



# A Meta-analysis on Children's Trust in Social Robots

Rebecca Stower<sup>1</sup> · Natalia Calvo-Barajas<sup>2</sup> · Ginevra Castellano<sup>2</sup> · Arvid Kappas<sup>1</sup>

Accepted: 2 December 2020 / Published online: 9 February 2021  
© The Author(s) 2021

## Abstract

Although research on children's trust in social robots is increasingly growing in popularity, a systematic understanding of the factors which influence children's trust in robots is lacking. In addition, meta-analyses in child–robot–interaction (cHRI) have yet to be popularly adopted as a method for synthesising results. We therefore conducted a meta-analysis aimed at identifying factors influencing children's trust in robots. We constructed four meta-analytic models based on 20 identified studies, drawn from an initial pool of 414 papers, as a means of investigating the effect of robot embodiment and behaviour on both social and competency trust. Children's pro-social attitudes towards social robots were also explored. There was tentative evidence to suggest that more human-like attributes lead to less competency trust in robots. In addition, we found a trend towards the type of measure that was used (subjective or objective) influencing the direction of effects for social trust. The meta-analysis also revealed a tendency towards under-powered designs, as well as variation in the methods and measures used to define trust. Nonetheless, we demonstrate that it is still possible to perform rigorous analyses despite these challenges. We also provide concrete methodological recommendations for future research, such as simplifying experimental designs, conducting a priori power analyses and clearer statistical reporting.

**Keywords** Trust · Social robot · Child robot interaction · Human robot interaction · Meta-analysis · Review · Robot errors · Pro-social attitudes · Developmental psychology

## 1 Introduction

As human–robot–interaction (HRI) research continues to grow in popularity, increasing focus is being placed towards children as a significant user group. Much of the research on child–robot–interaction (cHRI) in typically developing children focuses on educational settings, with the primary outcome of learning [33,50,99]. Other commonly studied

outcomes relate to children's engagement with the robot [1,18,57], or their perceptions of the robot as a social agent [7,45,87,103]. Nonetheless, there is a growing subset of research which also targets children's trust in robots (see [95] for a narrative review). However, studies in this area demonstrate a large amount of heterogeneity, both in how trust is defined and measured. Terms such as closeness, rapport, and friendship on the one hand, and reliability, credibility, and competence on the other are used interchangeably, and often overlap with other constructs such as engagement, social presence, and anthropomorphism. Consequently, developing a comprehensive understanding of what factors actually influence children's trust in robots, and how, is difficult.

Within psychology, multiple facets of trust can be defined. Several models conceptualising trust in organisations have been proposed (e.g., [70,71]). Similarly, game theory and behavioural economics also use trust-based paradigms such as the ultimatum game [76] or the prisoner's dilemma [2] to determine trust in an interaction partner. Models of interpersonal trust have also been described in [81] where trust is defined as “an expectancy held by an individual or a group

---

Rebecca Stower and Natalia Calvo-Barajas have contributed equally to this study.

---

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 765955 ANIMATAS.

---

✉ Rebecca Stower  
r.stower@jacobs-university.de

Natalia Calvo-Barajas  
natalia.calvo@it.uu.se

<sup>1</sup> Department of Psychology and Methods, Jacobs University, Bremen, Germany

<sup>2</sup> Department of Information Technology, Uppsala University, Uppsala, Sweden

that the word, promise, verbal or written statement of another individual or group can be relied upon” (pg. 1).

Across these theories, two clear domains of trust emerge. These broadly coincide with the universal dimensions of social cognition; warmth and competence [28]. The first domain of trust pertains to the more social or interpersonal elements of trust, regarding perceived benevolence, or affective relationships [70,71,81]. The second relates more to perceived competency or reliability (in terms of expectations of performance), such as is captured in cognitive trust theories [54,71]. Occasionally a third dimension, integrity, is also identified, which is related to the perceived moral standards of the agent [46]. However, the extent to which this can be considered separate to social trust is questionable, as judgements of whether someone can keep a secret, for example, necessarily encompass a moral judgement [10].

This distinction between social and competency trust has also been echoed in HRI [68]. Malle et al. [68,92] have identified four main dimensions of trust in robots (reliable, capable, ethical, and sincere), which they argue can be further grouped into two main dimensions; capacity trust (reliable, capable) and moral trust (ethical, sincere). To a certain degree, trust in a (social) robot’s capacity is akin to trust in other types of autonomous intelligent devices, such as autonomous cars [38,56]. However, in such devices social trust typically does not play a role, unlike with social robots, where interpersonal elements begin to be introduced. By social robot, we mean robots that exhibit human social characteristics such as emotional expression, dialogue, gaze, gestures, and/or personality [29]. As such, when understanding trust in social robots, both social trust and competency trust need to be taken into account. Although research in this area is growing more popular with adults (e.g., [30,73,82]), how to conceptualise children’s trust in social robots remains ambiguous.

There are two main issues related to the current state of research which makes determining children’s trust in social robots difficult. The first relates to variation in definitions and measurements used, in particular regarding the distinction between social and competency trust, the second relates to inconsistent findings regarding the effects of robot behaviour and embodiment on trust. Both of these are potentially able to be addressed by a meta-analysis, where the true magnitude of an effect is able to be determined. Subsequently, the goal of this meta-analysis is both to begin bridging the gap between qualitative and quantitative syntheses in cHRI, as well as develop a comprehensive understanding of the role of trust in children’s relationships with social robots.

### 1.1 Studies Investigating Trust in cHRI

Within cHRI, trust is somewhat difficult to define. Some cHRI studies (e.g., [31,95]) draw inspiration from Rotter [81] in regards to conceptualising interpersonal trust as relying

on another’s word or promises. Van Straten et al. [95] further combine this definition with “the reliance on another person’s knowledge and intent” [53], thus encompassing elements of both social and competency trust. In other work they also define interpersonal and technological trust separately [96]. Henkel et al. [39] distinguish between three sub-categories of what they define as social confidence, of which “general trust” is one, the others being “social judgement” and “privacy”, which potentially incorporates some elements of integrity or morality. Several cHRI studies do not provide a definition of trust at all, and contain only generic questions regarding trust such as “I would trust [the robot]” [20,65]. This lack of context makes it difficult to identify how children are truly interpreting the meaning of trust, which might differ from the experimenters’ intentions. The tendency in these works is to group trust with other social constructs such as empathy and acceptance, suggesting trust in cHRI so far has been conceptualised largely based on social factors, rather than reliability or competency. Studies which do focus more on the competency element of trust in cHRI, although less prevalent, typically refer to “trust [in] the system” [44] or belief in the system’s helpfulness [57].

It is also likely that the two domains of social and competency trust are related, with positive judgements creating a so called “halo effect” [51,55]. That is, children may use multiple sources of information to make an evaluation about trustworthiness [17,36,53]. Social trust may also overlap with other elements of social cognition and social learning. This means that evaluations of children’s social trust, competency trust, and other social judgements—liking or friendship—might be difficult to disentangle. Understanding the extent to which each exists as its own distinct construct can therefore help create a clearer understanding of what outcomes are actually being affected, and are important, in children’s interactions with social robots.

For the purposes of this review we define social trust according to the definitions set out by Refs. [31,95], that is, as *a belief that the person/agent will keep their word or promises*. Comparatively, we define competency trust as *the perceived competence or reliability of the person/agent* [53]. Although perceived integrity of robots is an interesting topic, within the current state of cHRI research it falls more under the domain of anthropomorphism and mind perception (as regards evaluating agents intentions) [45,85], and as such is outside the scope of the current review.

### 1.2 Factors Affecting Trust

In their seminal 2011 meta-analysis, Hancock et al. [34] identified 3 factors relating to human–robot trust; human-related, robot-related, and environmental factors. Of these, robot-related factors were identified as the most important. In a recent revision of this meta-analysis, this distinction was

again echoed, with robot attributes and performance identified as key elements in the development of human–robot trust [35]. However, although these meta-analyses provide a useful overview of the definitions and distinctions of trust within HRI, they have yet to be replicated within the context of cHRI.

Many studies on trust in cHRI manipulate either the embodiment or behaviour of the robot and measure subsequent effects on trust [44,64,65,91]. Other studies focus more on comparisons between a human and a robot [11,12,39,106,107]. There is also increasing focus towards understanding interactions with robots who commit errors (‘faulty robots’). All of these manipulations fall under the category of ‘robot-related factors’ as defined by Hancock et al. [34,35]. Some studies have also investigated child-related factors, such as age [45,72], or context-related factors, such as the role of the robot [19], the presence of other children, [86] or multiple interactions over time [57,104]. However, as there are too few studies to properly investigate these factors in the context of a meta-analysis, and robot-related factors have been identified as the most important in adult human–robot trust, this meta-analysis will focus on robot-related factors.

### 1.2.1 Robot Embodiment

Studies on robot embodiment can be broken down into two categories; studies comparing a human and a robot, and studies comparing different robot embodiments. Studies comparing human and robot interviewers often fail to find any concrete or systematic differences between a human interviewer [11,12,39,106,107]. Findings regarding the effect of robot embodiment on trust can be equally uninformative, with studies reporting little to no differences between physical and virtual embodiments [44,64] or between different robot embodiments [20]. The extent to which these findings reflect methodological problems (e.g., small sample sizes, ceiling effects[5]) rather than a true lack of effect is currently unclear (that is, there is absence of evidence rather than evidence of absence). Thus, combining the results of studies on embodiment in the form of a meta-analysis can provide insight into the actual effect of robot embodiment on trust.

### 1.2.2 Robot Behaviour

Studies which evaluate a robot’s behaviour, such as emotional expression, are also somewhat inconsistent. Tielman et al. [91] found that a non-affective robot was rated higher on social trust than an affective robot. Ligthart et al. [63] also found that children were less intimate with an energetic robot compared to a less energetic robot. Conversely, Breazeal et al. [14] found that a contingent robot was trusted more than a non-contingent robot. One core difference in these studies, however, is the former both dealt with measures of social

trust, and the latter with competency. As such, differentiating between social and competency trust when evaluating these experiments may provide some insight into these inconsistent findings.

### 1.2.3 Robot Errors

A subset of research in cHRI also focuses on the impact of robot errors. For example, Geisskovitch et al. [31] did not find consistent differences between reliable and unreliable robots on measures of both social and competency trust. Weiss et al. [101] investigated the credibility of a robot which provided a hint during a game that was later revealed to be either correct or incorrect, and found a significant drop in credibility when the robot was incorrect. Lemaignan et al. [60] compared different types of robot errors (getting lost, making a mistake, or disobeying) on children’s engagement and anthropomorphism, and found that a robot which disobeyed was less anthropomorphised than one which made a mistake or got lost. Yadollahi et al. [108] manipulated the type of error a robot made and identified children’s ability to recognise the robot’s mistakes (learning-by-teaching). Finally, Zguda et al. [109] performed a qualitative analysis of children’s reactions to a robot which showed a malfunction, suggesting that children either did not notice or did not care about the error of the robot. However, although robot errors were a manipulated variable in all these studies, quantitative measurements of trust were only included in the first two [31,101]. Nonetheless, it may still be interesting to review robot errors to determine if an effect on social trust, competency trust, or pro-social attitudes can be identified.

## 1.3 Current Meta-analysis

The aforementioned research demonstrates that comparing findings across cHRI studies is difficult, as any difference (or lack thereof) between a robot and a human, two different robots, the same robot displaying different behaviours, or a robot making errors cannot be directly attributed to any one robot-related feature. One way of overcoming this is by considering a general category of more human-like versus more robot-like features (similar to the categorisation used in [6]). In the human-like category, we consider more social conditions (embodiment, behaviour or errors), compared to more mechanical or baseline/control conditions in the robot-like category. In doing so, an understanding of the precise effects of robot behaviour, embodiment, and errors on social and competency trust can be developed.

With this research we aim to set a common ground for children’s trust in social robots by focusing on five main contributions, (i) defining competency and social trust and their role in cHRI, (ii) distinguishing between social trust and other pro-social attitudes such as liking and friendship, (iii) analy-

sis of robot-related factors that may influence both children's social and competency trust in social robots, (iv) identification of different methods used to measure children's trust in robots and the interpretation of them by the researchers, and (v) insights into how children interact with and perceive social robots, therefore developing a deeper understanding of how trust may potentially relate to other high-level outcomes such as learning and engagement.

## 2 Method

We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Protocols for this meta-analysis.<sup>1</sup> We also used the example excel template and R-scripts provided by the Stanford MetaLab.<sup>2</sup> The documentation outlining the decision process, inclusion and exclusion criteria, included papers, extracted statistics and R-scripts can be found on Github.<sup>3</sup>

### 2.1 Data Sourcing

Searches were conducted on a variety of multidisciplinary databases, including Web of Science, Scopus, IEEE Xplore and PsycINFO. Both journal articles and conference proceedings were considered. We also scanned the reference lists of relevant studies, as well as the recent review by Van Straten et al. [95] to identify additional papers. Papers already known to the authors were also consulted, and included when relevant. Authors who had published frequently in the field were contacted to ask for any unpublished or file-drawer data, however, this did not yield any new data. We also considered non-peer reviewed sources such as unpublished PhD theses and studies published on arXiv/PsyArXiv, however, this did not uncover any additional results.

Given the relatively recent emergence of cHRI as a field, we did not restrict our searches by year. Searches were constrained to English language papers. The following key terms were decided on for the final searches: [robot AND child\* AND (trust OR secret OR share OR disclos\* OR friend OR error OR mistake\*) AND NOT (autism OR asd OR diabetes)]. See Table 1 for a list of specific search parameters and number of papers identified from each database. The last search was conducted in January 2020.

