Trust and childhood maltreatment: evidence of bias in appraisal of unfamiliar faces

Louise Neil,[1] Essi Viding,[1] Diana Armbruster-Genc,[1] Matteo Lisi,[2] Isabelle Mareshal,[3] Georgia Rankin,[1] Molly Sharp,[1] Harriet Phillips,[1,4] Jessica Rapley,[1] Peter Martin,[5] and **Eamon McCrory[1,4]**

[1]Division of Psychology and Language Sciences, University College London, London, UK; [2]Department of Psychology, University of Essex, Colchester, UK; [3]Department of Psychology, Queen Mary University London, London, UK; [4]Anna Freud National Centre for Children and Families, London, UK; [5]Department of Applied Health Research, University College London, London, UK

Background: Child maltreatment is associated with poorer social functioning and increased risk of mental health problems in adolescence and adulthood, but the processes underlying these associations remain unclear. Although crucial for establishing and maintaining relationships, trust judgements have not been experimentally investigated in children who have experienced abuse and neglect. **Methods:** A community-based sample of 75 children aged 8–16 years with maltreatment documented on the basis of social services records, and a group of 70 peers matched on age, gender, cognitive ability, socioeconomic status, and ethnicity took part in the study. Children completed a trustworthiness face-judgement task in which they appraised the trustworthiness of unfamiliar facial stimuli varying along a computationally modelled trustworthiness dimension. **Results:** In line with clinical observations that childhood maltreatment is associated with an atypical pattern of trust processing, children with maltreatment experience were significantly less likely than their peers to rate unfamiliar faces as trustworthy. Moreover, they were more variable in their trust attributions than their peers. **Conclusions:** The study provides compelling experimental evidence that children with documented maltreatment perceive others as less trustworthy than their peers and are less consistent in their estimates of trustworthiness in others. Over time, alterations in trust processing may disrupt the development of social bonds and contribute to 'social thinning' (a reduction in the extent and quality of social relationships), leaving children more vulnerable to environmental stressors, increasing risk of mental health difficulties. **Keywords:** Maltreatment; childhood adversity; peer relationships; face processing; trust.

Introduction

Early maltreatment is robustly associated with an increased risk of mental health problems and relationship difficulties, even many years after the event (Gunnar & Reid, 2019; Jaffee, 2017). Understanding the pathogenesis of psychiatric disorder following childhood maltreatment is necessary if we are to intervene in a timely and targeted way in order to offset this risk. Considerable efforts are now being made to delineate maltreatment-driven alterations within specific neurocognitive processes thought to facilitate the later emergence of mental health difficulties (McCrory, Gerin, & Viding, 2017; McLaughlin, Weissman, & Bitrán, 2019). In particular, we have postulated that such alterations may confer latent vulnerability to future psychiatric disorder via their impact on social functioning (McCrory, Ogle, Gerin, & Viding, 2019; McCrory & Viding, 2015).

One critical aspect of social functioning is the ability to trust others (Fonagy & Allison, 2014). Trust is commonly defined as 'the intention to accept vulnerability based upon positive expectations of the intentions or behaviour of another' (Rousseau, Sitkin, Burt, & Camerer, 1998, p. 39). Higher levels of generalised beliefs in others' trustworthiness have

been associated with improved mental health and well-being across the life span (Chan, Hamamura, Li, & Zhang, 2017; Poulin & Haase, 2015) including reduced symptoms of depression (Lester & Gatto, 1990) although extremely high as well as extremely low levels of trust beliefs appear to be associated with poorer well-being in children (Rotenberg, Boulton, & Fox, 2005). The manner in which individuals perceive and judge trustworthiness has also been associated with mental health difficulties (Fertuck, Grinband, & Stanley, 2013). It is plausible that maladaptive trust judgements could facilitate the development of mental health problems. For example, perceptions of others' trustworthiness guide social approach and avoidance behaviours (Todorov, 2008; Winston, Strange, O'Doherty, & Dolan, 2002), and subsequently the establishment and maintenance of social support, which can protect against the development of mental health difficulties (van Harmelen et al., 2016).

It is thought that early interactions with responsive or emotionally available caregivers inform generalised expectancies of others' trustworthiness (Bowlby, 1980; Sharp & Fonagy, 2008), which are disrupted in individuals who do not receive adequate childhood care. Childhood maltreatment has been linked to reduced generalised trust beliefs (Gobin & Freyd, 2014) and a reticence to trust unfamiliar

others (see Bernath & Feshbach, 1995 for a review). Conversely, patterns of disinhibited social engagement identified in some individuals following abuse and neglect (Zeanah & Gleason, 2015), appear to demonstrate increased willingness to trust others. Such behaviours are variously hypothesised to arise from motivated forgetting of untrustworthy judgements in order that necessary attachments be formed in infancy (Freyd, 1996), difficulties discriminating the trustworthiness of caregivers from unfamiliar others following insecure attachment (Lieberman, 2003), or a preference for interactions with unfamiliar others following discordant or unsatisfying social interactions with caregivers (Minnis, Marwick, Arthur, & McLaughlin, 2006).

