Sex differences in children's toy preferences: A systematic review, meta-regression and meta-analysis.

Brenda K. Todd, Department of Psychology, City, University of London, London, U.K.

Rico A. Fischer, Glasgow Caledonian University, Glasgow, UK.

Steven Di Costa, Institute of Cognitive Neuroscience, University College London, London, U.K.

Amanda Roestorf, Department of Psychology, City, University of London, London, U.K.

Kate Harbour, City, University of London, London, U.K.

Paul Hardiman, Institute of Women's Health, University College London, London, U.K.

John A. Barry, PhD, Institute of Women's Health, University College London, London, U.K.

Abstract

From an early age, most children typically choose to play with toys typed to their own gender. In order to identify variables that predict toy preference, we conducted a meta-analysis of observational studies of the free selection of toys by boys and girls aged between one and eight years. From an initial pool of 1788 papers, 16 studies (787 boys and 813 girls) met our inclusion criteria. We found that boys played with male-typed toys more than girls did (Cohen's d = 1.03; $p < .0001$) and girls played with female-typed toys more than boys did (Cohen's d = -0.91; $p < .0001$). Meta-regression showed no significant effect of presence of an adult, study context, geographical location of the study, publication date, child's age, or the inclusion of 'gender-neutral' toys. However, further analysis of data for boys and girls separately revealed that older boys played more with 'male-typed' toys relative to 'female-typed' toys than younger boys ($\beta = .68$, $p < .0001$). Additionally, an effect of the length of time since study publication was found: girls played more with 'female-typed' toys in earlier than in later studies ($\beta = .70$, $p < .0001$) whereas boys played more with 'male-typed' toys ($\beta = 0.46$, $p < .05$) in earlier than in more recent studies. Boys also played with 'male-typed' toys less when observed in the home than in a laboratory ($\beta = -.46$, $p < .05$). Findings are discussed in terms of possible contributions of environmental influences and age-related changes in boys' and girls' toy preferences.

Keywords: toy preference; gender differences; play.

Introduction

Decades of study, beginning with Benjamin's (1932) paper, find evidence of sex differences in children's toy preference; data has been collected in a range of social contexts, using visual preference, experimental and observational methodologies. Sex differences in children's object preferences may originate in biological predispositions which are subsequently influenced by social processes. Alternatively, they may be solely or primarily attributed to social factors. In either case, the impact of socialisation is likely to be modified as cognition develops and boys and girls become aware of their gender group affiliation and associated norms.

Despite a wealth of pertinent data, there is still uncertainty about the degree to which given variables contribute to outcomes and vary with child age. This topic is of theoretical and everyday interest as parents, educators, marketers and the media question how children's 'gender-typed' interests are best supported or diverted.

Hyde (2005) recommends meta-analyses to examine the influence of context on behavioural sex-differences and in identifying age trends in their magnitude, yet her review of meta-analyses pertaining to sex differences does not cover sex-differences in children's play. We therefore present a meta-analysis of systematic observations of children's free selection of toys spanning several decades. Our aim is to estimate the effect of the social and cultural context of testing, child age, and methodologies where boys and girls are offered 'gender-neutral' as well as 'gender-typed' toys. Whereas many meta-analyses focus on small sex-differences, we aim to investigate whether sex differences in toy choice are as large as they are perceived to be and as reported to be in some individual studies.

1.1. Biological predispositions

Aptitudes which may impact on boys' toy choice include their typical advantage over girls in gross motor skills (Touwen, 1976) and propulsive movement (Benenson, Tennyson &

Wrangham, 2011), their higher activity levels (Campbell & Eaton, 1999) and lower impulse control (Else-Quest, Hyde, Goldsmith & Van Hulle, 2006). Toys affording construction and mechanical movement may appeal more to boys because of their early advantage in mental rotation of figures (Moore & Johnson, 2008; Quinn & Liben, 2008, 2014) and event-mapping (Schweinle & Wilcox, 2004; Wilcox, 2003). In contrast, girl's greater attraction to social stimuli may account for their preference for toys which afford nurturance; girls engage more than boys in mutual gaze (Lavelli & Fogel, 2002; Leeb & Rejskind, 2004) and have a small advantage in processing facial expressions (McClure, 2000). Girls' typical advantage in fine motor finger control (Nagy, Kompagne, Orvos & Pal, 2007; Touwen, 1976) may also drive toy selection.