### 2.2 Study Selection

Similar to [95], due to the heterogeneity of key words in cHRI, we conducted a two-step data screening process. First,

<sup>1</sup> <http://www.prisma-statement.org/>.

<sup>2</sup> <http://metalab.stanford.edu/>.

<sup>3</sup> <https://github.com/natycalvob/meta-analysis-trust-cHRI.git>.

**Table 1** Search parameters and number of results for each database

Database	Parameters	Number of results
Web of Science	Title/abstract/keywords	208
Scopus	Title/abstract/keywords	216
IEEE	Metadata	101
psycINFO	All fields	19

the title and the abstract of all extracted papers were screened, and irrelevant papers discarded. Any papers which could not be evaluated from the title and abstract alone had the full-text paper checked for eligibility. Any disagreements or ambiguities around papers were resolved through discussion between the two principal authors.

After the initial screening, 100 papers were retained. At this step, we conducted in-depth full-text screening to confirm eligibility and refine inclusion criteria. Studies which dealt with qualitative data, contained only proposals of studies, or were theoretical or technical were excluded. For papers which reused data (e.g., a conference paper that was later published as a journal article) we used only the paper with the most detailed statistical reporting. We further excluded papers where children did not directly interact with a robot (e.g., watching videos), or only interacted with a virtual agent. We retained papers which compared virtual and physical embodiment, so long as the experiment included at least one condition with a physically embodied social robot. Studies with functional robots (e.g., [27]) were not considered.

Studies which dealt with adults or only teenagers were excluded. Some studies included data from both children and teenagers, in these cases either only the child data was included, or if this data was not available, the means for the combined age groups were used. Non-typically developing children (e.g., children diagnosed with ASD) or children with medical conditions such as diabetes or cerebral palsy were excluded. Although children with diabetes are typically-developing, we identified two reasons for the exclusion of these studies. First, diabetes is a chronic disease that affects the physical and emotional state of a child causing anxiety, stress, and discomfort [47], and second, the studies were carried out in hospitals, which is a very different experimental setting from the other included studies. Studies from these populations sometimes also included a control group comparison. We decided to also exclude these groups, as we were primarily interested in studies manipulating factors related to the interaction with the robot, rather than studies which compared different populations of users. As such, we updated the search terms to exclude any children with ASD or diabetes, see Sect. 2.1.

Next, we divided the remaining papers into those eligible for qualitative analyses and those for quantitative analysis.

Studies without a manipulation or independent variable were retained only for qualitative analyses. Studies which did not explicitly measure trust, or measured trust only as part of a broader construct like engagement, anthropomorphism, or perceived social other, as well as papers which only contained observations from teachers or parents rather than direct measurements from the children themselves were also only considered qualitatively. Following this process, 20 papers were retained for quantitative analysis. See Fig. 1 for a flow chart depicting the screening process (adapted from [74]).

### 2.3 Data Extraction

Papers were at first divided between the principal authors and data extraction was completed independently, with consultation for any unsure cases. Later, all eligible papers were revised by the principal authors together.

All citation information was extracted, including the authors, title, and year of the study. The age range and mean age of children in the study, the country in which the study

was conducted, and the location of the study (e.g., school, laboratory) were also recorded. The independent variables, dependent variables, and design of the studies (between or within groups) were then identified. For the independent variables, the levels of the variable were also recorded. For the dependent variables, we also considered how the data was collected (i.e., via subjective or objective methods), the type of data (e.g., questionnaire, interview, behavioural), the length of the interaction (in minutes), and the structure of the interaction (e.g., game, learning task, interview between robot and child). We also classified the studies as either short-term (single interaction, including multiple time points within a single interaction session) or long-term (multiple interactions over multiple time points). For studies which included multiple interactions (either within groups or multiple time points) we considered only the length of a single interaction session.

The type of robot used in the study (robot type), as well as how the robot was operated (robot operation) was also recorded. We categorised the robots used as either humanoid

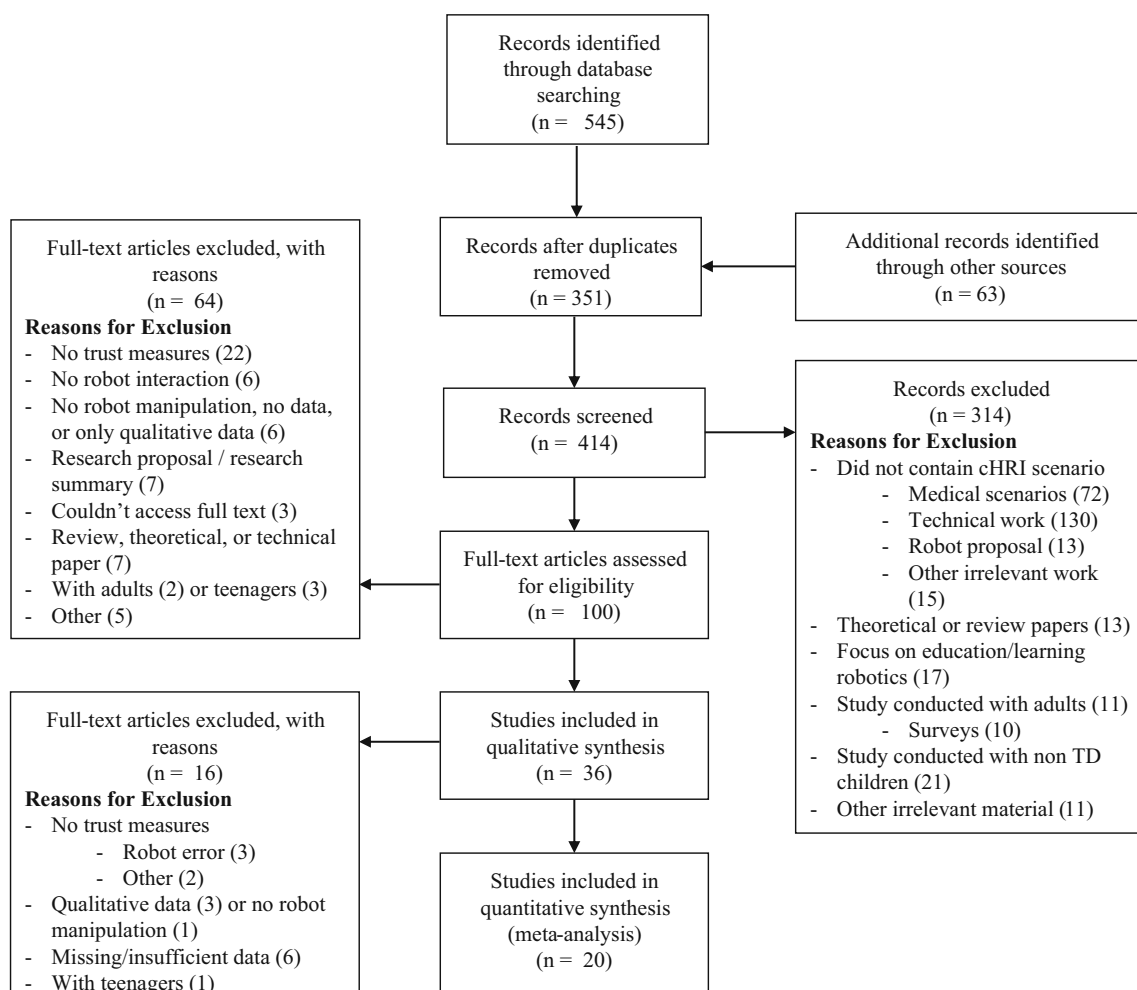


Fig. 1 PRISMA flow chart of included studies



or non-humanoid. The non-humanoid category was comprised of both zoomorphic and caricatured robots (as defined by Fong et al. [29]), due to the low number of studies in each of these categories, and the fact that all the robots concerned still contained some degree of anthropomorphic features. We classified the operation of the robot as autonomous, semi-autonomous, or Wizard of Oz (WoZ) depending on the level of control needed to operate the robot. Autonomous meant with the exception of starting the program, the robot acted without interference from the experimenters during the interaction, whereas semi-autonomous meant some elements of the interaction still required manual input. WoZ indicates a complete remote operation of the robot by a person (usually the experimenter).

As described in Sect. 1, trust varies in its definition across fields and purposes. This variation also affects the way trust is measured. We classified the type of trust measurement into two groups, objective measures, which refers to the assessment of the participant behaviour in the sense that participants are not aware of what the measurement outcome reflects (e.g., behavioural observation) [24], and subjective measures, where participants are aware of the outcome that is assessed (e.g., self-report, interviews).

All relevant reported statistical information was then extracted. Where the mean age of children in each group (for between groups studies) or the total mean age (for within groups studies) was reported, we used this in our analyses. However, a number of studies only reported the age range. For these cases, we estimated a mean based on the age range. For studies with a between groups design, we then used either the reported or estimated mean age for all groups, depending on what was available. Where the mean length of the interaction was reported, we extracted this, otherwise we again calculated an average based on the maximum and minimum range reported.

For studies which did not report effect sizes, contained insufficient information to compute them, or were otherwise missing relevant information, the corresponding author was contacted to request either the summary scores (means and SDs) for the relevant variables, or where available, the raw data. We adapted the email templates for author contact suggested by Moreau and Gamble [75]. If we did not receive a response within 1 month, we sent follow up emails both to the corresponding author and any other authors with up-to-date contact information. Papers where the authors did not reply, no longer had access to the data, or were otherwise unable to fulfill our request were then excluded from the quantitative review (one paper, [101], was still included although the data was not available for all variables as it reported other effect sizes of interest). Data was classified as extracted from either the paper, figures, or the authors. See Table 2 for a summary of the 20 included studies.

### 2.3.1 Measures of Social Trust

Social trust was found to be primarily grouped into two main methods; self-disclosure and secret-keeping, which mirrors the developmental literature on trust in general [10,80]. Measures of self-disclosure were often based on behavioural measurements such as length and type of responses [19,63], number of words or pieces of information [104,106,107], or revealing a secret to the robot [12]. Secret keeping, on the other hand, was more based on the robots perceived social fidelity using self report questions such as “do you think [the robot] could keep your secrets” or “would you tell a secret to [the robot]” [91]. For our analyses, we combined both self-disclosure and secret-keeping together, as both fit the definition of “*belief that [the robot] will keep their word or promises*”.

### 2.3.2 Measures of Competency Trust

Measures of competency trust showed a similar division between observational methods, such as forced-choice measures between agents in endorsing labels [14,31] or following instructions [48] and self-report measures asking about the perceived helpfulness of the robot [11,39,44]. We again decided to combine these measures as both fell under the definition of competency trust relating to “*perceived competency or reliability of the robot*”.

Studies on trust in cHRI may also do so in the context of the selective trust paradigm [51]. In this task, children are typically presented with two agents; one reliable and one unreliable. The reliable agent gives correct labels for familiar items, whilst the unreliable agent gives incorrect labels (e.g., labelling a “cup”, a “ball”). Then, the two agents give conflicting names for a novel, unfamiliar item and the child is asked which label they endorse. For a review on the use of the selective trust paradigm in psychology see [66].

### 2.3.3 Measures of Pro-social Attitudes

We also made the decision to create a separate category of effects which asked about pro-social attitudes such as being friends with the robot or liking the robot. While it is likely that social trust is at least partially captured by these constructs [10], they are not direct measurements and constitute separate constructs on their own. Nonetheless, as the overlap between these concepts and social trust is theoretically interesting, we included studies from the screening process which used measures of pro-sociality to conduct preliminary analyses comparing these with measures of social trust. For example, we included measures that asked if the child liked the robot (e.g., [11]) or if the child could be friends with the robot (e.g., [60]). Liking and friendship both represent general pro-social attitudes towards the robot, and can

**Table 2** Studies contributing effect sizes to the meta-analysis

	Study details			Interaction details			Robot details		Trust measurement details		
	Sample size	Age range (years)	Country	Experiment location	Interaction type	Study type	Interaction length (min)	Robot type	Robot operation	Outcome	Trust classification
Bethel et al. [12]	14 and 29	4–5	USA	Daycare center	Game	Short-term	15	RoboKind R25 and NAO	WoZ	Self-disclosure	Social
Bethel et al. [11]	60	8–12	USA	Laboratory	Interview	Short-term	20	NAO	WoZ	Helpfulness	Competency
Breazeal et al. [14]	17	3–5	USA	Preschool	Interaction	Short-term	6	DragonBot	WoZ	Self-disclosure	Social
Chandra et al. [19]	40	6–8	Portugal	School	Learning task	Short-term	15–20	NAO	WoZ	Approval	Competency
Geiskkovitch et al. [31]	17	3–5	Canada	Laboratory	Interaction	Short-term	10	NAO	WoZ	Liking	Liking*
Henkel et al. [39]	86	8–12	USA	Laboratory	Interview	Short-term	30	RoboKind R25 and NAO	WoZ	Self-disclosure	Social
Jones et al. [44]	51	11–13	England	School	Learning task	Short-term	4	NAO	Semi-autonomous	Reliability task	Competency
Kahn et al. [45]	60	9–15	USA	Laboratory	Game	Short-term	15	Robovie	WoZ	Liking	Liking*
Kennedy et al. [48]	26	7–8	England	Primary school	Learning task	Short-term	7–10	NAO	Semi-autonomous	Follow instructions	Competency
Leite & Lehman [58]	28	4–10	USA	Laboratory	Game	Short-term	60	Jimmy (Torso)	WoZ	Liking	Liking*
Leite et al. [57]	16	8–9	Portugal	School	Game	Long-term	10–25	iCAT	Autonomous	Helpfulness	Competency
Lemaignan et al. [60]	26	4–5	Switzerland	Laboratory	Game	Short-term	13	Ranger	WoZ	Friendliness	Liking
Lighthart et al. [63]	75	8–11	Netherlands	School	Interview	Short-term	15	NAO	Autonomous	Self-disclosure	Social
Melson et al. [72]	72	7–15	USA	Laboratory	Game	Short-term	45	AIBO	Autonomous	Social companion	Liking*
Tielman et al. [91]	18	8–9	Netherlands	School	Game	Short-term	10–20	NAO	WoZ	Self-disclosure	Social
Weiss et al. [101]	171	6–17	Austria	Shopping Mall	Game	Short-term	–	RS Media	WoZ	Honesty	Competency
Westlund & Breazeal [52]	86	3–8	USA	Laboratory	Learning task	Short-term	20	Tega	Semi-autonomous	Helpfulness	Liking*
Westlund et al. [104]	44	4–7	USA	Laboratory	Learning task	Long-term	20	Tega	Autonomous	Rapport Acceptance	Social
Wood et al. [106]	21	7–9	England	School	Interview	Short-term	4–12	KASPAR	WoZ	Self-disclosure	Social
Wood et al. [107]	20	7–9	England	School	Interview	Short-term	5.8	KASPAR	WoZ	Self-disclosure	Social

\*Reported measures for liking-social

be considered a reflection of underlying positive attitudes. Although liking and friendship themselves may also be considered separate constructs, given the high overlap between these terms this category is intended as a general first step towards understanding how trust relates to other relevant relationship-formation constructs.