Although trust has long been linked to mental health outcomes (Erikson, 1963), it is only recently that component mechanisms which contribute to trust judgements have been explored within a neurocognitive perspective (Krueger & Meyer-Lindenberg, 2019). This line of work can complement previous self-report (Bernath & Feshbach, 1995) and observational (Galdston, 1968) research by elucidating information processing patterns, that may be associated with atypical trust relationships. Such work is in its infancy in populations who have experienced maltreatment despite its potential importance in shedding light on social functioning and mental vulnerability in this group. In a relatively recent experimental study Pitula, Wenner, Gunnar, and Thomas (2017) found that previously institutionalised children were less trusting of anonymous peers in an established economic paradigm compared with children who were never institutionalised or adopted at an earlier age. These children also adapted their behaviours to signs of untrustworthiness in others more readily. In contrast to work on institutionalised samples, no prior study has investigated trustworthiness decisions in children with documented maltreatment recruited from a community sample using an experimental approach. One recent study with adults by Hepp, Schmitz, Urbild, Zauner, and Niedtfeld (2021) found that higher self-reports of childhood maltreatment in adults were associated with lower levels of trust in a hypothetical distrust game (Hepp et al., 2021). Collectively, these findings indicate that notwithstanding patterns of disinhibited social engagement in some individuals following abuse and neglect (Zeanah & Gleason, 2015), we would expect on balance such experiences to be associated with judgements characterised by reduced trust.

One potentially crucial component of trust processing previously unexplored in maltreated children is the attribution of trustworthiness to facial cues. There is an established experimental literature demonstrating that from infancy, children show preferential attention to the human face (Bushnell, Sai, & Mullin, 1989; Johnson, Dziurawiec, Ellis, & Morton, 1991; Meltzoff & Moore, 1989). This is thought to play a critical role in the foundation of reciprocal social interactions. Certain structural, global, and emotional qualities have a pervasive influence on social attributions and a broad consensus of what constitutes a trustworthy face appears to develop early in early childhood (Cogsdill, Todorov, Spelke, & Banaji, 2014). Atypical environments characterised by physical abuse as well as general features of adversity have been associated with differences in how individuals subsequently process information from faces. For example, increased sensitivity to threatening faces has been documented in individuals who have experienced early adversity (Ardizzi et al., 2015), physical abuse (Pollak, 2003), and family violence (McCrory et al., 2011). By focusing on facial trust attributions as a specific component of trust processing, we may be able to further understand the impact of maltreatment on social functioning at a mechanistic level. An extant study using facial stimuli in a related population indicates that this may be a fruitful avenue of research. Miellet, Caldara, Gillberg, Raju, and Minnis (2014) reported reduced consistency in trustworthiness judgements in children with a diagnosis of disinhibited reactive attachment disorder and maltreatment history (Miellet et al., 2014). However, the sample was comprised of only ten children, all of whom also had disinhibited social behaviour, making it unclear whether the findings could be generalised to maltreated children in general. We currently do not know whether biases in facial trust processing are present in children with documented histories of maltreatment, before mental health difficulties emerge. Identifying any such alterations in trust processing at an earlier stage may be relevant in the development of targeted approaches to prevention.

The aim of this study was to investigate facial trust attribution in children with prior maltreatment experience. In light of the extant evidence indicating that children with experience of institutionalisation and adults with experiences of childhood maltreatment present with lower levels of trust attributions, we hypothesised that a community sample of children with a history of documented maltreatment would rate unfamiliar faces, as less trustworthy overall, in comparison to ratings made by a group of carefully matched peers. As maltreatment will likely have been characterised by increased variability compared with peers in experiences of others' trustworthiness, we also hypothesised that these children may be less consistent in their trust judgements. In addition, we investigated whether biases in trust processing accompanying maltreatment experience were associated with manifest social and mental health difficulties. As biased facial trustworthiness judgements may reduce social approach, we also included a measure of apathy (a loss of motivation in order to seek pleasure).

Methods

Participants

75 children with documented maltreatment (MT) and 70 non-maltreated (NMT) children aged 8–16 years were included in this study. Children in the maltreatment group were recruited through one of three London social services (SS) departments. Children in the non-maltreatment group were recruited through schools in similar areas. Children with IQ scores <70, a diagnosed learning disability, or autism spectrum disorder were excluded from both groups while children with previous SS contact due to concerns over their care at home were excluded from the non-maltreated group.

A pairwise matching approach based on age and IQ scores was used to select the two samples using the Matching package (Sekhon, 2011, v.4.9-7) for R (R Core Team, 2019; please see Appendix S1 for further details). All children were assigned a weight of '1' except five of the non-maltreated sample who were each matched to two children from the maltreated sample and therefore assigned a weight of '2' in the analysis. There were no significant differences between the groups on age, socioeconomic status, cognitive ability, gender, or ethnicity (see Table 1).

