close to zero and the association between executive functioning and Theory of Mind ability prevailed.⁴

Additionally, we report the partial correlation between DRD4 VNTR 48 bp and Theory of Mind performance to rule out that correlations between variants of different genetic markers drove our null result for DRD4 VNTR 48 bp in the regression framework. Also without controlling for the additional genetic markers (and executive functioning), we found no association between allelic variants of DRD4 VNTR 48 bp and Theory of Mind performance. Table 5 reports the results from the partial analysis using *t*-tests for the comparison of means. Pearson's correlations close to zero ($r = [-0.026, 0.111]$) for the different outcomes of the main analysis corroborated the null findings. Additionally, unlike Lackner et al. (2012; but see Pappa et al., 2015), we did not find any evidence for an association between DRD4 VNTR 48 bp and executive functioning ($r = -0.046$).

3.2.2 | Genetic impact on scaled theory of mind score

In a follow-up analysis, we used additionally available data from the longitudinal data set. To rule out that our null findings were a

consequence of poor estimates of children's Theory of Mind ability, we investigated the associations between the genetic markers and scaled Theory of Mind scores (weighted likelihood estimates from Rasch analysis at 50 months and averaged across all three measurement points; see above). Again, there was no significant association at the pre-specified significance level of $p \leq 0.05$ between genetic markers and either the scaled Theory of Mind score at age 50 or the average score (Table 6). The analysis showed marginally significant associations between the scaled Theory of Mind score averaged across all three measurement points and OXTR rs53576 when controlling for executive functions. For DRD4 VNTR 48 bp, estimates remained consistently insignificant at any conventional level.

3.2.3 | Bayesian linear regression

Following a reviewer's suggestion, we additionally translated our main confirmatory analysis into a Bayesian framework to estimate evidence in favor of the null hypothesis that there is no influence of the sampled candidate genes. Corroborating our frequentist analysis, we found moderate evidence for the null hypothesis that DRD4 is not predictive for false belief understanding at the age of 50–70 months (BayesFactor₀₁ = 3.398). We present the detailed results of this exploratory Bayesian analysis in S7 of the supplemental material.

TABLE 4 Regression analysis of genetic markers on Theory of Mind abilities at 50 months.

	(1)	(2)	(3)	(4)	(5)	(6)
	Content FB	Location FB	Content FB	Location FB	Combined FB	Combined FB
COMT_rs4680 _{A/G}	0.014 (0.147)	-0.083 (0.146)	-0.011 (0.145)	-0.007 (0.155)	-0.004 (0.114)	0.009 (0.116)
COMT_rs4680 _{G/G}	-0.050 (0.196)	0.320 (0.192)	-0.140 (0.192)	0.353 [*] (0.202)	0.160 (0.152)	0.115 (0.153)
OXTR_rs53576 _{A/A}	-0.087 (0.312)	0.342 (0.300)	-0.412 (0.346)	0.244 (0.362)	0.136 (0.241)	-0.090 (0.276)
OXTR_rs53576 _{A/G}	-0.026 (0.120)	0.069 (0.118)	-0.036 (0.120)	0.079 (0.128)	0.023 (0.093)	0.01 (0.095)
DRD4 _{long}	-0.025 (0.136)	0.055 (0.134)	-0.007 (0.136)	0.087 (0.143)	0.003 (0.105)	0.036 (0.108)
EF			0.030 ^{***} (0.009)	0.011 (0.010)		0.021 ^{***} (0.007)
Constant	0.415 ^{***} (0.140)	0.352 ^{**} (0.142)	0.253 (0.152)	0.201 (0.168)	0.365 ^{***} (0.108)	0.219 [*] (0.121)
Observations	79	74	70	66	79	70
R-squared	0.005	0.116	0.172	0.117	0.028	0.128

Notes: Standard errors in parentheses.

The reference category of OXTR_rs53576 is adjusted due to a low cell frequency in OXTR_rs53576_{A/A}.

Columns (1), (2), and (5) only include the genetic markers as explanatory variables; columns (3), (4), and (6) also control for executive functions. Executive functions are measured by *Simon Says* at the age of 60 months.