The interests and aptitudes which pertain to differential attraction to objects may relate to hormonal characteristics. A large body of evidence from research with mammals shows strong effects of exposure to pre- and peri-natal sex hormones on sex-related behaviours, including juvenile play (Meyer-Bahlburg, Dolzal, Baker, Carlson, Obeid & New, (2004). Both prenatally and during the first six months of life, boys are typically exposed to higher levels of androgens than girls, resulting in masculinisation of the neural systems and of behaviour (Collaer & Hines, 1995; Hines, 2004).There is some evidence that levels of androgen exposure affects object preference in very young infants; for example, in a study of the visual preferences of 3-8 month-old infants, boys made more fixations to a truck than a doll ($d = .78$) whereas girls had more interest in the doll than the truck ($d > 1.0$). Stronger evidence of hormonal influence on toy preference comes from clinical populations (e.g. Cohen-Bendahan, van de Beek & Berenbaum, 2005; Hines, Brook, & Conway, 2004). A review of studies of the effect of testosterone concentrations on human development indicates the contribution of this hormone to the masculinisation of behaviour (Alexander, 2014).Girls with congenital adrenal hyperplasia who are exposed to higher levels of androgens than other

girls show correspondingly greater interest in ‘male-typed’ over ‘female-typed’ toys (Hines, 2004; Meyer-Bahlberg et al.,2004) and girls more affected by CAH showed greater interest in ‘male-typed’ toys than those less affected by the condition (van de Beek et al., 2009).

Further indications of relevant biological predispositions arise from the findings of sex-differences in object preference among non-human primates: female vervet monkeys make more contact with a doll and a cooking pot than males do, whereas males spend longer than females manipulating a car and a ball (Alexander & Hines, 2002) and male rhesus monkeys prefer wheeled over non-wheeled plush toys (Hassett, Siebert & Wallen, 2008).

In general, biological explanations predict continuity in boys’ and girls’ object preferences over late infancy and childhood and across contexts of testing.

1.2 The Effects of Socialisation and Study Setting

Various environmental processes involving the attitudes and behaviour of parents and other children may modify children’s toy preferences and is likely to vary by age and cultural norms.

Although parents’ reported opinions on gender stereotypes do not correlate strongly with their children’s behaviour (Fagot, 1974; Smith & Daglish, 1977), parent behaviour may; a meta-analysis indicates that parents encourage gender-typed activities, though decreasingly so as children grow older (Lytton & Romney, 1991). A social context where a parent is present may therefore incline a younger child to exhibit stereotypical preferences. However, some recent studies show that infants engage more in ‘gender-typed’ play when alone than with a parent (Zosuls et al., 2009), as do typically-developing girls, but not boys (Pasterski et al., 2005). The effect of socialisation may depend on parent and child sex, and vary as parents redefine their roles in interaction with the knowledge and perceived needs of younger and older children.

When the play-partner is another child, ‘gender-typed’ play is typically increased; pre-school children spend less time playing with ‘other-gender-typed’ toys with a peer than in solitary play (Serbin, Connor, Burchardt & Citron, 1979) and children’s behaviour is more highly stereotyped in group than dyadic situations (Fabes, Martin & Hanish, 2003; Banerjee & Lintern, 2000). Children engage more with same-sex than other-sex peers (Hines & Kaufman, 1994; Martin & Fabes, 2001) and, in such contexts, their toy preferences may reflect their motivation to conform to sex-role stereotypes as much as individual preference. Typically, parents’ differential socialisation of boys and girls as well as children’s own attitudes, are likely to reflect their cultural context. Lytton & Romney's (1991)’s meta-analysis of parental socialisation showed differences between North American studies and those from other Western countries. Few individual studies of children’s toy preference include regional comparisons. Although Turner & Gervai (1995) found substantial replication of parental effects on toy preference between the UK and Hungary, the finding that both girls and boys in Sweden preferred ‘female-typed’ toys less as age increased (Servin et al., 1999) contrasts with similar but earlier studies, conducted in the US, which found increases in girls’ preference for ‘female-typed’ toys with age (Blakemore, LaRue, & Olejnik, 1979; O’Brien & Huston, 1985). Servin et al. (1999, p.40) speculated that the difference they observed may be explained by Sweden’s “equal-roles family model”.