An additional sample of non-maltreated participants ($n = 18$) carried out the task on two separate occasions, two weeks apart, in order to obtain an indicator of test-retest reliability on the task. These children (11 girls, 7 boys) were all

Table 1 Sample characteristics for children with (MT) and without maltreatment experience (NMT)

Measures	MT	NMT	p
Gender			
n Female (% female)	36 (48%)	40 (53.3%)	.53
Age			
M (*SD*)	12.73 (2.34)	12.70 (2.17)	.92
IQ[a]			
M (*SD*)	100.05 (9.42)	100.59 (9.40)	.74
SES[b]			
M (*SD*)	3.27 (0.96)	3.05 (0.96)	.18
Ethnicity[c]			
Black	18 (24%)	14 (19.2%)	.09
White	31 (41.3%)	34 (46.6%)	
Asian	5 (6.7%)	12 (16.4%)	
Mixed	21 (28%)	11 (15.1%)	
Other Ethnicity	0 (0%)	2 (2.7%)	
SDQ[d]			
Total difficulties score	12.75 (4.96)	9.25 (4.32)	<.001
Prosocial behaviour	7.73 (1.42)	8.63 (1.00)	<.001
SCPQ[e]	14.01 (3.92)	16.00 (3.12)	.002
CASSS[f]	56.52 (12.45)	57.59 (10.92)	.61
AES[g]	18.67 (3.69)	15.70 (3.30)	.002

[a]IQ was measured using the vocabulary and matrices subtests of the WASI-II (Wechsler, 2011), $n = 150$.
[b]SES = Socioeconomic status: Highest level of education rated on a 6-point scale from 0 (no formal qualifications) to 5 (postgraduate qualifications), $n = 145$.
[c]Ethnicity $n = 148$.
[d]SDQ = Strengths and Difficulties Questionnaire. Scores reflect parent and child reports averaged, $n = 125$.
[e]SCPQ = Social Competence with Peers Questionnaire, $n = 125$.
[f]CASSS = Child and Adolescent Social Support Scale, close friend (frequency scale), $n = 137$.
[g]AES = Apathy Evaluation Scale, shortened form, $n = 63$.

either 15 or 16-years-old ($M = 16.06$, $SD = 0.29$). Further test-retest of young participants was cut short by the pandemic.

Written consent was provided by the child's legal guardian. Written and verbal assent was provided by all children in the study. The procedures used in this study were approved by the University Research Ethics Committee.

Measures

Trustworthiness face-judgement task. A trustworthiness face-judgement task was used to measure children's first impressions of face trustworthiness. In the task, children were shown 70 adult face stimuli, each one presented separately, and invited to answer yes or no to the question 'Is this face trustworthy.' Participants indicated their answer by pressing either a yes or a no key. No time limit was imposed on trial responses. A central fixation cross was presented at the beginning of each of the 70 trials.

Ten face identities were randomly selected by our group from a set of twenty-five maximally distinct computer-generated identities created with FaceGen software (Singular Inversions, Toronto, Canada) as described by Todorov, Dotsch, Porter, Oosterhof, and Falvello (2013). Each of these ten identities had been manipulated to yield seven morphs within each identity at 1 standard deviation increments from −3 to +3 standard deviations along a computationally modelled trustworthiness dimension (Todorov et al., 2013; Todorov & Oosterhof, 2011). The resulting coloured facial image stimuli were presented in randomised order over 70 trials.

Four practice trials comprising two additional identities at maximum and minimum levels of trustworthiness (+ and −3*SD* from the mean) were administered at the start of the task.

Maltreatment history. For children referred to Social Services (SS), maltreatment history was provided by social workers on the basis of SS records, using a validated maltreatment scale (Kaufman, Jones, Stieglitz, Vitulano, & Mannarino, 1994). The maltreatment categories included neglect ($n = 49$, 65%), emotional abuse ($n = 67$, 89%), sexual abuse ($n = 8$, 11%), and home violence ($n = 61$, 81%). A table detailing the onset, duration and severity by category is provided in the Supporting Information (Table S1).

Child- and parent-reported questionnaires. Questionnaire-based measures of social and psychological functioning included the parent and child versions of the Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997), the Social Competence with Peers Questionnaire (parent report) (SCPQ; Spence, 1995), the Child and Adolescent Social Support Scale, close friend scale (child-report) (CASSS; Malecki, Demaray, Elliott, & Nolten, 1999), and a shortened version of the Apathy Evaluation Scale (child-report) (AES; Marin, Biedrzycki, & Firinciogullari, 1991). Scores are reported in Table 1 (for questionnaire details, please see Appendix S2).