*** $p < 0.01$,

** $p < 0.05$,

* $p < 0.1$

TABLE 5 Partial analysis of links between DRD4 VNTR 48 bp and false belief task performance at 50–70 months. Columns 1 and 2 report means of the scores, split by the genetic variant of DRD4. Column 3 shows the p -values for a t -test, testing the null hypothesis that the difference is zero

	(1)	(2)	(3)
	DRD4 _{short}	DRD4 _{long}	p -value
Contents FB	0.684	0.667	0.810
Location FB	0.673	0.739	0.327
Combined FB	0.678	0.703	0.680

4 | DISCUSSION

Our study aimed at replicating and extending the limited evidence on molecular genetic influences on Theory of Mind development. We found no association between either of the sampled genetic markers (DRD4 VNTR 48 bp, COMT rs4680, OXTR rs53576) on Theory of Mind ability at the age of 50–70 months. This null finding was corroborated by additional exploratory analyses. The results stand in contrast to previous work documenting associations between DRD4 VNTR

48 bp, OXTR rs53576 and Theory of Mind ability. It is important to note that Lackner et al.'s predictions were specifically for 50-month-olds and not older children. Finding no links between DRD4 VNTR 48 bp and Theory of Mind task performance in older children is thus not inconsistent with their claim that this genetic marker promotes the trajectory of Theory of Mind development at around 50 months. Yet, in our follow-up analysis focusing only on 50-month-olds, we also did not find that DRD4 VNTR 48 bp relates to Theory of Mind performance. While we refrain from interpreting the coefficients of COMT rs4680 due to deviations from Hardy-Weinberg, our null result for this marker is in line with the findings of Lackner et al. (2012) and Xia et al. (2012). Further, our results do not corroborate the finding of Wu and Su (2015), who reported an association between OXTR rs53576 and Theory of Mind ability. The two marginally significant associations between OXTR rs53576 we find in Table 6 (one of them based on $n = 2$ children with the specific genotype) go –if anything– in the opposite direction as reported in Wu and Su (2015). Overall, our replication attempt of Lackner et al. (2012), our replication attempts of studies on the influence of OXTR rs53576 and COMT rs4680, and our follow-up analyses relying on robust task scores and several measurement points did not provide any evidence for associations between the three sampled genetic markers and Theory of Mind ability.



	(1)	(2)	(3)	(4)
	Scaled ToM (50–70)	Scaled ToM (50–70)	Scaled ToM (50)	Scaled ToM (50)
COMT_rs4680 _{A/G}	0.143 (0.296)	0.013 (0.294)	0.268 (0.375)	0.117 (0.356)
COMT_rs4680 _{G/G}	0.050 (0.391)	-0.145 (0.385)	0.158 (0.494)	-0.224 (0.467)
OXTR_rs53576 _{A/A}	1.132 (0.726)	1.245 [†] (0.682)	1.116 (0.919)	1.167 (0.828)
OXTR_rs53576 _{A/G}	0.324 (0.236)	0.463 [†] (0.236)	0.351 (0.298)	0.421 (0.287)
DRD4 _{long}	0.016 (0.262)	-0.034 (0.262)	-0.303 (0.331)	-0.232 (0.318)
EF		0.070 ^{***} (0.019)		0.073 ^{***} (0.023)
Constant	-0.175 (0.281)	-0.612 ^{**} (0.305)	-0.952 ^{***} (0.355)	-1.328 ^{***} (0.370)
Observations	77	69	77	69
R-squared	0.056	0.248	0.059	0.203

TABLE 6 Regression analysis of genetic markers on scaled Theory of Mind scores

Notes: Standard errors in parentheses.

The reference category of OXTR_rs53576 is adjusted due to a low cell frequency in OXTR_rs53576_{A/A}.

Columns (1) and (2) report scaled scores averaged across measures with 50, 60, and 70 months, while (3) and (4) report the score with 50 months. Columns (1) + (2) only include the genetic markers as explanatory variables; columns (3) + (4) also control for executive functions. Executive functions are measured by Simon Says at the age of 60 months.