However, as these variables were not a focus of our inclusion criteria, these studies are not a complete representation of the work in this area and thus the results presented here regarding liking and friendship cannot be considered as a comprehensive meta-analysis.

## 2.4 Meta-analytic Models

We planned four separate multivariate mixed-effects meta-regression models. First, we ran meta-analyses for social trust and competency trust individually, examining the effects of moderators on each. We then also ran a third, combined multivariate meta-analysis where we examined social and competency trust together. Finally, we ran an additional multivariate meta-analysis comparing measures of social trust with those of liking and friendship. We conducted all analyses in R using the *metafor* package [98].<sup>4</sup>

All effect-size comparisons were done using Cohen's *d* as the standardised measure to compare means. Effect sizes were calculated first from means and SDs, or when these were not available, from *t* or *F* values. We computed effect size variance by using sample size to weight the effect size according to study precision, wherein studies with larger samples were given higher weight. Most of the studies did not report correlations to calculate effect sizes, for these cases we imputed the correlations following the methodology proposed by Stanford Metalab.<sup>5</sup>

Several studies measured trust by using different outcomes (e.g., secret-keeping, following suggestions), or used multiple comparison groups, thereby creating dependencies between effect sizes. In order to account for multiple dependent effect sizes per study, we designed a multilevel structure where effect sizes from the same study were nested within a study-level factor, and effect sizes from different studies were assumed to be independent [9,32,77]. To capture these dependencies, we computed an approximation of the variance–covariance matrix based on the correlations between the effect sizes for each study [100]. We imputed the variance–covariance matrix for correlated effects by using the *clubSandwich* package in R.<sup>6</sup>

We formulated a random-effects multivariate meta analytic model including the weights of the effect sizes and the variance–covariance matrix for sampling errors. In other

words, we implemented a hierarchical linear model using random variables as the model parameters, where we analyzed the outcomes dependency and variance per study and how those are related to their population. To do so, we used the [inner | outer] formula, where the outcomes with the same values of the *outer* grouping factor (study) are assumed to be dependent, and therefore share correlated random effects corresponding to the values of the *inner* group factor (measures) [98]. To specify the variance structure corresponding to the *inner* factor, we implemented a compound symmetric structure for a single variance component corresponding to all the levels of the *inner* factor.

In addition, we conducted analyses for each of our identified moderators—age, interaction type, interaction length, robot type, robot operation, robot-related factors, and type of measure. The analyses followed the same structure described before, but including the interaction between the dependent variable (i.e., competency and/or social trust and/or liking) and the moderator. Note that for categorical moderators, this is equivalent to conducting a subgroup analysis on the individual papers within each category.

For all analyses, we first created a funnel plot as a method for estimating if there is bias in the existing publications (i.e., the tendency for only significant results to be published) [13,23,26,61]. The funnel plots each effect size against its standard error, where an asymmetrical funnel typically indicates a degree of publication bias. Larger (and therefore usually more precise) studies tend towards the top of the funnel, whereas less precise studies with greater variance are at the bottom. Studies missing from one side of the bottom of the funnel suggest that studies with smaller or non-significant effects and lower sample sizes are less likely to be published.

We also created forest plots which represent each effect size relative to its confidence intervals [13]. The effect size weight is represented by the size of the circle, and confidence intervals are shown by the length of the whiskers. The diamond reflects the overall effect size.

There are two sources which can cause variation between effect sizes; random error and true variation due to systematic differences in study design, sample and/or measurements, the latter which is known as heterogeneity [13]. Tests of heterogeneity therefore separate these causes and provide an indication of how much of the variance between studies is due to true differences in effect size. We conducted three estimates of heterogeneity, *Cochrane's Q*,  $\tau^2$  and  $I^2$ . A significant *Q* test suggests a significant amount of the variance between effect sizes can be explained by unaccounted for between study differences.  $I^2$  represents the proportion of variance in the model that can be explained by these unaccounted for factors, whereas  $\tau^2$  represents a more absolute value of between study variance [13,42,97].

We conducted post-hoc power analyses for each meta-analysis, determining both what the power is to detect an

<sup>4</sup> <http://www.metafor-project.org/doku.php>.

<sup>5</sup> <http://metalab.stanford.edu/>.

<sup>6</sup> <https://CRAN.R-project.org/package=clubSandwich>.



**Table 3** Summary of competency trust, social trust and liking variables

	Total	Independent variables			Robot related categories			
		Robot	Context	Child	Behaviour	Embodiment	Error	Human
Total studies	20	1	3	2**	6	3	3 <sup>†</sup>	5 <sup>†</sup>
Studies competency trust	7*	6**	1	1**	1	2	2 <sup>†</sup>	2 <sup>†</sup>
Studies social trust	10*	8	2	–	2	–	1 <sup>†</sup>	5 <sup>†</sup>
Studies liking-social	8*	7	–	1	3	1	1	2 <sup>†</sup>
Total measures	52	36	7	9	12	5	8	11
Measures competency trust	19	12	1	6	3	4	3	2
Measures social trust	20	14	6	–	5	–	2	7
Measures liking-social	13	10	–	3	4	1	3	2

\*Two studies reported measures for liking, competency trust (CT), and social trust (ST), and one study for both CT and ST

\*\*One study reported measures for robot and child related variables

<sup>†</sup>Studies which reported more than one robot-related measure

effect given the current median sample size, as well as, what sample size would be needed for each extracted effect to detect a significant effect with at least 80% power [8,23,37]. Power was calculated using the *pwr* package in R.<sup>7</sup> Lack of power is problematic because it can lead to type II error (false negatives), where an effect is incorrectly assumed to be absent when it may in fact be present in the population [15]. Somewhat counter-intuitively, low power also increases the likelihood of type I error (false positives) when significant effects are reported. When measuring psychological constructs in particular, larger sample sizes are often needed as effect sizes are typically smaller [78,84], potentially due to the abstract / indirect nature of many measurements.

### 3 Results

A total of 52 effect sizes were extracted from the 20 identified studies; corresponding to 20 effects for social trust, 19 for competency trust, and 13 for pro-social attitudes. The overall age of children ranged from 3–17 ( $M = 8.25$ ,  $SD = 2.81$ ). The length of interactions ranged from 4–60 min ( $Mdn = 17$ ). See Table 3 for a summary of the included studies.

After reviewing the eligible papers, we grouped the independent variables into three main categories; robot-related (e.g., embodiment and behaviour), child-related (e.g., age, gender), and context-related (e.g., time, role of the robot), see Table 3. However, due to the low number of studies on child-related and context-related factors, we decided to run the meta-analysis only for studies involving robot-related factors.

Within the robot related category, embodiment could be further broken down into effects of embodiment of the robot and effects comparing a robot and a human. Robot behaviour

was separated into effects of the behaviour of the robot, and effects of robot errors.

This method of comparing robot conditions with multiple alternatives was also used in [6]. In all cases, we defined the first condition as the more robot-like condition, and the second condition as the more human-like or social condition. For the case of robot errors, we classified the error condition with the human-like conditions, as previous studies have found robots which commit errors are perceived as more human-like [82]. Thus, a positive effect size reveals a positive effect of human-like embodiment or behaviour, whereas a negative effect size implies the opposite.

#### 3.1 Social Trust

We ran a multivariate mixed-effects meta-analytic regression model using the effect sizes for social trust. Effect sizes and their variances were calculated as described in Section 2.4. As moderators, we entered age, interaction type, interaction length, robot operation, and type of measure. All studies included for the social trust analysis used a humanoid robot, hence for this model we did not consider robot type as a moderator.

The overall meta-analytic model was non-significant, and adding moderators did not account for any additional variance. However, there was a trend of the type of measure suggesting that objective measures have a positive effect on social trust for human-like features ( $p = 0.080$ ). The results from all meta-analytic models for social trust including moderators can be seen in Table 4.

The funnel plot for social trust is depicted in Fig. 2. The funnel is slightly asymmetrical, suggesting some slight publication bias could exist. There was non-significant between study heterogeneity in effect sizes  $Q(13) = 14.97$ ,  $p = 0.309$ . Although overall estimates for  $\tau^2$  and  $I^2$  for all models were quite low, indicating a low between study variation in effect

<sup>7</sup> <https://CRAN.R-project.org/package=pwr>.

**Table 4** Meta-analytic regression for social trust

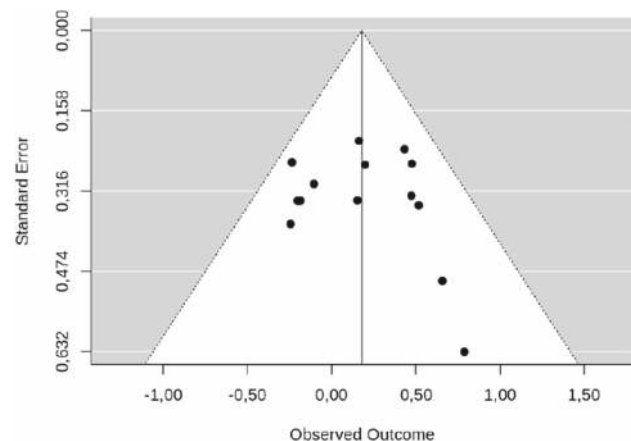
	Est. [CI 95%]	SE Est.	z	p value	QM(df)	$\tau^2$	$I^2$
Model social trust				0.077 <sup>†</sup>	–	0.023	20.991
Social trust	0.179 [–0.019, 0.377]	0.101	1.768	0.077			
Age				0.090 <sup>†</sup>	–	0.025	21.730
Age × Social trust	0.020 [–0.003, 0.042]	0.012	1.696	0.090			
Interaction type				0.147 <sup>†</sup>	5.359(3)	0.025	23.422
Game × Social trust	0.085 [–0.224, 0.393]	0.158	0.538	0.591			
Interaction × Social trust	0.694 [–0.168, 1.555]	0.440	1.578	0.115			
Interview × Social trust	0.208 [–0.046, 0.461]	0.129	1.607	0.108			
Interaction length				0.170 <sup>†</sup>	–	0.031	25.759
Length × Social trust	0.008 [–0.003, 0.019]	0.006	1.374	0.170			
Robot operation				0.153 <sup>†</sup>	3.758(2)	0.025	21.305
Autonomous × Social trust	0.432 [–0.120, 0.984]	0.282	1.534	0.125			
WoZ × Social trust	0.129 [–0.084, 0.341]	0.108	1.186	0.236			
Robot related factors				0.122 <sup>†</sup>	5.800(3)	0.023	21.456
Behaviour × Social trust	0.253 [–0.084, 0.590]	0.172	1.473	0.141			
Error × Social trust	0.694 [–0.167, 1.554]	0.439	1.580	0.114			
Human × Social trust	0.135 [–0.113, 0.383]	0.127	1.066	0.287			
Type of measure				0.147 <sup>†</sup>	3.835(2)	0.023	21.457
Objective × Social trust	0.268 [–0.032, 0.568]	0.153	1.751	0.080			
Subjective × Social trust	0.118 [–0.145, 0.581]	0.134	0.877	0.380			

\*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$

† Overall  $p$ -value for test moderators

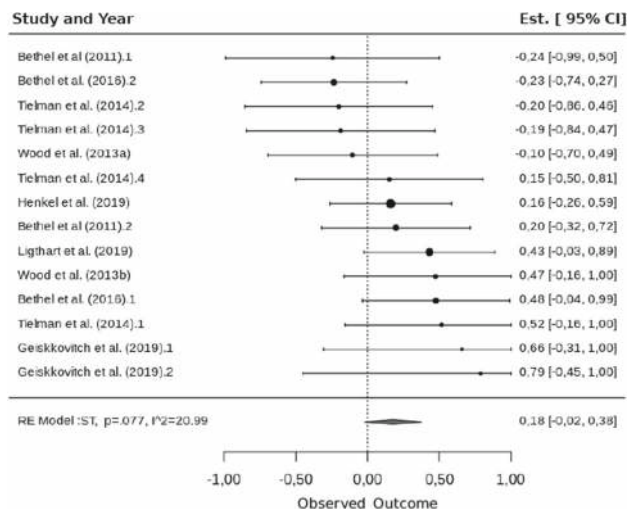
sizes, many of the included studies had low precision (i.e., large confidence intervals; see Fig. 3 which depicts the effect sizes relative to their standard variation) which may obscure the effect of between study differences in the true effect size [13].