Data analysis

Each child's keypress responses were converted to proportion of 'trustworthy' scores for the seven facial trustworthiness levels tested. A generalised-linear multilevel model (GLMM) with a probit link function was carried out in the R software for statistical computing (R Core Team, 2019) using the lme4 package to estimate parameters at the group level while assigning random effects to the face identities. The parameters of interest derived from the fitted psychometric functions were the bias (corresponding to the level of trustworthiness at which the child was equally likely to respond 'trustworthy' or 'not trustworthy') and variability (insensitivity to changes in the level of trustworthiness of the stimulus). Standard errors and confidence intervals were estimated using the parametric

bootstrap and accounted for the weighted cases (see Appendix S3 for further details). Participant-specific random intercepts, corresponding to log-odds of a participant judging the mid-range stimulus as trustworthy, were extracted from the model and correlated (Pearson's, two tailed) with parent and child-reported questionnaire measures of concurrent social functioning and psychological functioning. Power calculations can be found in Appendix S4.

Results

Test-retest reliability

In the 18 additional NMT children who carried out the task twice, trust bias scores had moderate to strong test-retest reliability ($r = .68$, $p < .001$, on participant-specific intercepts). The variability scores had poorer test-retest reliability ($r_s = .40$, $p = .10$) which may in part result from practice effects that reduce performance variability on the second administration.

Trust judgements

The bias in the maltreated group was significantly greater than in the group of non-maltreated counterparts with bootstrapped confidence intervals demonstrating that this difference was significant (see Table 2), meaning that children in the maltreatment group required a greater level of trustworthiness before offering a 'trustworthy' judgement. Figure 1 demonstrates that while both groups 'correctly' judged faces at the lower end of the trustworthiness continuum as untrustworthy, differences emerged when faces were manipulated to exhibit more 'trustworthy' characteristics. The model also demonstrated that there was a significant effect of group on the variability of judgements, with children in the maltreated group more likely to show more variability in their responses. Further details of results are provided in the Supporting Information (Table S2, Figures S1–S3).

Emotion recognition

In order to investigate whether generalised difficulties in facial emotion recognition could explain the
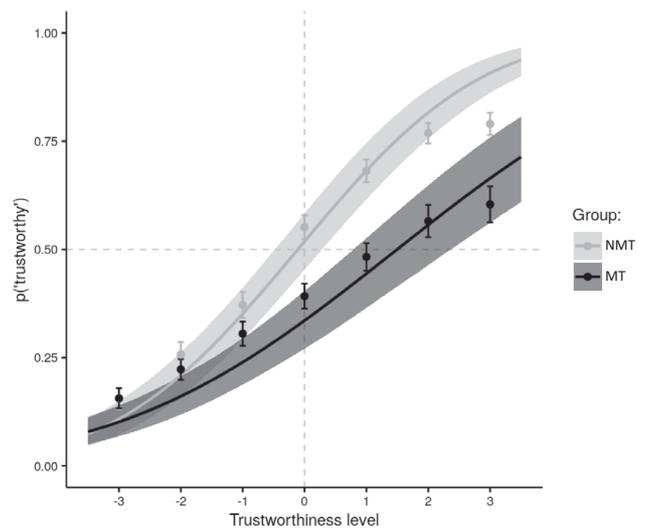

Figure 1 The proportion of faces judged 'trustworthy' at each level of trust, by group: children with maltreatment (MT) and without maltreatment experience (NMT). The error bars on the dots are standard errors across observers (note that dots were calculated without taking frequency weights into account) whereas lines and shaded bands are model predictions with 95% confidence intervals

observed differences in trust, a subsample of children who came into the university for a second testing session carried out an emotion recognition task ($N = 60$; MT = 38, NMT = 22). Children were first asked to identify the emotion of a series of fearful or joyful faces, and then to rate the intensity of the emotion on a scale from 1 to 10. Faces were presented to them as movies converging on either low, medium, or high levels of emotional intensity (Wingenbach, Ashwin, & Brosnan, 2016). First, a three-way mixed analysis of variance was conducted with group (MT or NMT) as a between-groups factor and facial emotion/level as within subjects' factors on participant's emotion recognition accuracy. There was no main effect of group on emotion recognition ($F(1, 58) = 0.04$, $p = .84$, $\eta p^2 = .001$. Nor was there a significant interaction between group and emotion ($F(1, 58) = 1.94$, $p = .17$, $\eta p^2 = .03$) or group and level ($F(1, 58) = 0.05$, $p = .83$, $\eta p^2 = .001$. Second, a three-way mixed analysis of variance was carried out on participants' 1–10 intensity ratings of correctly judged emotions, again with group (MT or NMT) as a between-groups factor and facial emotion/level as within-subjects' factors. There was no significant main effect of group on participant's intensity ratings ($F(1, 55) = 0.01$, $p = .92$, $\eta p^2 = .00$. In addition, there was no significant interaction between group and emotion ($F(1, 55) = 2.40$, $p = .13$, $\eta p^2 = .04$ or group and level ($F(1, 55) = 0.09$, $p = .92$, $\eta p^2 = .002$) on intensity ratings. In this same subsample, children with documented maltreatment had significantly greater trust bias scores than non-maltreated children ($t(58) = 3.77$, $p < .001$; two-tailed test on random intercepts) making it unlikely that generalised difficulties in facial

Table 2 Group-level estimates of the GLMM model for non-maltreated and maltreated children

	NMT	MT	Difference
Bias	-0.13 ± 0.19 95%CI [-0.48, 0.28]	1.55 ± 0.4 95%CI [0.79, 2.36]	1.69 ± 0.36 95%CI [0.99, 2.36]
Variability	2.26 ± 0.15 95%CI [2.06, 2.66]	3.49 ± 0.34 95%CI [2.97, 4.32]	1.23 ± 0.35 95%CI [0.59, 1.95]

The table shows the mean ($\pm$ bootstrapped standard errors) and the 95% confidence intervals calculated with the percentile method.

emotion recognition can explain the observed differences in trust. Please see the Supporting Information for further details of the task methodology and results (Appendix S5, Table S3, Figure S4).