The study of children’s toy preferences has spanned several decades, allowing for a comparison of boys’ and girls’ behaviour in different historical time periods. Historical changes in gender roles are likely to impact on socialisation; Karraker, Vogel and Lake's (1995) replication of a study by Rubin, Provenzano and Luria (1974) found a reduction in US parents’ stereotypical perception of newborns; similarly, Signorella, Bigler and Liben (1993) found smaller age and sex effects in children’s own nonstereotyped responses in earlier relative to later studies (conducted between 1963 and 1991). However, Zosuls et al. (2009)

suggest that many Western middle-class parents currently promote gender-neutral play. We suggest, therefore, the effect of time-period, can be used as a proxy measure for the effect of culture on children's gender-typed object preferences.

The immediate settings in which children's toy preferences are tested vary in the formality and social context of testing, factors which might affect behaviour. For example, laboratory and home contexts typically involve interaction with, or the presence of, a parent, whereas in school or nursery settings the focal child may be in the presence of their peers.

Comparing findings from studies conducted in different social settings (e.g. presence/absence of parent or peers) as well as geographical and historical contexts, while controlling for other factors, may inform us on specific environmental effects on children's sex-typed choices. Extrapolating from the studies described above, we might therefore expect to see the magnitude of sex differences in toy preference differ by: (a) social context of testing; (b) geographical region; (c) date since publication; and (d) child age. One specific aspect of context that has not received research attention is whether the immediate environmental setting of the study (home, nursery, laboratory, etc.) influences boys' and girls' toy choice.

1.3 The Development of 'Gender-Typed' Behaviour

Throughout this review, we refer to the potential impact of child age on the degree to which 'gender-typed' preferences for toys are demonstrated with the assumption that a child's conception of gender and associated stereotypes will expand with age.

Conceptualisation of gender is comprised of perceptual, cognitive and social components, acquired incrementally. By six months of age, infants distinguish between male and female faces (Quinn, Yahr, Kuhn, Slater & Pascalis, 2002) and voices (Miller, 1983). Explicit labelling of others by gender occurs between the ages of 18-24 months (for review

see Martin & Ruble, 2010) and it is likely that infants begin to understand that they too have a gendered identity during this period. As the concept of gender develops, selective modelling (Slaby & Frey, 1975) of parent and peer behaviour is expected to increase and the media become more influential (Diekman & Murnen, 2004; Leaper, Breed, Hoffman, & Perlman, 2002). Higher levels of gender category knowledge predict increased awareness of stereotypical preferences.

In contrast to the development of stereotyped cognition, Martin & Ruble (2004) argue that 'gender-typed' behaviour shows more mixed developmental patterns. A meta-analysis of studies of developmental change in gender schemata shows that, although knowledge of gender stereotypes increases with age, behaviour in accordance with the stereotypes does not necessarily follow (Signorella et al., 1993) and this aspect requires further exploration.

Studies that include infants and/or pre-school children in more than one age-group generally report increases in 'gender-typed' toy preference with age (e.g. O'Brien & Huston & Risley, 1983; Todd et al., 2016; Zosuls et al., 2009; Zosuls et al., 2014) but the distinct developmental patterns of boys and girls vary between studies (see Servin et al., 1999). Drivers of developmental change are multifaceted and a more systematic comparison of boys' and girls' stereotypical preferences at different ages and across different social settings and cultural contexts may add to our understanding.