We next ran a post-hoc power analysis. For between groups studies with average  $N = 34$ , observed power was 11%, and for within groups studies (average  $N = 19$ ) it was again 11%, suggesting these studies were severely under-



**Fig. 2** Funnel plot for social trust

powered. Using the extracted average effect size of  $d = 0.18$ , the number of participants that are required to detect a significant effect with 80% power is 491 per condition for between groups studies, and 247 per condition for within.



**Fig. 3** Forest plot for social trust

### 3.2 Competency Trust

We used the same multivariate mixed-effects meta-analytic regression model as described in Sect. 2.4 using the effect sizes for competency trust, see Table 5. A forest plot depicting the effect sizes relative to their standard variation can also be seen in Fig. 5.

The funnel for competency trust (see Fig. 4) is asymmetrical and shows a large amount of spread even at the tip of the funnel. This suggests both the presence of some bias, and that studies with higher sample sizes, which should theoretically have greater precision, still had large variation (this is also reflected in the wide confidence intervals in Fig. 5).

The between study heterogeneity for all analyses fell into the “high” range for  $I^2$  as defined by [43], and the overall test for heterogeneity was significant  $Q(11) = 174.70, p \leq 0.001$ . Given the large amount of variation in study designs, methods, and samples, this result is not surprising. Furthermore, high heterogeneity suggests the influence

of moderators on the effect of robot related factors on trust [13].

The overall meta analytic effect size estimate for competency trust was non-significant. However, after accounting for the effect of moderators, there was a significant effect of interaction type and robot related factors. Regarding interaction type, only children who interacted with the robot in a game context showed a significant effect of human-like attributes on competency trust. For the robot related factors, children who interacted with a robot who committed an error had a significant increase in competency trust, compared to those who interacted with a non-faulty robot.

We again conducted a post-hoc power analysis. Given the current average sample sizes of  $N = 40$  and  $N = 17$  for between and within groups, respectively, the observed power is 11% and 10%. In order to reach 80% power with the identified average effect size of  $d = 0.16$ , 558 participants are required per condition for between groups studies, and 281 per condition for within groups.

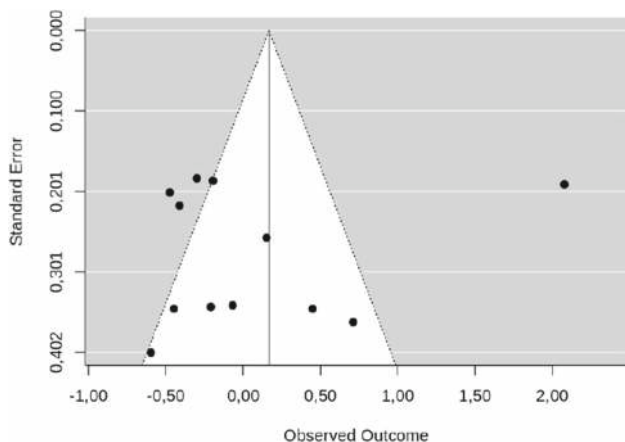


Fig. 4 Funnel plot for competency trust

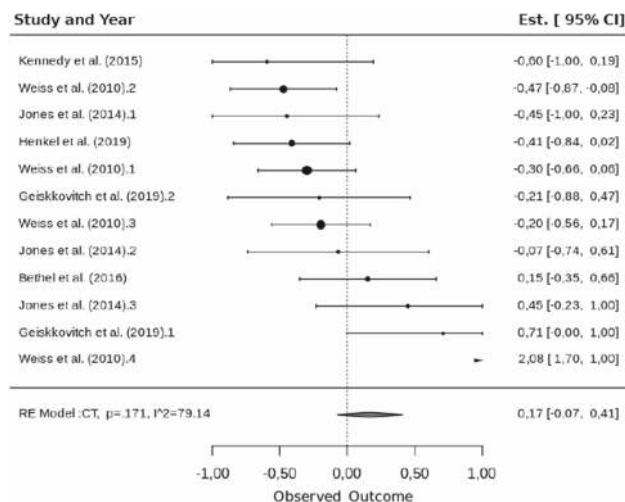


Fig. 5 Forest plot for competency trust

### 3.3 Social and Competency Trust

Next, we ran a combined multivariate mixed-effects regression model using both social and competency trust as outcomes, see Table 6. The variance–covariance matrix between social and competency trust was calculated as described in Sect. 2.4.

The funnel plot is again slightly asymmetrical, suggesting a small degree of bias, see Fig. 6. The between study heterogeneity in all cases was low-moderate, but still significant  $Q(23) = 53.42, p \leq 0.001$ .

Without moderators, the overall meta analytic effect size was significant for competency trust, but not social trust. Moderators were significant for competency trust only. Younger children showed greater competency trust in the robot than older ones. The type of interaction with the robot was again significant, however, the direction was reversed from the model with only competency, in this case suggesting that where children played a game with the robot, human-like attributes had a *negative* effect on competency trust. The length of the interaction was non-significant at the  $p < 0.05$  level, but suggested a trend towards shorter interactions leading to higher competency trust. The robot operation moderator was significant, with experiments where the robot was wizarded showing a negative effect of human-like attributes on trust. Also, when accounting for the manipulation of robot related factors, the behaviour of the robot influenced trust such that more human-like behaviour lead to significantly *lower* competency trust. Finally, the significant effect of type of measure regarding subjective measures also had a reverse direction suggesting these measures triggered a *negative* effect on competency trust for human-like attributes.

**Table 5** Meta-analytic regression for competency trust

	Est. [CI 95%]	SE Est.	z	p value	QM(df)	$\tau^2$	$I^2$
Model competency trust				0.171 <sup>†</sup>	–	0.212	79.144
Competency trust	0.168 [–0.073, –0.408]	0.123	1.368	0.171			
Age				0.147 <sup>†</sup>	–	0.208	78.739
Age × Competency trust	0.016 [–0.006, 0.037]	0.011	1.450	0.147			
Interaction type				0.217 <sup>†</sup>	5.764(4)	0.172	77.915
Game × Competency trust	0.274 [0.026, 0.523]	0.127	2.163	0.031*			
Interaction × Competency trust	0.202 [–0.448, 0.852]	0.332	0.610	0.542			
Interview × Competency trust	–0.220 [–0.909, 0.459]	0.352	–0.626	0.531			
Learning task × Competency trust	–0.111 [–0.494, 0.272]	0.196	–0.567	0.571			
Interaction length				0.351 <sup>†</sup>	–	0.218	79.256
Length × Competency trust	0.002 [–0.021, 0.025]	0.012	0.196	0.844			
Robot operation				0.270 <sup>†</sup>	2.620(2)	0.187	77.947
Semi-auto × Competency trust	–0.111 [–0.496, 0.274]	0.196	–0.595	0.572			
WoZ × Competency trust	0.186 [–0.054, 0.426]	0.123	1.517	0.129			
Robot related factors				0.001 <sup>†***</sup>	67.782(4)	0.648	91.954
Behaviour × Competency trust	–0.308 [–1.869, 1.253]	0.796	–0.387	0.699			
Embodiment × Competency trust	–0.111 [–1.499, 1.278]	0.708	–0.157	0.876			
Error × Competency trust	1.800 [0.397, 3.204]	0.716	2.515	0.012*			
Human × Competency trust	–0.220 [–1.438, 0.998]	0.621	–0.354	0.723			
Type of measure				0.399 <sup>†</sup>	1.838(2)	0.217	80.409
Objective × Competency trust	0.018 [–0.577, 0.613]	0.303	0.060	0.952			
Subjective × Competency trust	0.174 [–0.078, 0.426]	0.129	1.355	0.176			

\*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$

† Overall  $p$  value for test moderators

The observed power to detect effects was again low, 12% for between groups studies, average  $N = 38$ , and 11% for within groups, average  $N = 18$ . Conversely, to detect an effect of ( $d = 0.17$ ) as identified in the model, 491 participants are needed per condition for a between groups study, and 247 per condition for within groups.

### 3.4 Social Trust and Liking

For the final meta-analysis, we combined papers on social trust with those measuring liking and/or friendship and ran an additional multivariate mixed-effects regression model, see Table 7. Again, the variance–covariance matrix between social trust and liking was calculated as described in Sect. 2.4. A forest plot of effect sizes relative to their standard variation for only those studies which contained measures of liking can be seen in Figure 7.

The funnel plot is mostly symmetrical (See Figure 8), indicating low or minimal publication bias. The  $Q$ -test for heterogeneity was significant  $Q(22) = 84.24$ ,  $p < 0.001$ , suggesting the differences in effect sizes between studies may be accounted for by differences in study-specific variables (moderators).

The overall meta analytic effect size was significant only for liking. Introducing moderators into the model accounted for significant additional variance for liking. Age had a significant effect on liking, suggesting younger children liked the robots more. The type of interaction was again also significant, with studies where the children played a game with the robot showing a negative effect of human-like attributes on liking. Interaction length was also significant, with shorter interactions again leading to greater liking. For the robot operation, both the categories of WoZ and autonomous operation were significant moderators of liking, such that in both cases human-like attributes showed a negative effect on liking. Studies which compared a robot and a human showed that the robot was liked significantly more. Similarly, for studies which manipulated embodiment, a more robot-like embodiment was preferred. Finally, concerning type of measure, objective measures had a reverse direction suggesting that for studies where these kind of measures were used, more human-like features lead to lower liking.

Observed power for between groups studies with average  $N = 27$  was 49%, and for within groups with average  $N = 23$  was 69%. Whilst this is an improvement compared to the previous meta-analytic models, it still falls short of

**Table 6** Meta-analytic regression for social and competency trust

	Est. [CI 95%]	SE Est.	z	p-value	QM(df)	$\tau^2$	$I^2$
<i>Model competency + social trust</i>				0.010 <sup>†**</sup>	9.299(2)	0.041	48.044
Competency trust	−0.256 [−0.445, −0.067]	0.096	−2.655	0.008**			
Social trust	0.179 [−0.046, 0.404]	0.115	1.560	0.119			
<i>Age</i>				0.008 <sup>†**</sup>	9.637(2)	0.043	49.471
Age × Competency trust	−0.024 [0.041, −0.007]	0.009	−2.773	0.006**			
Age × Social trust	0.020 [−0.006, 0.045]	0.013	1.492	0.136			
<i>Interaction type</i>				0.067 <sup>†</sup>	13.213(7)	0.050	56.378
Game × Competency trust	−0.308 [−0.517, −0.099]	0.107	−2.888	0.004**			
Game × Social trust	0.085 [−0.284, 0.454]	0.188	0.450	0.653			
Interaction × Competency trust	0.202 [−0.425, 0.830]	0.320	0.632	0.528			
Interaction × Social trust	0.694 [−0.279, 1.667]	0.496	1.397	0.162			
Interview × Competency trust	−0.220 [−0.685, 0.244]	0.237	−0.929	0.353			
Interview × Social trust	0.208 [−0.083, 0.499]	0.149	1.399	0.162			
Learning task × Competency trust	−0.111 [−0.581, 0.359]	0.240	−0.462	0.644			
<i>Interaction length</i>				0.074 <sup>†</sup>	5.215(2)	0.050	52.463
Length × Competency trust	−0.016 [−0.032, 0.000]	0.008	−1.939	0.053			
Length × Social trust	0.008 [−0.005, 0.021]	0.007	1.198	0.231			
<i>Robot type</i>				0.010 <sup>†**</sup>	9.299(2)	0.024	48.044
Humanoid × Competency trust	−0.256 [−0.445, −0.067]	0.096	−2.655	0.008**			
Humanoid × Social trust	0.179 [−0.046, 0.404]	0.115	1.560	0.119			
<i>Robot operation</i>				0.041 <sup>†*</sup>	9.967(4)	0.040	49.085
Autonomous × Social trust	0.432 [−0.172, 1.036]	0.308	1.402	0.161			
Semi-auto × Competency trust	−0.111 [−0.589, 0.368]	0.244	−0.454	0.650			
WoZ × Competency trust	−0.268 [−0.468, −0.067]	0.102	−2.615	0.009**			
WoZ × Social trust	0.129 [−0.113, 0.370]	0.123	1.044	0.297			
<i>Robot related factors</i>				0.071 <sup>†</sup>	13.059(7)	0.047	54.950
Behaviour × Competency trust	−0.308 [−0.524, −0.092]	0.110	−2.791	0.005**			
Behaviour × Social trust	0.253 [0.144, 0.650]	0.203	1.249	0.212			
Embodiment × Competency trust	−0.111 [−0.583, 0.362]	0.241	−0.460	0.645			
Error × Competency trust	0.202 [−0.425, 0.830]	0.320	0.632	0.527			
Error × Social trust	0.694 [−0.278, 1.666]	0.496	1.399	0.162			
Human × Competency trust	−0.220 [−0.679, 0.238]	0.234	−1.941	0.347			
Human × Social trust	0.135 [−0.151, 0.420]	0.149	0.925	0.355			
<i>Type of measure</i>				0.035 <sup>†*</sup>	10.323(4)	0.043	49.909
Objective × Competency trust	0.018 [−0.506, 0.542]	0.267	0.068	0.946			
Objective × Social trust	0.268 [−0.067, 0.603]	0.171	1.567	0.117			
Subjective × Competency trust	−0.271 [−0.465, −0.076]	0.099	−2.731	0.006**			
Subjective × Social trust	0.118 [−0.187, 0.422]	0.155	0.758	0.448			

\*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$

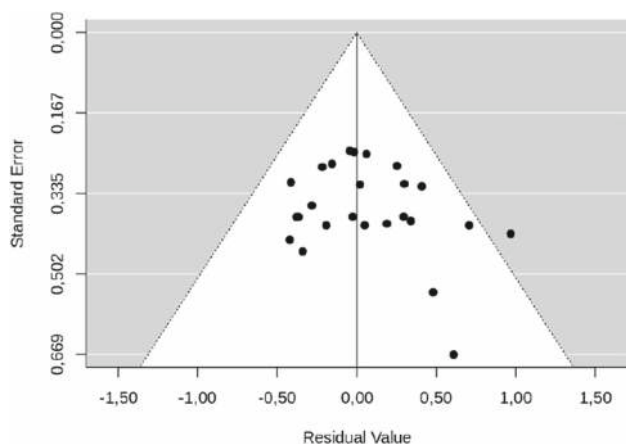
† Overall  $p$ -value for test moderators

the required 80% power threshold. To reach 80% power with an identified moderate effect of  $d = 0.54$ , 55 participants are needed per condition for between groups, and 29 per condition for within groups. The number of required participants is lower than in previous models, as the overall effect size is larger [21].