*** $p < 0.01$,

** $p < 0.05$,

* $p < 0.1$

There are several explanations of why we found no evidence for the association between DRD4 VN TR 48 bp and Theory of Mind task performance reported by Lackner et al. (2012). In the remainder, we discuss the two most extreme interpretations.

One possible interpretation is that the results of the present study are false-negative due to methodological differences from the original study. Le Bel et al. (2018) offered a classification tool to evaluate the closeness of replications. According to their framework, we would categorize our replication of Lackner et al. (2012) as “close direct replication”, despite the methodological differences: We addressed the same effect and hypothesis, targeted the same independent and dependent variable construct, the same population (i.e., age group), and used the same independent and dependent operationalizations.

However, even when using such a classification tool, some ambiguity remains. Specifically, one could argue that our operationalization of the construct of interest (false belief reasoning) differed from the original study. First, our content false belief task was identical with the one used by Lackner et al., but Lackner et al. excluded it from their analyses because it decreased the internal consistency of their false belief measures. Second, we had not employed a deceptive pointing task. Third, the false belief about location task we used

was a slightly different operationalization of the same task type. Lackner et al. used a task in which the protagonist's belief about the current location of a critical object had to be inferred (Wimmer & Perner, 1993), whereas in the present task (Wellman & Liu, 2004) the story character's belief about the location of an object was explicitly stated. Yet, these two tasks, albeit differing in superficial task demands, have been shown to be of similar difficulty and children's performance in both tasks is consequently highly comparable (Wellman & Bartsch, 1988). In sum, we employed identical or highly comparable false belief tasks to measure the same construct of interest like Lackner et al. We thus maintain that we conducted a close and direct replication attempt.

Nosek and Errington (2020) recently offered a definition of replication that rather focuses on the interpretation of the results. They suggested that a replication is defined by two criteria: First, results consistent with the original claim increase confidence in this claim. Second, inconsistent results decrease confidence in the original claim. A study meeting the first but failing the second criterion is a “generalizability test”. We are convinced that our conceptualization and operationalization is close enough to Lackner et al. to provide diagnostic evidence about their prior claim in the case consistent and inconsistent results. In our view it is therefore very unlikely that our



findings are false negatives due to methodological differences. Our failed replication attempt shows that the reliability of the original findings is significantly more constrained than previously assumed. Future replication studies will show whether confidence in the original claim can be reinstated, or whether further replication failures will gradually render this claim irrelevant.

An advantage of our study is that we could rely on findings from several tasks at several measurement point in our longitudinal data set. Our further analyses based on this data strongly suggest that our results are not driven by noisy measurement of the behavioral tasks. The theory-compliant association between performance in our Theory of Mind and executive functioning tasks further points to the accuracy of our measures (Carlson & Moses, 2001; Devine & Hughes, 2014).

A circumstance that might speak for the interpretation that our findings are false-negative is the small sample size of our study. It could have been under-powered to detect the effects reported by Lackner et al. (2012), as replication attempts are advised to have substantially larger sample sizes than the original study (Simonsohn, 2015). We were, however, able to run a comprehensive replication attempt of some of the leading papers in our field by making use of two alternative ways to increase statistical power. First, in our confirmatory analysis, we aggregated Theory of Mind scores across three measurement points at 50, 60, and 70 months, a repeated measures design that substantially increases statistical power (e.g., Vickers, 2003). Second, in our exploratory analyses, we aggregated across not only the three measurement points but also across several Theory of Mind tasks to obtain an even more precise estimate of Theory of Mind competence (cf., McClelland, 2000). The fact that the series of analyses we conducted did not produce noisy results but consistently converged on the null finding of no association between the sampled genetic markers and Theory of Mind abilities supports the argument that our approach resulted in meaningful findings.

An alternative conclusion is that the results of our study are true-negative. The combined inspection of our and previous work on the links between genetic markers and Theory of Minds abilities leads us to favor this interpretation. The researcher's degrees of freedom in this kind of studies are unreasonably large. Numerous arbitrary methodological analytical decisions have to be made. For example, genetic variants are not straightforward to code. As already pointed out by Pappa et al. (2015) in their quantitative review of studies addressing DRD4 polymorphisms, different classification of the data due to many variants or the collapsing of certain genetic types is frequent. To be clear, these decisions might be justified in each study. In a novel field without well-established procedural standards, researchers have to navigate through this garden of forking paths by finding a balance between adhering to what has been done in previous studies and adapting the procedure for their current purpose. Yet, without preregistration, it is hard for the field to retrace the rationale behind these decisions. Further, the resulting idiosyncratic analyses make it hard to compare findings across studies. In this context, it is not surprising that replication, testing very specific

predictions with preregistered analyses derived from previous literature did not find evidence for the predicted gene-behavior links.