1.4 Methodological Aspects of Previous Research and Rationale for Choice of Studies in the Present Meta-Analysis

Toy preferences have been measured in a variety of ways. Parental and child self report, sometimes retrospective, is valuable when working with large samples but may be prone to bias in recall (Yarrow, Campbell & Burton, 1970) or demand characteristics (see Wilansky-Traynor & Lobel, 2008, for a discussion). Visual preference techniques can test infants whose motor development is insufficient to demonstrate preference in other ways. However, it may

be the child's ability to act on an object, as well as its visual properties, that is appealing. This meta-analysis covers only observational studies testing children's free-preferences in various controlled settings because we believe that these best indicate a child's free preference in a play situation. Meta-regression is used to control for any influence of child age and study context on toy preference. Where extraneous variables could not be controlled in any meaningful way, studies were excluded; for example, we exclude studies measuring play with peers because competition for toys may limit choice: studies involving play with the child's own toys, on the basis that available toys differ between participants and are already familiar to them. Included studies use a wide range of 'gender-typed' toy stimuli and many also include toys defined as 'gender-neutral'. This variation in methodology is also controlled as it may impact on the effect size of found sex-differences.

In summary, studies reporting observations of children in free play were considered the best indicator of toy preference because, (a) they typically offer a range of 'gender-typed' toys and, (b) they allow manipulation and movement of the stimuli and have the potential for symbolic play.

1.5 The Present Review

This meta-analysis was designed to test the following hypotheses:

1) Gender-typed preferences will be observed in children's free play with toys.
2) The setting in which the child plays, for example the presence of a parent or location in a laboratory rather than home, will affect the magnitude of 'gender-typed' behaviour.
3) Effect sizes for toy preference will be smaller in more egalitarian compared with less egalitarian societies.
4) Effect sizes for toy preference will be larger in historically earlier studies rather than later studies.

5) Sex-differences in toy preference will increase with age.
6) Effect sizes for toy preference will be smaller when the children are presented with 'gender-neutral' toys at the same time as they are presented with 'gender-typed' toys.

2 Method

2.1 Selection of Studies

The selection of studies followed the guidelines issued by Stroup et al. (2000). Qualified librarians advised on the search strategy and assisted where papers were difficult to access. The study protocol is available from the authors.

2.2 Inclusion and Exclusion Criteria

Five inclusion criteria were applied: first, studies included typically developing children; second, duration of play a toy was reported in mean and SD or SE in seconds, or this information could be extracted indirectly from the paper e.g. from a histogram with error bars; third, the study provided observational data, or provided baseline data prior to experimental manipulation; fourth, the study reported other relevant data, e.g. the number of children in each group; fifth, stimuli toys were equally available to all participants and not shared with other children at the time of testing.

Papers not published in English were translated and reviewed.

Published studies were assessed from the earliest possible dates for each database up to 24th Sept 2016. The search terms used, in topic, were: "("toy choice*" OR "toy preference*" OR "toy play") OR ("toy" AND "play behav*") OR ("gender typ* play") OR ("sex-typ* play") OR ("gender stereotyp* toy") OR ("gender stereotyp* play") OR ("sex-typ* toy") OR ("sex stereotyp* toy") OR ("gender typ* activit*") OR ("sex typ* activit*")". As Figure 1 shows, this yielded 520 hits from Web of Knowledge / Web of Science (from 1900), 205 from Embase Classic+Embase (from 1947), 11 from Maternity and Infant Care (from

1971), 200 from Ovid MEDLINE(R) (from 1946), and 852 from PsycINFO (from 1806). A hand search of the resulting papers produced no further relevant papers. Authors were contacted where additional information was needed. For example, the Lamminmäki & Hines (2012) study did not include the mean and SD seconds of play, so the authors of that paper were contacted and provided this information.

2.3 Data Extraction

Data were collected and coded based on relevance to child development research and toy choice paradigm methodology. Five variables were identified as having a possible impact on the child's toy choice, and were used as covariates in the meta-regression. The following data were extracted from each study:

(i) The presence of another person during the play task. This was categorized such that higher scores indicate more potential influence of the presence of another person on the child: 1 = Child completely alone; 2 = Child engaged in sole play with minimal parental or researcher interaction; 3 = Combination of 50% sole play with inactive parent and 50% parent present and actively involved in play with child.