## 4 Discussion

Neither the meta-analysis for competency trust nor social trust individually showed overall significant effects. Looking at the moderators, competency trust showed some effects of interaction type and robot errors, whereas social trust showed





**Fig. 6** Funnel plot for competency and social trust

a trend only for the moderator of measurement type. When we combined both social and competency trust in a multivariate meta-analysis the overall model for competency trust was significant, including for all moderators. The combined model for social trust and liking revealed a significant overall effects on liking only. Measurement type was again identified as a potential moderator for social trust.

There was an *overall negative effect of human-like attributes on competency trust*. This could suggest that in all cases, more human-like behaviour or embodiment negatively impacts competency trust. This effect is not totally surprising, and aligns with previous findings which suggest social behaviour in robots may not always be beneficial for outcomes such as learning [49,99]. It is possible that robots which are seen as more human-like create greater expectations for performance, thus leading to a decrease in competency trust if or when these expectations are not met. Alternatively, robots with more machine-like embodiment could improve impressions of competence due to associations with machine-like intelligence, especially in cases where there are no social elements associated with the task.

Competency trust could also overlap to some extent with learning [54]. Thus, the negative relationship between human-like behaviours and competency trust may provide a potential explanation for some of the previously somewhat counter-intuitive findings regarding the effect of robot social behaviours on learning. Future work should focus on the disentanglement of competency trust from objective learning outcomes, in determining how robot behaviour and embodiment affects both separately.

The absence of any effect for social trust is surprising, given the predominant focus on social and interpersonal trust in the literature (e.g., [14,39,40,52]). One explanation for this is that there was more variation in studies measuring social trust, which can be seen by comparing Figs. 3 and 5. This variation could mask any effects on social trust, as

studies are less precise regarding the true effect size. This variance could also reflect the overlap of social trust with other, related social constructs, suggesting that these studies were not successfully tapping into this element of trust.

## 4.1 Moderators

All of the models except for social trust indicated moderate-high heterogeneity, suggesting differences in effect sizes between studies are due to systematic differences rather than random error. We attempted to account for these differences by identifying a set of study-specific moderators which could be influencing the results. Each moderator (age, length of interaction, interaction type, robot type, robot operation, type of measure, and robot-related factors) is discussed individually below.

### 4.1.1 Age

In the combined social and competency model, *age was significantly negatively associated with competency trust*. If this result is reflective of a true effect, then it would be unsurprising, as younger children (ages 6–10) previously have been found to score higher on domains such affective reactions [69], conceptualising the robot as a social other [45], and persuasion and credibility [101] when interacting with social robots. This would suggest a general trend of younger children being more susceptible to manipulations of robot embodiment and behaviour. However, the fact that this effect was seen in the competency trust outcome alone, and that the effect size is very small ( $-0.024$ ) could suggest the presence of a spurious effect, especially given low power can not only lead to type II error, but increases the possibility of type I error being present where significant effects are detected [15].

### 4.1.2 Interaction Type

In the model focusing on competency trust alone, studies which involved the child playing a game with the robot showed a positive effect of human-like attributes on competency trust. However, in the combined social and competency trust model, the direction of this effect was reversed. Due to the lack of any stable effect, no strong conclusion regarding the effect of interaction type can be drawn as to how it influences trust.

In the combined social trust and liking model, the negative effect of human-like attributes was mimicked for liking. This result is surprising, as although social behaviours may be detrimental for learning [49,99], they are typically associated with other positive factors such as valence and enjoyment [33,44,104]. Only two studies reported effects of both competency trust and liking together, [11,39], and as such, we

**Table 7** Meta-analytic regression for social trust and liking

	Est. [CI 95%]	SE Est.	z	p-value	QM(df)	$\tau^2$	$I^2$
<i>Model social trust + liking</i>				0.042 <sup>†*</sup>	6.332(2)	0.216	74.632
Social Trust	0.179 [−0.186, 0.544]	0.186	0.961	0.336			
Liking Social	−0.540 [−0.995, −0.085]	0.232	−2.326	0.020*			
<i>Age</i>				0.003 <sup>†***</sup>	11.431(2)	0.151	66.815
Age × Social trust	0.020 [−0.016, −0.055]	0.018	1.069	0.285			
Age × Liking social	−0.080 [−0.128, −0.031]	0.024	−3.246	0.001***			
<i>Interaction type</i>				0.041 <sup>†*</sup>	14.637(7)	0.192	73.087
Game × Social trust	0.085 [−0.296, 0.465]	0.194	0.436	0.663			
Game × Liking social	−0.937 [−1.779, −0.094]	0.430	−2.180	0.029*			
Interaction × Social trust	0.694 [−0.336, 1.723]	0.525	1.320	0.187			
Interaction × Liking social	−0.052 [−1.143, 1.038]	0.556	−0.094	0.925			
Interview × Social trust	0.208 [−0.221, 0.637]	0.219	0.949	0.343			
Interview × Liking social	−1.009 [−1.715, −0.302]	0.361	−2.797	0.005**			
Learning task × Liking social	0.080 [−0.429, 0.589]	0.260	0.307	0.759			
<i>Interaction length</i>				0.039 <sup>†*</sup>	6.466(2)	0.240	76.063
Length × Social trust	0.008 [−0.013, 0.029]	0.010	0.765	0.444			
Length × Liking social	−0.020 [−0.036, −0.004]	0.008	−2.507	0.012*			
<i>Robot type</i>				0.003 <sup>†***</sup>	13.845(3)	0.129	63.270
Humanoid × Social trust	0.179 [−0.093, 0.451]	0.139	1.291	0.197			
Humanoid × Liking social	−0.877 [−1.446, −0.307]	0.290	−3.018	0.003**			
Zoomorphic × Liking social	−0.434 [−0.828, −0.039]	0.201	−2.155	0.031*			
<i>Robot operation</i>				0.025 <sup>†*</sup>	12.798(5)	0.185	70.255
Autonomous × Social trust	0.432 [−0.528, 1.392]	0.490	0.882	0.378			
Autonomous × Liking social	−1.097 [−2.019, −0.175]	0.471	−2.331	0.020*			
Semi-auto × Liking social	0.080 [−0.523, 0.682]	0.307	0.259	0.795			
WoZ × Social trust	0.129 [−0.224, 0.481]	0.180	0.716	0.474			
WoZ × Liking social	−0.712 [−1.263, −0.161]	0.281	−2.533	0.011*			
<i>Robot related factors</i>				<0.001 <sup>†***</sup>	30.010(7)	0.074	51.104
Behaviour × Social trust	0.253 [−0.205, 0.711]	0.234	1.084	0.279			
Behaviour × Liking social	0.095 [−0.258, 0.447]	0.180	0.526	0.599			
Embodiment × Liking social	−1.097 [−1.748, −0.446]	0.332	−3.301	<0.001***			
Error × Social trust	0.694 [−0.287, 1.674]	0.500	1.386	0.166			
Error × Liking social	0.286 [−0.486, 1.058]	0.394	0.727	0.467			
Human × Social trust	0.135 [−0.216, 0.486]	0.179	0.753	0.451			
Human × Liking social	−1.009 [−1.595, −0.422]	0.299	−3.370	<0.001***			
<i>Type of measure</i>				0.186 <sup>†</sup>	6.182(4)	0.250	76.918
Objective × Social trust	0.268 [−0.304, 0.840]	0.292	0.919	0.358			
Objective × Liking social	−0.943 [−1.861, −0.026]	0.468	−2.015	0.044*			
Subjective × Social trust	0.118 [−0.414, 0.650]	0.271	0.434	0.664			
Subjective × Liking social	−0.290 [−0.846, 0.265]	0.283	−1.025	0.305			

\*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$

† Overall  $p$ -value for test moderators

did not conduct a meta-analysis combining these factors. Consequently, further research is needed investigating the relationship between social trust, competency trust, and lik-

ing to determine whether the negative effect of human-like attributes replicates.

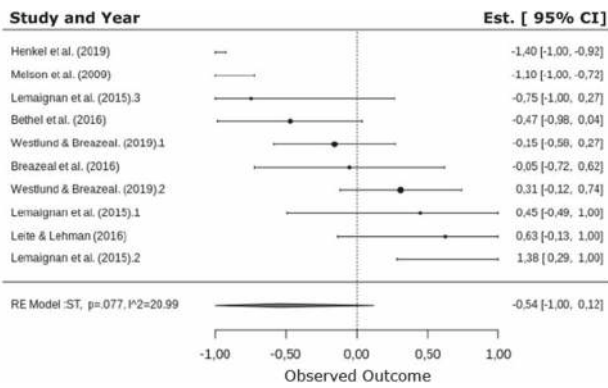


Fig. 7 Forest plot for liking

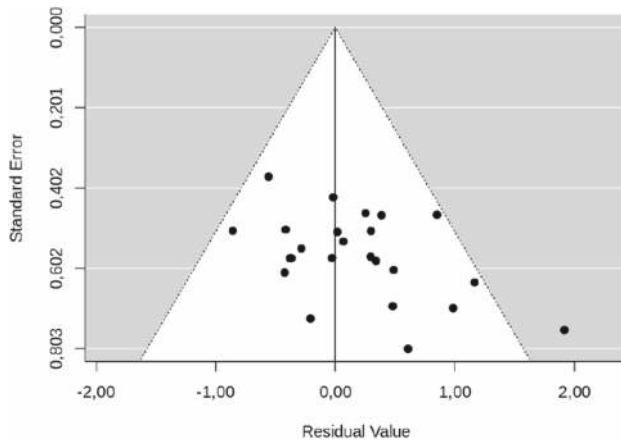


Fig. 8 Funnel plot for social trust and liking

#### 4.1.3 Interaction Length

Interaction length also influenced competency trust, such that *shorter interactions may lead to higher competency trust*. It is possible that during shorter analyses initial expectations are able to be maintained, whereas as interactions increase in length there are more opportunities for the robot to commit an error or otherwise perform sub-optimally, therefore leading to a decrease in competency trust. Manipulating the length of a single interaction session as a variable could therefore be interesting, as it may provide an indication of how long it takes for impressions of trust in robots to form. It is also possible that, should there be multiple, repeated interactions, trust would either increase, or revert to baseline levels. However, as only two studies included in the analysis were conducted long-term, we were unable to test this hypothesis. In the future, more long-term studies which focus on the development of trust over several repeated interactions would be beneficial.

#### 4.1.4 Robot Type

The majority of studies were conducted with humanoid robots, primarily using NAO (SoftBank Robotics). As such, we were unable to conduct a proper investigation into how physical robot embodiment influences trust. This limits the extent to which our results can be generalised, as how children interact with NAO may not be representative of social robots in general. This is relevant given the finding that more human-like attributes decrease competency trust, as its possible that more zoomorphic or functional robot designs have different effects on trust, especially if expectations regarding performance change for these embodiments. We therefore motivate future researchers to test different robot embodiments and evaluate their effects on trust perception.

#### 4.1.5 Robot Operation

Robot operation as a moderator was only significant in the combined social and competency trust model. This result suggested that *for studies where the robot was wizarded, human-like attributes lead to a negative effect on competency trust*. However, again this result should be interpreted with caution. Wizard of Oz was the most common method of operation, and thus we cannot rule out the possibility of type I error (false positive). However, should this result reflect a true effect, it could suggest that wizarded robots embody some features which autonomous or semi-autonomous robots do not, and which lead to an incongruity between human-like attributes and competency trust. For example, wizarded robots may be more contingent or socially responsive compared to autonomous systems, thus echoing the idea that human-like behaviours are not always beneficial for competency trust.