Associations between trust and social and psychological functioning

In the maltreatment group, neither prosocial behaviour, as measured by a subscale of the SDQ ($r = .04$, $p = .77$) nor scores on the SCPQ ($r = -.03$, $p = .83$) showed significant associations with children's propensity to trust. In terms of concurrent mental health functioning, there was no significant association with the SDQ total difficulties score ($r = .07$, $p = .57$). There was, however, a moderate association with AES scores, which fell just short of significance after applying a Benjamini-Hochberg corrected alpha of .013 ($r = -.40$, $p = .017$), indicating that greater levels of apathy may be linked to a reduced propensity towards trusting unfamiliar faces in children with histories of maltreatment. There was a modest negative relationship, which did not reach statistical significance ($r = -.21$, $p = .07$), between severity of abuse (as indicated by children's summed Kaufman subscale scores) and children's propensity to trust in the maltreated group. In the non-maltreated group, there were no significant associations between propensity to trust and scores on the SCPQ ($r = .20$, $p = .17$), the SDQ ($r = -.21$, $p = .16$) or the AES ($r = -.05$, $p = .80$). A small association between trust and prosocial behaviour ($r = .30$, $p = .04$), indicating that children in this group who were more likely to judge unfamiliar faces as trustworthy engaged in greater levels of prosocial behaviour, did not reach significance after applying a Benjamini-Hochberg corrected alpha of .013.

Discussion

This is the first experimental study to investigate how children, who have experienced maltreatment, make judgements regarding the trustworthiness of others on the basis of facial cues. From a clinical perspective, there has been longstanding consensus that trust processing is atypical in this group (Martin, 1976) but rather surprisingly, this aspect of trust processing has not been experimentally investigated. In line with our predictions, participants with documented childhood maltreatment were significantly more likely to rate unfamiliar facial stimuli as untrustworthy and showed more variability in their trust judgements than peers of a similar age, cognitive ability, and background. These findings indicate that maltreated children perceive others as less trustworthy than their peers, and are less consistent in how they use or judge facial cues when attributing trustworthiness to others.

The study provides experimental evidence in line with the widespread assumption that violations of

trust by key attachment figures, such as threatening, inconsistent, or neglectful caregiving, can translate to reduced trust in unfamiliar others (Bowlby, 1980; Sharp & Fonagy, 2008). Our findings are consistent with those of Pitula et al., (2017) who found that children adopted after 12 months of being institutionalised were less likely to trust anonymous peers in the context of an economic game, and previous research utilising self-report or observational methodologies, which has identified greater mistrust in children with experience of abuse or neglect (see Bernath & Feshbach, 1995 for a review). Crucially, this study contributes to our understanding of the impact of maltreatment on trust, by experimentally examining one specific aspect of trust, facial judgements on first impressions, and demonstrating that this particular process is disrupted in children who have experienced maltreatment. The finding that these children did not differ from their peers in emotion recognition or intensity ratings assigned to emotional faces in a separate task suggests that the observed differences in trust judgements are not merely a by-product of broader difficulties in emotion recognition.

The observed pattern of atypical trust processing may contribute to our understanding of the established link between maltreatment and later social and mental health difficulties (McCrory et al., 2019). Abusive and neglectful environments by definition are characterised by relationships which fail to protect and meet the needs of a child. In such a context, a child might be expected to develop an atypical internal model of trustworthiness, affecting how facial configurations are processed. We hypothesise that, over time, reduced perceptions of trustworthiness as well as inconsistent trust attributions will impact an individual's ability to cultivate and maintain stable social relationships with others. This, alongside other neurocognitive processes and environmental factors (McCrory & Viding, 2015) may contribute to a reduced number of social bonds with others, and a reduction in the quality of those bonds. Prospective longitudinal studies have shown that the experience of childhood maltreatment is associated with reduced social support from others, even decades later when individuals enter adulthood (Sperry & Widom, 2013). Other studies have demonstrated that prior maltreatment experience is also associated with both social isolation and loneliness over time (Hanlon et al., 2020). This pattern of attenuated social networks following maltreatment (including both extent and quality of social relationships), relative to peers, has been termed 'social thinning' (McCrory, 2020; Viding & McCrory, 2020) and is postulated to be one factor contributing to long term psychiatric risk (Wielaard, Hoyer, Rhebergen, Stek, & Comijs, 2018). There is a need for prospective studies, incorporating experimental measures of trust, to investigate whether trust

processing is one mechanism which may contribute to social thinning over time.