So, what role do genes play in Theory of Mind development? Recent research drawing on large consortium data sets strongly suggests a polygenetic influence on cognition. For example, Davies et al. (2018) assessed links between genes and general cognitive function in a sample of over 300.000 participants. They identified 709 genes that were associated with general cognitive function. Further, there is well-documented environmental influence on Theory of Mind development, for example parental mental state talk (Tompkins et al., 2018), which constitutes an additional source of variance in our and previous studies. This, in combination with potential gene \times environment interactions, speaks for a complex multicausal influence on Theory of Mind development and against an unidirectional impact of single genotypes. Notably, a polygenetic influence, interacting with environmental features, does not preclude the possibility that particular genes play a detectable role in Theory of mind development. Yet, our replication attempt does not provide evidence for an influence of previously discussed candidate genes.

Our results suggest a more cautious interpretation of previous results in the domain of molecular genetic influences on Theory of Mind and call for further investigation. Evaluating the currently available evidential basis, we conclude that research on the molecular basis of Theory of Mind currently faces substantial methodological issues that prevent firm conclusions on genetic influences on Theory of Mind development. Our work highlights the importance of replication studies for estimating the conclusiveness of findings and demonstrates that they are an essential tool to support progress in developmental science.

ACKNOWLEDGMENT

This work was funded by a grant from the German Research Foundation (SO 213/27-1,2,3). We thank Nicole Schraud for excellent technical assistance and help with data analysis.

DATA AVAILABILITY STATEMENT

The preregistration and data of this study are available at <https://osf.io/vujgs/>.

ORCID

Tobias Schuwerk  <https://orcid.org/0000-0003-3720-7120>

Christopher Osterhaus  <https://orcid.org/0000-0002-1262-2427>

Klaus-Peter Lesch  <https://orcid.org/0000-0001-8348-153X>

ENDNOTES

¹ Note that Lackner et al. (2012) refer to a split between at least one long alleles (≥ 6 repeats) and short alleles (≤ 4 repeats) as no child has 5 repeats in their sample.

² We corroborated the (null) results of Table 3 in an additional analysis where we restricted the sample to complete observations (i.e., children without any missing data for one of the tasks at either measurement point). While this decreased the sample size slightly [e.g., from 71 to 65 children in column (6)], point estimates remained insignificant and close to zero. For the sake of brevity, we refrained from reporting

the results here, but shared the corresponding code in the analysis file.

³ We calculated the compound score despite a relatively low Cronbach's alpha of .28 of the two false belief tasks because the pattern of results remains stable when analyzing the content false belief task and the location false belief task separately.

⁴ We conducted two additional robustness checks for Table 4. First, we estimated columns (1)-(4) with logistic regressions, confirming the results. Second, we estimated columns (5)+(6) summing z-scores instead of simply taking the mean of the individual scores and found the same (null) results. We also included the corresponding code for these analyses in the analysis file.