Given the current technological limitations of social robots, and especially with children as a user group [5], using a wizarded robot is not inherently problematic [88]. Wizard of Oz designs can allow for manipulation of features that are not yet technically possible, as well as guide towards features to focus on in future autonomous programming. Nonetheless, human control over the robot could allow for combinations of behaviour that are not necessarily realistic [79], as well as introduce issues with transparency and perceived agency, especially as regards children as users [40,102,103]. Thus, as technological capacities progress, comparing WoZ to autonomous or semi-autonomous robot designs will allow for a more nuanced understanding of how and if having a human controller affects different aspects of the interaction, such as trust.

#### 4.1.6 Robot Related Factors

Another moderator we considered was the type of robot related factor which was manipulated. *Robots which committed an error showed a significant positive effect on trust in the competency trust model only, such that robots which made an error were trusted more.* This is a surprising finding, and conflicts with studies in adults that suggest errors violate competency trust [73,82,83]. It is possible that children perceive and react to errors differently than do adults [5]. However, there was only a very small number of effect sizes which contributed to this relationship ( $N = 3$ ), which, given the high heterogeneity and small sample sizes, may not have been sufficient to conduct specific subgroup analyses. Thus, more investigation needs to be undertaken regarding the specific effects of robot errors on children's trust.

The other robot related factor to show an effect was robot behaviour, which showed a *negative effect of human-like behaviour on competency trust* in the combined social and competency trust model. This finding again aligns with the overall idea that human-like behaviour is not necessarily always beneficial in child robot interaction.

#### 4.1.7 Type of Measure

The type of measure showed a trend for social trust. This result revealed that *children who are not aware of the fact that an outcome is being measured show a higher degree of social trust in the robot.* The majority of the studies classified as objective measures are related to self-disclosure, suggesting that robotic systems promote interactions where children feel more comfortable revealing personal information toward the robot. This attitude contributes to the development of a closer relationship with the robot [19,22].

Potentially, self-disclosure itself may be an imprecise measurement of social trust, as it is difficult to quantify and may show significant individual variation between participants [22,105]. Developing more precise measurements of social trust, and contrasting them with other, related measures should thus be a focus of future research. In fact, [94] recently published a series of scales intended to distinguish between closeness to, trust in, and perceived social support in a robot. This study represents a first step towards the development of standardised scales, however, as with any measure, its success will depend on the adoption and use by other researchers in the field.

Measure type also had a significant effect on the combined social and competency trust model, suggesting that *for studies where children's trust was measured objectively, more human-like robots were perceived as less competent.* Finally, the significant effect found on the combined social trust and liking model suggest that there is a *negative effect of objec-*

*tive measures that decreases the perceived likability of the robot for more human-like features.*

The findings from the 'measurement type' moderator demonstrate the importance of using adequate procedures and measures at assessing trust in cHRI studies. These results are consistent with social psychology studies suggesting that subjective and objective measures elicit different responses [24]. We believe that objective measures of trust allow for capturing more natural interactions between the child and the robot, as well as reducing biases introduced by subjective measures.

#### 4.2 Qualitative Observations

From a qualitative perspective, all of the studies relating to robot errors focused on the robot's reliability, with errors directly linked to task performance. No studies were found which related to non-task related errors or social errors. A possible exception is [58], where the robot could be argued to make a social error by revealing information it was not supposed to know. However, in this study they only analysed children's affective responses, and did not measure trust directly. As such, an additional valuable direction for future research relates to how social errors, such as social faux pas, impact trust. For example, it is possible that social errors would be more related to social trust, and by manipulating these or other social norm violations we would begin to see more of an effect on social trust [89]. Consequently, a distinction between social and competency errors, and how these differentially affect both social and competency trust would be useful.

In the two studies conducted long term [57,104], both found children's perceptions of the robot to remain positive or increase over time. The 'novelty effect' in HRI is gathering increasingly more attention [4,59], as researchers try to determine whether the effects found in short-term interactions generalise across repeated encounters. This is especially relevant in the domain of trust, where children may form immediate impressions of trustworthiness [16]. As such, long-term studies will be of crucial importance to understand the trajectory of trust development, breakdown, and recovery in cHRI.

#### 4.3 Methodological Considerations and Recommendations

A common theme among these analyses was a serious lack of power. The median sample size in the included studies for within participants was  $N = 18$  and for between participants  $N = 60$ , which is far too low to reach the recommended threshold of 80% power, especially considering the average effect sizes found in the data range from small to medium

[21]. Thus, these analyses reveal a trend in cHRI publications towards under-powered analyses.

Consequently, rather than suggesting that any given predictor or moderator does (or does not) affect trust, these results are more indicative of a broader problem relating to insufficiently powered study designs in cHRI. That being said, recruiting children for HRI studies is often difficult, which is undoubtedly the reason behind the universally low sample sizes. Studies in cHRI could instead benefit from simpler experimental designs, where a smaller sample size is required to detect effects. In the case of more complex designs, sufficient statistical power should be accounted for a priori in the design and conceptualisation of the study (e.g., through software such as G\*Power [25] or the *pwr* package in R). In all cases, power analyses should be conducted a priori. cHRI researchers could also focus more towards international collaborations, such as the Many Labs project [90].

Notwithstanding, the required sample sizes determined by the post-hoc power analyses may still be an over-estimation, as all effect size estimates were quite small [21]. As such, it is possible that if true effects do exist, they would be larger and thus detectable with a smaller sample size (as is seen in the post-hoc power analysis for social trust and liking).

A second theme which emerged amongst the results relates to the high confidence intervals, reflecting low precision around effect size estimates. Again, this is telling of a broader underlying issue in cHRI relating to inconsistent study designs, measurements, and moderators. One reason for this could be the lack of common ground in defining constructs, which is discussed as a general problem in HRI in [4]. This lack of consistency means both that studies which on the surface appear to measure the same outcome may in fact be measuring separate constructs, and that replication of any one finding is difficult. Thus, the field of cHRI would benefit from a corpus of standardised definitions and measurements, especially as it relates to more psychological constructs such as trust.

As is also discussed in [95], the reporting of statistical information in cHRI studies can be lacking, in particular for non-significant findings. We further confirm this assessment through a quantitative evaluation of cHRI studies in the domain of trust. To facilitate clear and transparent interpretation of findings, for all hypotheses the relevant statistical tests and effect sizes should be reported. Where possible, all means and standard deviations for identified conditions and outcomes should also be available [either in the paper itself, or, if space is limited, published elsewhere such as in supplementary materials, on GitHub, or the Open Science Framework (OSF)].

When conducting and reporting analyses, a clear distinction should also be made between hypotheses which were made a priori, and post-hoc exploratory analyses. This

includes reporting significance for individual items in a scale versus the total scale, the reliability of scales, splitting data into groups, or reporting effects or interactions that were not initially hypothesised. Transparency around reporting of non-hypothesised results will minimise the chance of type-II error (false-negatives) and allow for greater reproducibility of results. This has the benefit of increasing precision of reported effects, and thus comparing results across studies and determining true effect sizes will be easier.

Finally, all studies included in the review came from WEIRD populations (Western, Educated, Industrialised, Rich, Democratic) [41]. One study was considered which compared populations of Dutch and Pakistani children [86], however, it contained insufficient statistical information to be included in the quantitative review. Notwithstanding, given that children's trust can show cultural variation [67], and that there is cultural variation in attitudes towards robots [3,62], comparing children's trust in robots cross culturally is a valuable direction for future work.

## 5 Limitations

One of the major benefits of a meta-analysis is the ability to combine the results of several studies to conduct more powerful analyses [13]. However, like any other analysis a meta-analysis itself still has limits to its statistical power [37,93]. Given the small number of eligible studies, the high heterogeneity between them, and the small sample sizes, specific subgroup analyses were more difficult to conduct. This meant not all of the identified moderators (e.g. robot operation, type of robot, interaction length) were able to be investigated with sufficient detail. In addition, we were unable to investigate other factors which may influence trust, such as child related factors or context related factors, due to a lack of studies in these areas.

Trust itself is also a highly complex and nuanced concept. Making clear distinctions between the two identified domains of social and competency trust is not always straightforward. Similarly, the boundary between social trust, liking, and friendship is also blurry, especially when it comes to behavioural measures such as self-disclosure, which potentially tap into several underlying constructs [22,105]. The analysis of pro-social attitudes such as liking and friendship was also not comprehensive, as these terms were not the focus of our search criteria. As such, we welcome other researchers in the field to come together to create a more cohesive understanding of the role of trust in cHRI as a whole, and to continue the development of methods and measures through which such relationships can be captured.

The field of cHRI is still in its infancy, which made conducting a comprehensive meta-analysis at this point



difficult, and limited our ability to draw more outspoken conclusions. Nonetheless, we hope the methodological and conceptual findings from this research will form the groundwork for more future high-quality research, upon which a more thorough meta-analysis can build on. In addition, we hope the issues highlighted here can also guide researchers towards which elements are important in the design and realisation of future experiments in this area.

## 6 Conclusion

In sum, this meta-analysis provides a concrete conceptualisation of trust in two domains, social and competency. There was a general finding that human-like behaviours do not always positively influence competency trust or liking. These results contribute to a growing body of evidence which challenges the idea that embodying robots with human-like behaviour is always beneficial. Several moderating factors were also found to influence competency trust and/or liking; including age, interaction length, and robot operation. For social trust, only one moderator showed a trend towards significance, type of measure, which revealed that whether the measure used to capture trust was subjective or objective influences the direction of effect.

In addition, the meta-analysis uncovered several directions for improvement when it comes to the design and implementation of cHRI experiments. Whilst any one individual study may comprise good experimental design (as many of the studies included here have), generalising across these studies is difficult due the low sample sizes, low power, and high standard variations. Current and future researchers in cHRI should take on the challenge of improving results and replication by designing studies with larger sample sizes, consistent definitions and transparent statistical reporting.

**Acknowledgements** We would like to thank Sho Tsuji for help and advice with conducting the meta-analysis. In addition, we thank all of the authors who shared their data with us and/or responded to our email questions in order to conduct the meta-analysis, including Zachary Henkel and Cindy Bethel, Jacqueline Kory-Westlund, Myrthe Tielman, Shruti Chandra, Astrid Weiss, Séverin Lemaignan, Mike Lighthart, and Denise Geiskkovitch. We also thank everyone who took the time to give feedback on the initial versions of this manuscript.

**Funding** This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 765955 ANIMATAS. Open Access funding enabled and organized by Projekt DEAL.

**Data Availability** All materials used for the preparation of this paper can be found at <https://github.com/natycalvob/meta-analysis-trust-cHRI>.

## Compliance with Ethical Standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Ahmad MI, Mubin O, Shahid S, Orlando J (2019) Robot's adaptive emotional feedback sustains children's social engagement and promotes their vocabulary learning: a long-term child-robot interaction study. *Adapt Behav* 27(4):243–266. <https://doi.org/10.1177/1059712319844182>
- Axelrod R (1980) Effective choice in the Prisoner's Dilemma. *J Conflict Resolut* 24(1):3–25. <https://doi.org/10.1177/002200278002400101>
- Bartneck C, Nomura T, Takayuki K, Suzuki T, Kato K (2005) A cross-cultural study on attitudes towards robots. In: *Human-computer interaction*. Las Vegas
- Baxter P, Kennedy J, Senft E, Lemaignan S, Belpaeme T (2016) From characterising three years of HRI to methodology and reporting recommendations. In: 2016 11th ACM/IEEE international conference on human-robot interaction (HRI). IEEE, pp 391–398. <https://doi.org/10.1109/HRI.2016.7451777>
- Belpaeme T, Baxter P, de Greeff J, Kennedy J, Read R, Looije R, Neerinx M, Baroni I, Zelati MC (2013) Child-robot interaction: perspectives and challenges. In: *International conference on social robotics (ICSR)*. pp 452–459. [https://doi.org/10.1007/978-3-319-02675-6\\_45](https://doi.org/10.1007/978-3-319-02675-6_45)
- Belpaeme T, Kennedy J, Ramachandran A, Scassellati B, Tanaka F (2018) Social robots for education: a review. *Sci Robot* 3(21):eaat5954. <https://doi.org/10.1126/scirobotics.aat5954>
- Beran TN, Ramirez-Serrano A, Kuzyk R, Fior M, Nugent S (2011) Understanding how children understand robots: perceived animism in child-robot interaction. *Int J Hum Comput Stud* 69(7–8):539–550. <https://doi.org/10.1016/j.ijhcs.2011.04.003>
- Bergmann C, Tsuji S, Piccinini PE, Lewis ML, Braginsky M, Frank MC, Cristia A (2018) Promoting replicability in developmental research through meta-analyses: insights from language acquisition research. *Child Dev* 89(6):1996–2009. <https://doi.org/10.1111/cdev.13079>
- Berkey CS, Hoaglin DC, Antczak-Bouckoms A, Mosteller F, Colditz GA (1998) Meta-analysis of multiple outcomes by regression with random effects. *Stat Med* 17(22):2537–2550. [https://doi.org/10.1002/\(SICI\)1097-0258\(19981130\)17:22<3c3c2537::AID-SIM953>3e3e3.0.CO;2-C](https://doi.org/10.1002/(SICI)1097-0258(19981130)17:22<3c3c2537::AID-SIM953>3e3e3.0.CO;2-C)
- Bernath MS, Feshbach ND (1995) Children's trust: theory, assessment, development, and research directions. *Appl Prev Psychol* 4(1):1–19. [https://doi.org/10.1016/S0962-1849\(05\)80048-4](https://doi.org/10.1016/S0962-1849(05)80048-4)