Reduced propensity to trust was not associated with concurrent social functioning in maltreated children. It may be that the questionnaire-based measures of social functioning used in this study were insufficiently sensitive to those domains of social processing most implicated by trust processing. Longitudinal social network measures that objectively quantify the degree to which a child is integrated with their peer group over time would potentially represent a more sensitive marker of social functioning that could be incorporated in a future study. Moreover, trust bias may constitute a marker of latent vulnerability and compromise social competence at later stages of adolescence in the relationships with both adults and peers. Over time, the cumulative impact of reduced trust may compromise the quality and longevity of relationships leading to impoverished social relationships and vulnerability to mental health problems. This interpretation is in line with the association of moderate effect size between trust attributions and children's self-reports of apathy found in this study. Reduced responsivity to reward has previously been identified as a marker of latent vulnerability to future depression (Stringaris et al., 2015). Although this association did not survive stringent correction for multiple comparisons, it suggests that apathy may have a role in diminishing the perceived benefits of engaging and initiating interactions with others. We believe this association warrants further systematic investigation.

There are a number of limitations to our study. First, the cross-sectional approach adopted here means we can only theorise on the impact of trust bias on children's future social and psychological functioning. Prospective longitudinal data are necessary to investigate the impact of trust bias on the development of children's social relationships as they navigate later adolescence and early adulthood. However, this study highlights trust as an important focus for future research aimed at undercovering the potential mechanisms underlying social and mental health difficulties following childhood maltreatment. Second, we focussed only on one specific aspect of trust processing. Future studies could employ auditory stimuli, or study trust perception in the context of decision-making and threat processing to gain a fuller understanding of how trust is processed in children following maltreatment experience. Investigating children's judgements of facial trustworthiness using child as well as adult facial stimuli could also help shed further light on the extent of altered trust processing in this population. Third, the emotion recognition control task was only administered to a subsample of the participants when it would have been preferable to include the whole sample.

In conclusion, this study is the first to experimentally investigate facial trust attributions in children who have experienced maltreatment. The findings demonstrate that these children find unfamiliar faces significantly less trustworthy than their non-maltreated peers, and that they are also less consistent when deciding what constitutes facial trustworthiness. We suggest this pattern of atypical trust processing may be one factor that contributes to latent vulnerability for the development of mental health problems, and alongside elevated levels of apathy, may compromise the cultivation and maintenance of social relationships. Future investigations using sensitive measures of social relationships within a longitudinal framework are needed to explore the impact of trust bias on relationships and mental health over time, and delineate how this pathway unfolds.

Supporting information
Additional supporting information may be found online in the Supporting Information section at the end of the article:

Appendix S1. Matching procedure.
Appendix S2. Questionnaires.
Appendix S3. Bootstrapping approach.
Appendix S4. Power calculations.
Appendix S5. Emotion Recognition Task methodology.
Table S1. Maltreatment severity, estimated duration, and age of onset.
Table S2. Table output for more standard GLM parametrisation.
Table S3. Emotion Recognition Task descriptive statistics by group.
Figure S1. 'Null' and standard bootstrapped distributions of the bias and variability parameters.
Figure S2. Psychometric functions for individual participants.
Figure S3. Distribution of biases after excluding subjects with $|\mu| >$ or <3.
Figure S4. Distribution of trust bias scores in Emotion Recognition Task sub-sample in comparison to the main sample.

Correspondence
Eamon McCrory, Division of Psychology and Language Sciences, University College London, 26 Bedford Way, London WC1H 0AP, UK; Email: e.mccrory@ucl.ac.uk

Key points

- Childhood maltreatment is a significant risk factor for mental health, social and relationship difficulties.
- This study provides the first experimental evidence that children with documented maltreatment judge unfamiliar faces as less trustworthy and are also less consistent in their trust judgements, than their non-maltreated peers.
- Trust bias of unfamiliar faces may index latent vulnerability for the development of mental health problems. This vulnerability may operate via diminishing social networks, a phenomenon known as 'social thinning'.

References

Ardizzi, M., Martini, F., Umiltà, M.A., Evangelista, V., Ravera, R., & Gallese, V. (2015). Impact of childhood maltreatment on the recognition of facial expressions of emotions. *PLoS One, 10*, e0141732.

Bernath, M.S., & Feshbach, N.D. (1995). Children's trust: Theory, assessment, development, and research directions. *Applied and Preventive Psychology, 4*, 1–19.

Bowlby, J. (1980). Attachment and loss: Volume III: Loss, sadness and depression. In *Attachment and Loss: Volume III: Loss, Sadness and Depression* (pp. 1–462). London: The Hogarth press and the institute of psycho-analysis.

Bushnell, I.W.R., Sai, F., & Mullin, J.T. (1989). Neonatal recognition of the mother's face. *British Journal of Developmental Psychology, 7*, 3–15.