REFERENCES

- Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development, 72*(4), 1032-1053.
- Carpendale, J. I., & Lewis, C. (2004). Constructing an understanding of mind: The development of children's social understanding within social interaction. *Behavioral and Brain Sciences, 27*(1), 79-96.
- Coull, G. J., Leekam, S. R., & Bennett, M. (2006). Simplifying second-order belief attribution: What facilitates children's performance on measures of conceptual understanding? *Social Development, 15*(3), 548-563.
- Davies, G., Lam, M., Harris, S. E., Trampush, J. W., Luciano, M., Hill, W. D., Hagenaaers, S. P., Ritchie, S. J., Marioni, R. E., Fawns-Ritchie, C., Liewald, D. C. M., Okely, J. A., Ahola-Olli, A. V., Barnes, C. L. K., Barnes, C. L. K., Bertram, L., Bis, J. C., Burdick, K. E., Christoforou, A., & ... Deary, I. J. (2018). Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nature Communications, 9*(1), 2098.
- 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.
- Dodell-Feder, D., Lincoln, S. H., Coulson, J. P., & Hooker, C. I. (2013). Using fiction to assess mental state understanding: a new task for assessing theory of mind in adults. *PLoS One, 8*(11), e81279.
- Ebstein, R. P., Israel, S., Chew, S. H., Zhong, S., & Knafo, A. (2010). Genetics of human social behavior. *Neuron, 65*(6), 831-844.
- Gong, P., Fan, H., Liu, J., Yang, X., Zhang, K., & Zhou, X. (2017). Revisiting the impact of OXTR rs53576 on empathy: A population-based study and a meta-analysis. *Psychoneuroendocrinology, 80*, 131-136.
- Grossmann, T., Vaish, A., Franz, J., Schroeder, R., Stoneking, M., & Friederici, A. D. (2013). Emotional voice processing: investigating the role of genetic variation in the serotonin transporter across development. *PLoS One, 8*(7), e68377.
- Happé, F. G. (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.
- Heyes, C. (2018). *Cognitive gadgets: The cultural evolution of thinking*. Cambridge, MA: Harvard University Press.
- Hughes, C., & Cutting, A. L. (1999). Nature, nurture, and individual differences in early understanding of mind. *Psychological Science, 10*(5), 429-432.
- Hughes, C., Jaffee, S. R., Happé, F., Taylor, A., Caspi, A., & Moffitt, T. E. (2005). Origins of individual differences in theory of mind: From nature to nurture? *Child Development, 76*(2), 356-370.
- Hughes, C., & Plomin, R. (2000). Individual differences in early understanding of mind: Genes, non-shared environment and modularity. In P. Carruthers & A. Chamberlain (Eds.), *Evolution and the human mind: Modularity, language, and meta-cognition* (pp. 47-61). Cambridge University Press.
- Killen, M., Mulvey, K. L., Richardson, C., Jampol, N., & Woodward, A. (2011). The accidental transgressor: Morally-relevant theory of mind. *Cognition, 119*(2), 197-215.
- Kloo, D., & Sodian, B. (2017). The developmental stability of inhibition from 2 to 5 years. *British Journal of Developmental Psychology, 35*, 582-595.
- Knafo, A., Israel, S., & Ebstein, R. P. (2011). Heritability of children's prosocial behavior and differential susceptibility to parenting by variation in the dopamine receptor D4 gene. *Development and Psychopathology, 23*(1), 53-67.
- Kristen, S., Thoermer, C., Hofer, T., Aschersleben, G., & Sodian, B. (2006). Skalierung von "Theory of Mind"-Aufgaben. *Zeitschrift Für Entwicklungspsychologie Und Pädagogische Psychologie, 38*(4), 186-195.
- Lackner, C., Sabbagh, M. A., Hallinan, E., Liu, X., & Holden, J. J. (2012). Dopamine receptor D4 gene variation predicts preschoolers' developing theory of mind. *Developmental Science, 15*(2), 272-280.
- LeBel, E. P., McCarthy, R. J., Earp, B. D., Elson, M., & Vanpaemel, W. (2018). A unified framework to quantify the credibility of scientific findings. *Advances in Methods and Practices in Psychological Science, 1*(3), 389-402.
- Lichter, J. B., Barr, C. L., Kennedy, J. L., Van Tol, H. H. M., Kidd, K. K., & Livak, K. J. (1993). A hypervariable segment in the human dopamine receptor D4 (DRD4) gene. *Human Molecular Genetics, 2*(6), 767-773.
- Lucht, M. J., Barnow, S., Sonnenfeld, C., Ulrich, I., Grabe, H. J., Schroeder, W., Freyberger, H. J., John, U., Herrmann, F. H., Kroemer, H., & Roskopf, D. (2013). Associations between the oxytocin receptor gene (OXTR) and "mind-reading" in humans—An exploratory study. *Nordic Journal of Psychiatry, 67*(1), 15-21.
- McClelland, G. H. (2000). Increasing statistical power without increasing sample size. *American Psychologist, 55*(8), 963-967.
- Moore, C., Bryant, D., & Furrow, D. (1989). Mental terms and the development of certainty. *Child Development, 60*(1), 167-171.
- Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences, 115*(11), 2600-2606.
- Nosek, B. A., & Errington, T. M. (2020). What is replication? *PLoS Biology, 18*(3), e3000691.
- Osterhaus, C., Kloo, D., Kristen-Antonow, S., Sodian, B. (2019). Scaling of (advanced) ToM and modeling children's ToM competencies: Longitudinal findings in 4- to 8-year-olds. Talk presented at the 2019 Biennial Meeting of the Society for Research in Child Development. Baltimore, MD.
- Osterhaus, C., Koerber, S., & Sodian, B. (2016). Scaling of advanced theory-of-mind tasks. *Child Development, 87*(6), 1971-1991.
- Pappa, I., Mileva-Seitz, V. R., Bakermans-Kranenburg, M. J., Tiemeier, H., & van Ijzendoorn, M. H. (2015). The magnificent seven: A quantitative review of dopamine receptor d4 and its association with child behavior. *Neuroscience & Biobehavioral Reviews, 57*, 175-186.
- Rodrigues, S. M., Saslow, L. R., Garcia, N., John, O. P., & Keltner, D. (2009). Oxytocin receptor genetic variation relates to empathy and stress reactivity in humans. *Proceedings of the National Academy of Sciences, 106*(50), 21437-21441.
- Rohwer, M., Kloo, D., & Perner, J. (2012). Escape from meta-ignorance: How children develop an understanding of their own lack of knowledge. *Child Development, 83*(6), 1869-1883.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*(11), 1359-1366.
- Simonsohn, U. (2015). Small telescopes: Detectability and the evaluation of replication results. *Psychological Science, 26*(5), 559-569.
- Skuse, D. H., & Gallagher, L. (2009). Dopaminergic-neuropeptide interactions in the social brain. *Trends in Cognitive Sciences, 13*(1), 27-35.
- Sodian, B., & Kristen-Antonow, S. (2015). Declarative joint attention as a foundation of theory of mind. *Developmental Psychology, 51*(9), 1190.