11. Bethel CL, Henkel Z, Stives K, May DC, Eakin DK, Pilkinton M, Jones A, Stubbs-Richardson M (2016) Using robots to interview children about bullying: Lessons learned from an exploratory study. In: 2016 25th IEEE international symposium on robot and human interactive communication (RO-MAN). IEEE, pp 712–717. <https://doi.org/10.1109/ROMAN.2016.7745197>
12. Bethel CL, Stevenson MR, Scassellati B (2011) Secret-sharing: interactions between a child, robot, and adult. In: 2011 IEEE international conference on systems, man, and cybernetics. IEEE, pp 2489–2494. <https://doi.org/10.1109/ICSMC.2011.6084051>
13. Borenstein M, Hedges LV, Higgins JP, Rothstein HR (2009) Introduction to meta-analysis. Wiley, Chichester. <https://doi.org/10.1002/9780470743386>
14. Breazeal C, Harris PL, DeSteno D, Kory Westlund JM, Dickens L, Jeong S (2016) Young children treat robots as informants. *Top Cognit Sci* 8(2):481–491. <https://doi.org/10.1111/tops.12192>
15. Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, Munafò MR (2013) Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci* 14(5):365–376. <https://doi.org/10.1038/nrn3475>
16. Calvo-Barajas N, Perugia G, Castellano G (2020) The effects of robot's facial expressions on children's first impressions of trustworthiness. In: The 29th IEEE international conference on robot and human interactive communication. pp 165–171. <https://doi.org/10.1109/ROMAN47096.2020.9223456>
17. Carpenter M, Nielsen M (2008) Tools, TV, and trust: introduction to the special issue on imitation in typically-developing children. *J Exp Child Psychol* 101(4):225–227. <https://doi.org/10.1016/j.jecp.2008.09.005>
18. Castellano G, Leite I, Pereira A, Martinho C, Paiva A, McOwan PW (2012) Detecting Engagement in HRI: an exploration of social and task-based context. In: 2012 International conference on privacy, security, risk and trust and 2012 international conference on social computing. IEEE, pp 421–428 (2012). <https://doi.org/10.1109/SocialCom-PASSAT.2012.51>
19. Chandra S, Alves-Oliveira P, Lemaignan S, Sequeira P, Paiva A, Dillenbourg P (2016) Children's peer assessment and self-disclosure in the presence of an educational robot. In: 2016 25th IEEE international symposium on robot and human interactive communication (RO-MAN). IEEE, pp 539–544. <https://doi.org/10.1109/ROMAN.2016.7745170>
20. Cohen I, Looije R, Neerinx MA (2014) Child's perception of robot's emotions: effects of platform, context and experience. *Int J Soc Robot* 6(4):507–518. <https://doi.org/10.1007/s12369-014-0230-6>
21. Cohen J (1992) A power primer. *Psychol Bull* 112(1):155–159. <https://doi.org/10.1037/0033-2909.112.1.155>
22. Collins NL, Miller LC (1994) Self-disclosure and liking: a meta-analytic review. *Psychol Bull* 116(3):457
23. Cristia A (2018) Can infants learn phonology in the lab? A meta-analytic answer. *Cognition* 170(9):312–327. <https://doi.org/10.1016/j.cognition.2017.09.016>
24. De Houwer J (2006) What are implicit measures and why are we using them. In: The handbook of implicit cognition and addiction. pp 11–28
25. Faul F, Erdfelder E, Lang AG, Buchner A (2007) G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods* 39(2):175–191. <https://doi.org/10.3758/BF03193146>
26. Field AP, Gillett R (2010) How to do a meta-analysis. *Br J Math Stat Psychol* 63(3):665–694. <https://doi.org/10.1348/000711010X502733>
27. Fior M, Nugent S, Beran TN, Ramírez-Serrano A, Kuzyk R (2010) Children's relationships with robots: robot is child's new friend. *J Phys Agents (JoPha)* 4(3):9–17. <https://doi.org/10.14198/JoPha.2010.4.3.02>
28. Fiske ST, Cuddy AJ, Glick P (2007) Universal dimensions of social cognition: warmth and competence. *Trends Cognit Sci* 11(2):77–83. <https://doi.org/10.1016/j.tics.2006.11.005>
29. Fong T, Nourbakhsh I, Dautenhahn K (2003) A survey of socially interactive robots. *Robot Autonom Syst* 42(3–4):143–166. [https://doi.org/10.1016/S0921-8890\(02\)00372-X](https://doi.org/10.1016/S0921-8890(02)00372-X)
30. Gaudiello I, Zibetti E, Lefort S, Chetouani M, Ivaldi S (2016) Trust as indicator of robot functional and social acceptance. An experimental study on user conformation to iCub answers. *Comput Hum Behav* 61:633–655. <https://doi.org/10.1016/j.chb.2016.03.057>
31. Geiskkovitch DY, Thiessen R, Young JE, Glenwright MR (2019) What? That's not a chair!: how robot informational errors affect children's trust towards robots. In: 2019 14th ACM/IEEE international conference on human–robot interaction (HRI). IEEE, pp 48–56. <https://doi.org/10.1109/HRI.2019.8673024>
32. Gleser LJ, Olkin I (1994) Stochastically dependent effect sizes. In: Cooper JCVH, Hedges LV (eds) The handbook of research synthesis and meta-analysis, 2nd edn. Russell Sage Foundation, New York, pp 357–376. <https://doi.org/10.7758/9781610448864.16>
33. Gordon G, Spaulding S, Westlund JK, Lee JJ, Plummer L, Martinez M, Das M, Breazeal CL (2016) Affective personalization of a social robot tutor for children's second language skills. In: Proceedings of the 30th conference on artificial intelligence (AAAI 2016). pp 3951–3957
34. Hancock PA, Billings DR, Schaefer KE, Chen JYC, de Visser EJ, Parasuraman R (2011) A meta-analysis of factors affecting trust in human–robot interaction. *Hum Factors J Hum Factors Ergon Soc* 53(5):517–527. <https://doi.org/10.1177/0018720811417254>
35. Hancock PA, Kessler TT, Kaplan AD, Brill JC, Szalma JL (2020) Evolving trust in robots: specification through sequential and comparative meta-analyses. *Hum Fact*. <https://doi.org/10.1177/0018720820922080>
36. Harris PL, Koenig MA, Corriveau KH, Jaswal VK (2018) Cognitive foundations of learning from testimony. *Annu Rev Psychol* 69(1):251–273. <https://doi.org/10.1146/annurev-psych-122216-011710>
37. Hedges LV, Pigott TD (2001) The power of statistical tests in meta-analysis. *Psychol Methods* 6(3):203–217. <https://doi.org/10.1037/1082-989X.6.3.203>
38. Hegner S, Beldad A, Brunswick GJ (2019) In automatic we trust: investigating the impact of trust, control, personality characteristics, and extrinsic and intrinsic motivations on the acceptance of autonomous vehicles. *Int J Hum Comput Interact* 35:1769–1780
39. Henkel Z, Baugus K, Bethel CL, May DC (2019) User expectations of privacy in robot assisted therapy. *Paladyn J Behav Robot* 10(1):140–159. <https://doi.org/10.1515/pjbr-2019-0010>
40. Henkel Z, Bethel CL (2017) A robot forensic interviewer. In: Proceedings of the companion of the 2017 ACM/IEEE international conference on human–robot interaction—HRI '17. ACM Press, New York, New York, USA, pp 10–20. <https://doi.org/10.1145/3029798.3034783>
41. Henrich J, Heine SJ, Norenzayan A (2010) Beyond WEIRD: towards a broad-based behavioral science. *Behav Brain Sci* 33(2–3):111–135. <https://doi.org/10.1017/S0140525X10000725>
42. Higgins JP, Thompson SG (2002) Quantifying heterogeneity in a meta-analysis. *Stat Med* 21(11):1539–1558. <https://doi.org/10.1002/sim.1186>
43. Higgins JPT (2003) Measuring inconsistency in meta-analyses. *BMJ* 327(7414):557–560. <https://doi.org/10.1136/bmj.327.7414.557>
44. Jones A, Castellano G, Bull S (2014) Investigating the effect of a robotic tutor on learner perception of skill based feedback. In: Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics),

- vol 8755. Springer International Publishing, pp 186–195. [https://doi.org/10.1007/978-3-319-11973-1\\_19](https://doi.org/10.1007/978-3-319-11973-1_19)
45. Kahn PH, Kanda T, Ishiguro H, Freier NG, Severson RL, Gill BT, Ruckert JH, Shen S (2012) “Robovie, you’ll have to go into the closet now”: children’s social and moral relationships with a humanoid robot. *Dev Psychol* 48(2):303–314. <https://doi.org/10.1037/a0027033>
  46. Kahn PH, Turiel E (1988) Children’s conceptions of trust in the context of social expectations. *Merrill Palmer Q* 34(4):403–419
  47. Kalra S, Jena BN, Yeravdekar R (2018) Emotional and psychological needs of people with diabetes. *Indian J Endocrinol Metab* 22(5):696
  48. Kennedy J, Baxter P, Belpaeme T (2015) Comparing robot embodiments in a guided discovery learning interaction with children. *Int J Soc Robot* 7(2):293–308. <https://doi.org/10.1007/s12369-014-0277-4>
  49. Kennedy J, Baxter P, Belpaeme T (2015) The robot who tried too hard. In: *Proceedings of the tenth annual ACM/IEEE international conference on human–robot interaction—HRI ’15*, 801. ACM Press, New York, New York, USA, pp 67–74. <https://doi.org/10.1145/2696454.2696457>
  50. Kennedy J, Baxter P, Senft E, Belpaeme T (2016) Social robot tutoring for child second language learning. In: *2016 11th ACM/IEEE international conference on human–robot interaction (HRI)*. IEEE, pp. 231–238. <https://doi.org/10.1109/HRI.2016.7451757>
  51. Koenig MA, Harris PL (2005) Preschoolers mistrust ignorant and inaccurate speakers. *Child Dev* 76(6):1261–1277. <https://doi.org/10.1111/j.1467-8624.2005.00849.x>
  52. Kory-Westlund JM, Breazeal C (2019) Exploring the effects of a social robot’s speech entrainment and backstory on young children’s emotion, rapport, relationship, and learning. *Front Robot AI*. <https://doi.org/10.3389/frobot.2019.00054>
  53. Landrum AR, Eaves BS, Shafto P (2015) Learning to trust and trusting to learn: a theoretical framework. *Trends Cognit Sci* 19(3):109–111. <https://doi.org/10.1016/j.tics.2014.12.007>
  54. Landrum AR, Mills CM, Johnston AM (2013) When do children trust the expert? Benevolence information influences children’s trust more than expertise. *Dev Sci* 16(4):622–638. <https://doi.org/10.1111/desc.12059>
  55. Landrum AR, Pflaum AD, Mills CM (2016) Inducing knowledge-ability from niceness: children use social features for making epistemic inferences. *J Cognit Dev* 17(5):699–717. <https://doi.org/10.1080/15248372.2015.1135799>
  56. Lee JD, See KA (2004) Trust in automation: designing for appropriate reliance. *Hum Factors* 46(1):50–80. [https://doi.org/10.1518/hfes.46.1.50\\_30392](https://doi.org/10.1518/hfes.46.1.50_30392)
  57. Leite I, Castellano G, Pereira A, Martinho C, Paiva A (2014) Empathic robots for long-term interaction: evaluating social presence, engagement and perceived support in children. *Int J Soc Robot* 6(3):329–341. <https://doi.org/10.1007/s12369-014-0227-1>
  58. Leite I, Lehman JF (2016) The robot who knew too much. In: *Proceedings of the the 15th international conference on interaction design and children—IDC ’16*. ACM Press, New York, New York, USA, pp 379–387. <https://doi.org/10.1145/2930674.2930687>
  59. Leite I, Martinho C, Paiva A (2013) Social robots for long-term interaction: a survey. *Int J Soc Robot* 5(2):291–308. <https://doi.org/10.1007/s12369-013-0178-y>
  60. Lemaignan S, Fink J, Mondada F, Dillenbourg P (2015) You’re doing it wrong! studying unexpected behaviors in child–robot interaction. In: *International conference on social robotics*, vol 1. Springer International Publishing, pp 390–400. [https://doi.org/10.1007/978-3-319-25554-5\\_39](https://doi.org/10.1007/978-3-319-25554-5_39)
  61. Lewis M, Braginsky M, Tsuji S, Bergmann C, Piccinini P (2016) A quantitative synthesis of early language acquisition using meta-analysis. <https://doi.org/10.17605/OSF.IO/HTSJM>
  62. Li D, Rau PLP, Li Y (2010) A cross-cultural study: effect of robot appearance and task. *Int J Soc Robot* 2(2):175–186. <https://doi.org/10.1007/s12369-010-0056-9>
  63. Lighthart M, Neerincx MA, Bindsbergen KLA V, Grootenhuis MA (2019) A child and a robot getting acquainted—interaction design for eliciting self-disclosure. In: *Proceeding of the 18th international conference on autonomous agents and multi-agent systems (AAMAS 2019)*. pp 8–11
  64. Looije R, Neerincx MA, de Lange V (2008) Children’s responses and opinion on three bots that motivate, educate and play. *J Phys Agents (JoPha)* 2(2):13–20. <https://doi.org/10.14198/JoPha.2008.2.2.03>
  65. Looije R, van der Zalm A, Neerincx MA, Beun RJ (2012) Help, I need some body the effect of embodiment on playful learning. In: *2012 IEEE RO-MAN: The 21st IEEE international symposium on robot and human interactive communication*. IEEE, pp 718–724. <https://doi.org/10.1109/ROMAN.2012.6343836>
  66. Lucas AJ, Lewis C (2010) Should we trust experiments on trust? *Hum Dev* 53(4):167–172. <https://doi.org/10.1159/000320044>
  67. Lucas AJ, Lewis C, Pala FC, Wong K, Berridge D (2013) Social-cognitive processes in preschoolers’ selective trust: three cultures compared. *Dev Psychol* 49(3):579–590
  68. Malle BF, Ullman D (2021) A multi-dimensional conception and measure of human–robot trust. Trust in human-robot interaction: research and applications. pp 3–25
  69. Martínez-Miranda J, Pérez-Espinosa H, Espinosa-Curiel I, Avila-George H, Rodríguez-Jacobo J (2018) Age-based differences in preferences and affective reactions towards a robot’s personality during interaction. *Comput Hum Behav* 84:245–257. <https://doi.org/10.1016/j.chb.2018.02.039>
  70. Mayer RC, Davis JH, Schoorman FD (1995) An integrative model of organizational trust. *Acad Manag Rev* 20(3):709–734. <https://doi.org/10.5465/amr.1995.9508080335>
  71. McAllister DJ (1995) Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. *Acad Manag J* 38(1):24–59. <https://doi.org/10.5465/256727>
  72. Melson GF, Kahn PH, Beck A, Friedman B, Roberts T, Garrett E, Gill BT (2009) Children’s behavior toward and understanding of robotic and living dogs. *J Appl Dev Psychol* 30(2):92–102. <https://doi.org/10.1016/j.appdev.2008.10.011>
  73. Mirnig N, Stollnberger G, Miksch M, Stadler S, Giuliani M, Tscheligi M (2017) To err is robot: how humans assess and act toward an erroneous social robot. *Front Robot AI* 4:1–15. <https://doi.org/10.3389/frobot.2017.00021>
  74. Moher D, Liberati A, Tetzlaff J, Altman DG, Altman D, Antes G, Atkins D, Barbour V, Barrowman N, Berlin JA, Clark J, Clarke M, Cook D, D’Amico R, Deeks JJ, Devereaux PJ, Dickersin K, Egger M, Ernst E, Gøtzsche PC, Grimshaw J, Guyatt G, Higgins J, Ioannidis JP, Kleijnen J, Lang T, Magrini N, McNamee D, Moja L, Mulrow C, Napoli M, Oxman A, Pham B, Rennie D, Sampson M, Schulz KF, Shekelle PG, Tovey D, Tugwell P (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. <https://doi.org/10.1371/journal.pmed.1000097>
  75. Moreau D, Gamble B (2020) Conducting a meta-analysis in the age of open science: tools, tips, and practical recommendations. <https://doi.org/10.31234/osf.io/t5dwg>
  76. Nowak M, Page K, Sigmund K (2000) Fairness versus reason in the ultimatum game. *Science* 289(5485):1773–5
  77. Pustejovsky JE, Tipton E (2016) Small sample methods for cluster-robust variance estimation and hypothesis testing in fixed effects models. <https://arxiv.org/pdf/1601.01981.pdf>. 8 Jan 2016