Chan, D.K.C., Hamamura, T., Li, L.M.W., & Zhang, X. (2017). Effect of human development on the relationship between generalised trust and health: an international cross-sectional investigation. *The Lancet, 390*, S105.

Cogsdill, E.J., Todorov, A.T., Spelke, E.S., & Banaji, M.R. (2014). Inferring character from faces: A developmental study. *Psychological Science, 25*, 1132–1139.

Erikson, E.H. (1963). *Childhood and society*. New York: W. W. Norton & Company.

Fertuck, E.A., Grinband, J., & Stanley, B. (2013). Facial trust appraisal negatively biased in borderline personality disorder. *Psychiatry Research, 207*, 195–202.

Fonagy, P., & Allison, E. (2014). The role of mentalizing and epistemic trust in the therapeutic relationship. *Psychotherapy, 51*, 372.

Freyd, J.J. (1996). *Betrayal trauma. Encyclopedia of psychological trauma*, 76.

Galdston, R. (1968). Dysfunctions of parenting: The battered child, the neglected child, the exploited child. In J.G. Howells (Ed.), *Modern perspectives in international child psychiatry* (pp. 571–584). Edinburgh: Oliver and Boyd.

Gobin, R.L., & Freyd, J.J. (2014). The impact of betrayal trauma on the tendency to trust. *Psychological Trauma: Theory, Research, Practice, and Policy, 6*, 505.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry, 38*, 581–586.

Gunnar, M.R., & Reid, B.M. (2019). Early Deprivation Revisited: Contemporary studies of the impact on young children of institutional care. *Annual Review of Developmental Psychology, 1*, 93–118.

Hanlon, P., McCallum, M., Jani, B.D., McQueenie, R., Lee, D., & Mair, F.S. (2020). Association between childhood maltreatment and the prevalence and complexity of multimorbidity: A cross-sectional analysis of 157,357 UK Biobank participants. *Journal of Comorbidity, 10*, 2235042X1094 4344.

Hepp, J., Schmitz, S.E., Urbild, J., Zauner, K., & Niedtfeld, I. (2021). Childhood maltreatment is associated with distrust and negatively biased emotion processing. *Borderline Personality Disorder and Emotion Dysregulation, 8*, 1–14.

Jaffee, S.R. (2017). Child maltreatment and risk for psychopathology in childhood and adulthood. *Annual Review of Clinical Psychology, 13*, 525–551.

Johnson, M.H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition, 40*, 1–19.

Kaufman, J., Jones, B., Stieglitz, E., Vitulano, L., & Mannarino, A.P. (1994). The use of multiple informants to assess children's maltreatment experiences. *Journal of Family Violence, 9*, 227–232.

Krueger, F., & Meyer-Lindenberg, A. (2019). Toward a model of interpersonal trust drawn from neuroscience, psychology, and economics. *Trends in Neurosciences, 42*, 92–101.

Lester, D., & Gatto, J.L. (1990). Interpersonal trust, depression, and suicidal ideation in teenagers. *Psychological Reports, 67*, 786.

Lieberman, A.F. (2003). The treatment of attachment disorder in infancy and early childhood: Reflections from clinical intervention with later-adopted foster care children. *Attachment & Human Development, 5*, 279–282.

Malecki, C.K., Demaray, M.K., Elliott, S.N., & Nolten, P.W. (1999). *The child and adolescent social support scale*. DeKalb, IL: Northern Illinois University.

Marin, R.S., Biedrzycki, R.C., & Firinciogullari, S. (1991). Reliability and validity of the Apathy Evaluation Scale. *Psychiatry Research, 38*, 143–162.

Martin, H.P. (Ed.) (1976). *The abused child: A multidisciplinary approach to developmental issues and treatment*. Cambridge, MA: Ballinger.

McCrory, E.J. (2020). The case for a preventative approach to mental health: Childhood maltreatment, neuroimaging, and the theory of latent vulnerability. In W. Davies, J. Savulescu & R. Roache (Eds.), *Psychiatry reborn: Biopsychosocial psychiatry in modern medicine* (pp. 175–189). Oxford, UK: Oxford University Press.

McCrory, E.J., De Brito, S.A., Sebastian, C.L., Mechelli, A., Bird, G., Kelly, P.A., ... & Viding, E. (2011). Heightened neural reactivity to threat in child victims of family violence. *Current Biology, 21*, R947–R948.

McCrory, E.J., Gerin, M.I., & Viding, E. (2017). Annual research review: childhood maltreatment, latent vulnerability and the shift to preventative psychiatry–the contribution of functional brain imaging. *Journal of Child Psychology and Psychiatry, 58*, 338–357.

McCrory, E., Ogle, J.R., Gerin, M.I., & Viding, E. (2019). Neurocognitive adaptation and mental health vulnerability following maltreatment: The role of social functioning. *Child Maltreatment, 24*, 435–451.

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

McLaughlin, K.A., Weissman, D., & Bitrán, D. (2019). Childhood adversity and neural development: A systematic review. *Annual Review of Developmental Psychology*, *1*, 277–312.