- Strommen, E. A. (1973). Verbal self-regulation in a children's game: Impulsive errors on "Simon Says". *Child Development*, 44(4), 849–853.
- Vickers, A. J. (2003). How many repeated measures in repeated measures designs? Statistical issues for comparative trials. *BMC Medical Research Methodology*, 3(1), 22.
- Wade, M., Hoffmann, T. J., Wigg, K., & Jenkins, J. M. (2014). Association between the oxytocin receptor (OXTR) gene and children's social cognition at 18 months. *Genes, Brain and Behavior*, 13(7), 603–610.
- Weisman, O., Pelphrey, K. A., Leckman, J. F., Feldman, R., Lu, Y., Chong, A., Chen, Y., Monakhov, M., Soo, H., & Ebstein, R. P. (2015). The association between 2D: 4D ratio and cognitive empathy is contingent on a common polymorphism in the oxytocin receptor gene (OXTR rs53576). *Psychoneuroendocrinology*, 58, 23–32.
- Wellman, H. M., & Bartsch, K. (1988). Young children's reasoning about beliefs. *Cognition*, 30, 239–277.
- 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.
- Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development*, 75(2), 523–541.
- Wimmer, H., & Perner, J. (1993). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13(1), 103–128.
- Wu, N., & Su, Y. (2015). Oxytocin receptor gene relates to theory of mind and prosocial behavior in children. *Journal of Cognition and Development*, 16(2), 302–313.
- Xia, H., Wu, N., & Su, Y. (2012). Investigating the genetic basis of theory of mind (ToM): The role of Catechol-O-Methyltransferase (COMT) gene polymorphisms. *PLoS One*, 7(11), e49768.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Opitz T, Schuwerk T, Paulus M, et al. No links between genetic variation and developing theory of mind: A preregistered replication attempt of candidate gene studies. *Dev Sci*. 2021;24:e13100. <https://doi.org/10.1111/desc.13100>