78. Richard FD, Bond CF, Stokes-Zoota JJ (2003) One hundred years of social psychology quantitatively described. *Rev Gen Psychol* 7(4):331–363. <https://doi.org/10.1037/1089-2680.7.4.331>
79. Riek L (2012) Wizard of Oz studies in HRI: a systematic review and new reporting guidelines. *J Hum Robot Interact* 1(1):119–136. <https://doi.org/10.5898/jhri.1.1.riek>
80. Rotenberg KJ (1980) “A promise kept, a promise broken”: developmental bases of trust. *Child Dev* 51(2):614. <https://doi.org/10.2307/1129307>
81. Rotter JB (1967) A new scale for the measurement of interpersonal trust. *J Personal* 35(4):651–665. <https://doi.org/10.1111/j.1467-6494.1967.tb01454.x>
82. Salem M, Eyssel FA, Rohlfing K, Kopp S, Joubin F (2013) To err is human(-like): effects of robot gesture on perceived anthropomorphism and likability. *Int J Soc Robot* 5(3):313–323. <https://doi.org/10.1007/s12369-013-0196-9>
83. Salem M, Lakatos G, Amirabdollahian F, Dautenhahn K (2015) Would you trust a (faulty) robot? In: Proceedings of the tenth annual ACM/IEEE international conference on human–robot interaction—HRI ’15. ACM Press, New York, New York, USA, pp 141–148. <https://doi.org/10.1145/2696454.2696497>
84. Schäfer T, Schwarz MA (2019) The meaningfulness of effect sizes in psychological research: differences between sub-disciplines and the impact of potential biases. *Front Psychol* 10(April):1–13. <https://doi.org/10.3389/fpsyg.2019.00813>
85. Severson RL, Carlson SM (2010) Behaving as or behaving as if? Children’s conceptions of personified robots and the emergence of a new ontological category. *Neural Netw* 23(8–9):1099–1103. <https://doi.org/10.1016/j.neunet.2010.08.014>
86. Shahid S, Kraemer E, Swerts M (2014) Child–robot interaction across cultures: How does playing a game with a social robot compare to playing a game alone or with a friend? *Comput Hum Behav* 40:86–100. <https://doi.org/10.1016/j.chb.2014.07.043>
87. Somanader MC, Saylor MM, Levin DT (2011) Remote control and children’s understanding of robots. *J Exp Child Psychol* 109(2):239–247. <https://doi.org/10.1016/j.jecp.2011.01.005>
88. Steinfeld A, Jenkins OC, Scassellati B (2009) The oz of wizard. In: Proceedings of the 4th ACM/IEEE international conference on Human robot interaction—HRI ’09. ACM Press, New York, New York, USA, p 101. <https://doi.org/10.1145/1514095.1514115>
89. Stower R, Kappas A (2020) “Oh no, my instructions were wrong!” An exploratory pilot towards children’s trust in social robots. In: The 29th IEEE international conference on robot and human interactive communication. pp 641–646. <https://doi.org/10.1109/ROMAN47096.2020.9223495>
90. Stroebe W (2019) What can we learn from many labs replications? *Basic Appl Soc Psychol* 41:103–91
91. Tielman M, Neerinx M, Meyer JJ, Looije R (2014) Adaptive emotional expression in robot–child interaction. In: Proceedings of the 2014 ACM/IEEE international conference on human–robot interaction—HRI ’14. ACM Press, New York, New York, USA, pp 407–414. <https://doi.org/10.1145/2559636.2559663>
92. Ullman D, Malle BF (2019) Measuring gains and losses in human–robot trust: evidence for differentiable components of trust. In: 2019 14th ACM/IEEE international conference on human–robot interaction (HRI). pp 618–619
93. Valentine J, Pigott T, Rothstein H (2011) How many studies do you need? A primer on statistical power for meta-analysis. *Qual Eng* 56:117–120
94. van Straten CL, Kühne R, Peter J, de Jong C, Barco A (2020) Closeness, trust, and perceived social support in child–robot relationship formation. *Interact Stud* 21(1):57–84. <https://doi.org/10.1075/IS.18052.str>
95. van Straten CL, Peter J, Kühne R (2019) Child–robot relationship formation: a narrative review of empirical research. *Int J Soc Robot*. <https://doi.org/10.1007/s12369-019-00569-0>
96. van Straten CL, Peter J, Kühne R, de Jong C, Barco A (2018) Technological and interpersonal trust in child–robot interaction. In: Proceedings of the 6th international conference on human-agent interaction—HAI ’18. ACM Press, New York, New York, USA, pp 253–259. <https://doi.org/10.1145/3284432.3284440>
97. Veroniki AA, Jackson D, Viechtbauer W, Bender R, Bowden J, Knapp G, Kuss O, Higgins JP, Langan D, Salanti G (2016) Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Res Synth Methods* 7(1):55–79. <https://doi.org/10.1002/jrsm.1164>
98. Viechtbauer W (2010) Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software* 36(3), 1–48. <https://doi.org/10.18637/jss.v036.i03>
99. Vogt P, van den Berghe R, de Haas M, Hoffman L, Kanero J, Mamus E, Montanier JM, Oranc C, Oudgenoeg-Paz O, Garcia DH, Papadopoulos F, Schodde T, Verhagen J, Wallbridgell CD, Willemsen B, de Wit J, Belpaeme T, Goksun T, Kopp S, Kraemer E, Kuntay AC, Leseman P, Pandey AK (2019) Second language tutoring using social robots: a large-scale study. In: 2019 14th ACM/IEEE international conference on human–robot interaction (HRI), vol 2019. IEEE, pp 497–505. <https://doi.org/10.1109/HRI.2019.8673077>
100. Wei Y, Higgins JP (2013) Estimating within-study covariances in multivariate meta-analysis with multiple outcomes. *Stat Med* 32(7):1191–1205. <https://doi.org/10.1002/sim.5679>
101. Weiss A, Scherndl T, Buchner R, Tscheligi M (2010) A robot as persuasive social actor a field trial on child–robot interaction. In: Proceedings of the 2nd international symposium on new frontiers in human–robot interaction—a symposium at the AISB 2010 convention. pp 136–142
102. Westlund JK, Breazeal C (2015) Deception, secrets, children, and robots: What’s acceptable? Workshop on the emerging policy and ethics of human–robot interaction, held in conjunction with the 10th ACM/IEEE international conference on human–robot interaction
103. Westlund JMK, Breazeal C (2016) Transparency, teleoperation, and children’s understanding of social robots. In: 2016 11th ACM/IEEE international conference on human–robot interaction (HRI). IEEE, pp 625–626. <https://doi.org/10.1109/HRI.2016.7451888>
104. Westlund JMK, Park HW, Williams R, Breazeal C (2018) Measuring young children’s long-term relationships with social robots. In: Proceedings of the 17th ACM conference on interaction design and children—IDC ’18. ACM Press, New York, New York, USA, pp 207–218. <https://doi.org/10.1145/3202185.3202732>
105. Wheelless LR, Grotz J (1977) The measurement of trust and its relationship to self-disclosure. *Hum Commun Res* 3(3):250–257. <https://doi.org/10.1111/j.1468-2958.1977.tb00523.x>
106. Wood LJ, Dautenhahn K, Lehmann H, Robins B, Rainer A, Syrdal DS (2013) Robot-mediated interviews: Does a robotic interviewer impact question difficulty and information recovery? In: Assistive technology: from research to practice, proceedings of the AAATE 2013 conference
107. Wood LJ, Dautenhahn K, Rainer A, Robins B, Lehmann H, Syrdal DS (2013) Robot-mediated interviews—How effective is a humanoid robot as a tool for interviewing young children? *PLoS ONE* 8(3):e59448. <https://doi.org/10.1371/journal.pone.0059448>
108. Yadollahi E, Johal W, Paiva A, Dillenbourg P (2018) When deictic gestures in a robot can harm child–robot collaboration. In: Proceedings of the 17th ACM conference on interaction design and children—IDC ’18. ACM Press, New York, New York, USA, pp 195–206. <https://doi.org/10.1145/3202185.3202743>
109. Zguda P, Sniezynski B, Indurkha B, Kolota A, Jarosz M, Sondej F, Izui T, Dziok M, Belowska A, Jedras W, Venture G (2019) On the role of trust in child–robot interaction\*. In: 2019 28th IEEE international conference on robot and human interactive commu-

nication (RO-MAN). IEEE, pp 1–6. <https://doi.org/10.1109/RO-MAN46459.2019.8956400>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Rebecca Stower** is a PhD student at Jacobs University Bremen. She has a Bachelor of Psychological Science (Hons I) from the University of Queensland, Australia (2016). Her research is part of the EU ITN ANIMATAS project. Her thesis centres on the conceptualisation and measurement of children's trust in robots, and in particular, the occurrence of robot errors during child-robot interactions.

**Natalia Calvo-Barajas** is a PhD student at Uppsala University. She obtained her master's degree in Robotics Engineering from the University of Genoa, Italy, and her bachelor's degree in Mechatronics Engineering from the Nueva Granada Military University, Colombia. She is part of the EU ITN ANIMATAS project. Her research focuses on modelling trust in child-robot educational interaction. She is interested in implementing machine learning models for the understanding of children's perception of trust in robots.

**Ginevra Castellano** is a Professor in Intelligent Interactive Systems at the Department of Information Technology, Uppsala University, where she leads the Uppsala Social Robotics Lab. By adopting an interdisciplinary approach grounded in principles from the social sciences, Castellano aims to develop computational social abilities that allow robots to behave in a socially intelligent way in natural interaction with humans. She has published more than 100 research papers on these topics, receiving over 3000 citations. Castellano is a recipient of the 10-Year Technical Impact Award at the ACM International Conference on Multimodal Interaction 2019.

**Arvid Kappas** is a professor of psychology and Dean at Jacobs University Bremen. He has been conducting research on emotions for over three decades. Having obtained his Ph.D. at Dartmouth College, NH, USA, he has held positions in Switzerland, Canada, the UK, and Germany. His current research interests focus on various aspects of human emotions, as well as Affective Computing and Social Robotics. He is a fellow of the Association for Psychological Science and the Society for Personality and Social Psychology. From 2013-2018 he was president of the International Society for Research on Emotion.