(ii) The study location. Whether the study was set in the home or at another location was coded as: 1=Home; 2=Lab; 3=Nursery.

(iii) The gender equality values of the region where a study was conducted. This was estimated using the gender inequality index (GII; Ferrant, 2010) in order to assess any effect of cultural gender stereotyping on children's toy preferences. GII scores represent gender inequality in: reproductive health, parliamentary seats, higher educational achievement, and participation in the workforce. GII scores exist for 137 countries, from 1995 in most cases ("Table 4: Gender Inequality Index |

Human Development Reports," n.d.). Lower scores indicate more gender equality. To give an idea of the range of scores, as of the latest date (2013), Sweden is third in the world table with a score of 0.049, and Yemen is at the bottom with 0.733. The GII is produced every few years and does not tend to vary much over time; for example, though the GII is falling in the UK over the five years measured (2000 = 0.228; 2005 = 0.216; 2010 = 0.207; 2012 = 0.205; 2013 = 0.193), this is a relatively slow decline. In the present meta-analysis, where no GII score in the year a study was conducted, the GII score in the year closest to the study was used.

(iv) Years since publication of a study. This measure tests the hypothesis that the prevalence of gender stereotyping has changed over time with the expectation that older studies would find more stereotypical toy preference.

(v) The age of the children. This information was extracted because of the relevance of children's developmental stage to preference for toys. For example, Zosuls et al. (2009) found that sex-difference in toy preference was higher in older than in younger infants.

(vi) The inclusion or not of toys defined as 'gender-neutral'. This variable tests for an effect of including 'gender-neutral' toys on sex differences in toy preference and was coded as: 1 = included 'gender-neutral' toys; 0 = no 'gender-neutral' toys included.

The outcome variable was operationalised in three ways, reflected in sections 3.2, 3.3, 3.5 and 3.6. Firstly, in the meta-analyses and meta-regressions the outcome was expressed as the standard mean difference between the duration with which boys and girls played with male-typed or female-typed toys. The effect size of this difference was estimated in units of

Cohen's *d.* A Cohen's *d* of 0.2 is considered a small difference between groups, 0.5 a moderate difference, and 0.8 or more a large difference. Secondly, regressions were performed separately for boys and girls, with outcomes expressed as the percentage of time played with male-typed toys relative to female-typed toys. Thirdly, the actual duration of play - rather than proportion or percentage of play – for each toy was assessed for boys and girls separately. Thus the raw number of seconds of play with male- and female-typed toys, rather than the percentage of play with male toys relative to female toys, was assessed.

Tables 1a and 1b show information regarding: (a) characteristics of the participants (mean age, sex); (b) the setting (presence of others; at child's home or nursery, etc.); (c) the country in which the study was conducted; (d) whether or not 'gender-neutral' as well as 'male-typed' and 'female-typed' toys were included ; (e) mean seconds played with toys); and (f) years since study publication . Table 2 shows the additional data that were extracted in order to assess the methodological quality of each study, for example, whether a study had controlled for relevant variables.

[Table 1a and Table 1b here]

Each article was assessed by authors SD, AR or KH, BT, RF and JB. Articles that fitted the main criteria (observing sex differences in children's choice of gender-typical toys) were accessed. Methodological quality was assessed by SD, AR, KH, RF and checked by JB based on the criteria of the Newcastle-Ottowa Quality Assessment Scale (NOS) for case–control studies (Wells et al. 2011). The Cochrane Non-Randomized Studies Methods Working Group consider the NOS an acceptable tool for assessment of non-randomized studies (Reeves & Higgins, 2011). The criteria considered were: (a) clear justification for gendered nature of toy e.g. based on research; (b) recruitment of consecutive participants; (c) whether boys and girls were comparable in terms of social background; (d) whether parents' views on gender were measured; (e) whether the toys were comparable (in size, shape etc.)

and the boys and girls ages were comparable; (f) play behaviour was clearly defined; (g) the measurement of the outcome (time spent playing with toy) was clearly defined; (h) whether non-uptake or dropout rates were reported. Inter-rater agreement of NOS scoring was assessed using Cohen's Kappa. A Kappa of 0.41 – 0.60 generally considered as demonstrating 'moderate agreement' and 0.61 - 0.80 as 'substantial agreement' (Landis and Koch 1977). The NOS scores of SD and AR and RF and KH were in substantial agreement, as shown by the Cohen's Kappa of 0.77 (p = .000013); the mean score for the 16 studies was 6.63 (SD = 0.81). Discrepancies were examined and discussed before a final score was agreed upon and assigned.