Meltzoff, A.N., & Moore, M.K. (1989). Imitation in newborn infants: Exploring the range of gestures imitated and the underlying mechanisms. *Developmental Psychology*, *25*, 954.

Miellet, S., Caldara, R., Gillberg, C., Raju, M., & Minnis, H. (2014). Disinhibited reactive attachment disorder symptoms impair social judgements from faces. *Psychiatry Research*, *215*, 747–752.

Minnis, H., Marwick, H., Arthur, J., & McLaughlin, A. (2006). Reactive attachment disorder—a theoretical model beyond attachment. *European Child and Adolescent Psychiatry*, *15*, 336–342.

Pitula, C.E., Wenner, J.A., Gunnar, M.R., & Thomas, K.M. (2017). To trust or not to trust: social decision-making in post-institutionalized, internationally adopted youth. *Developmental Science*, *20*, e12375.

Pollak, S.D. (2003). Experience-dependent affective learning and risk for psychopathology in children. *Annals of the New York Academy of Sciences*, *1008*, 102–111.

Poulin, M.J., & Haase, C.M. (2015). Growing to trust: Evidence that trust increases and sustains well-being across the life span. *Social Psychological and Personality Science*, *6*, 614–621.

R Core Team (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available from: https://www.R-project.org/

Rotenberg, K.J., Boulton, M.J., & Fox, C.L. (2005). Cross-sectional and longitudinal relations among children's trust beliefs, psychological maladjustment and social relationships: are very high as well as very low trusting children at risk? *Journal of Abnormal Child Psychology*, *33*, 595–610.

Rousseau, D.M., Sitkin, S.B., Burt, R.S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review*, *23*, 393–404.

Sekhon, J.S. (2011). Multivariate and propensity score matching software with automated balance optimisation: The Matching package for R. *Journal of Statistical Software*, *42*, 1–52.

Sharp, C., & Fonagy, P. (2008). Social cognition and attachment-related disorders. In C. Sharp, P. Fonagy & I. Goodyer (Eds.), *Social cognition and developmental psychopathology* (pp. 271–302). Oxford University Press.

Spence, S.H. (1995). *Social skills training: Enhancing social competence and children and adolescents*. Windsor, UK: The NFER-NELSON Publishing Company.

Sperry, D.M., & Widom, C.S. (2013). Child abuse and neglect, social support, and psychopathology in adulthood: A prospective investigation. *Child Abuse and Neglect*, *37*, 415–425.

Stringaris, A., Vidal-Ribas Belil, P., Artiges, E., Lemaitre, H., Gollier-Briant, F., Wolke, S., ... & Paillère-Martinot, M.-L. (2015). The brain's response to reward anticipation and depression in adolescence: dimensionality, specificity, and longitudinal predictions in a community-based sample. *American Journal of Psychiatry*, *172*, 1215–1223.

Todorov, A. (2008). Evaluating faces on trustworthiness: An extension of systems for recognition of emotions signaling approach/avoidance behaviors. *Annals of the New York Academy of Sciences*, *1124*, 208–224.

Todorov, A., Dotsch, R., Porter, J.M., Oosterhof, N.N., & Falvello, V.B. (2013). Validation of data-driven computational models of social perception of faces. *Emotion*, *13*, 724.

Todorov, A., & Oosterhoof, N.N. (2011). Modeling social perception of faces. *Signal Processing Magazine, IEEE*, *28*, 117–122.

van Harmelen, A.-L., Gibson, J.L., St Clair, M.C., Owens, M., Brodbeck, J., Dunn, V., ... & Goodyer, I.M. (2016). Friendships and family support reduce subsequent depressive symptoms in at-risk adolescents. *PLoS One*, *11*, e0153715.

Viding, E., & McCrory, E. (2020). Disruptive behavior disorders: The challenge of delineating mechanisms in the face of heterogeneity. *American Journal of Psychiatry*, *177*, 811–817.

Wechsler, D. (2011). *WASI-II: Wechsler abbreviated scale of intelligence* (2nd edn). San Antonio, TX: Psychological Corporation.

Wielaard, I., Hoyer, M., Rhebergen, D., Stek, M.L., & Comijs, H.C. (2018). Childhood abuse and late-life depression: Mediating effects of psychosocial factors for early-and late-onset depression. *International Journal of Geriatric Psychiatry*, *33*, 537–545.

Wingenbach, T.S., Ashwin, C., & Brosnan, M. (2016). Validation of the Amsterdam Dynamic Facial Expression Set-Bath Intensity Variations (ADFES-BIV): A set of videos expressing low, intermediate, and high intensity emotions. *PLoS One*, *11*, e0147112.

Winston, J.S., Strange, B.A., O'Doherty, J., & Dolan, R.J. (2002). Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nature Neuroscience*, *5*, 277–283.

Zeanah, C.H., & Gleason, M.M. (2015). Annual research review: Attachment disorders in early childhood–clinical presentation, causes, correlates, and treatment. *Journal of Child Psychology and Psychiatry*, *56*, 207–222.