The time difference in boys' and girls' play with 'gender-typed' toys in each age-group in each study (shown in Figures 2a and 2b) were calculated as a common unit, Cohen's *d*. The combined effect size for all studies was calculated as the estimated treatment difference, *Z,* with fixed or random effects model, as appropriate. Results were considered statistically significant where the probability value was below the 0.05 threshold. Statistical analyses were performed using Review Manager Version 5.1 (The Cochrane Collaboration, 2011), Stata Version 13.0 (StataCorp, Texas) and SPSS 22.

Heterogeneity was assessed using I^2 and χ^2 statistics. I^2 values of 30% or above were considered likely to represent moderate heterogeneity, and χ^2 p values <.10 were considered to represent significant heterogeneity, thus studies showing tolerably heterogeneity (I^2 values < 30% and χ^2 p values >.10) were analyzed using fixed effects models, and studies beyond these limits (I^2 values > 30% and χ^2 p values <.10) were analyzed using random effects models.

2.4 Eligible articles

Figure 1 summarises the search strategy used to identify appropriate studies. After duplicates were removed, the titles and abstracts of 961 records, from the 1788 initially retrieved, were

assessed. A total of 182 full texts were further assessed, including reference sections, and 166 were excluded for various reasons (for example, toy play data was not separated by child sex). One study was excluded (Berenbaum & Hines, 1992) because of overlap of participants with another (Berenbaum & Snyder, 1995); the more recent study was selected for inclusion because of the greater number of participants and the higher rating on the NOS scale.

Finally, 16 papers met all of the inclusion and exclusion criteria.

Some studies presented data in proportions of total time playing (e.g. Zosuls et al., 2009) or intervals of time (e.g. Todd et al., 2016) and these studies were included if the overall time in seconds was stated in the paper, from which an approximation of the mean time could then be calculated.

Some studies were excluded, for example where there was an experimental manipulation with no baseline data given (e.g. Wolf, 1973), or where a child was assessed playing together with other children (e.g. Serbin, Connor & Citron,1981), or when children were tested at home with their own toys rather than with toys supplied by the experimenter (e.g. Fagot & Leinbach, 1989).

In one study (Pasterski et al., 2005), 84 (71%) of the participants were assessed in the UK, and 34 (29%) were assessed in the US. In this case, the GII score for each country (0.216 for the UK and 0.288 for the US) was weighted according to the relative number of participants in each country, giving a combined GII score for the study of 0.228.

[Figure 1 here]

The 16 studies that met the inclusion criteria for the meta-analysis (787 boys and 813 girls, in 27 groups) are listed in Tables 1a and 1b (raw data publicly available here DOI: 10.6084/m9.figshare.5047660, DOI: 10.6084/m9.figshare.5047663). The children had a mean (SD) age of 43.25 (27.62) months old (minimum age group mean 12 months, maximum 93 months). Five of the 16 studies presented their findings in subgroups based on age. The mean

(SD) size of each of these subgroups was 33.33 (26.69) for the boys and 38.75 (27.10) for the girls. Fifteen studies were conducted in Western countries (US, Canada, Europe or Israel) and one in Hong Kong with children of Chinese ethnicity, and the findings were published between 1980 and 2016. Three studies were conducted in the child's home, ten in a laboratory setting, and three in a nursery. In four studies the child played alone, in seven an adult was present but not interacting with the child and in five studies there was moderate or full interaction between parent and child. Thirteen studies included 'gender-neutral' toys as well as 'gender-typed' toys in the stimuli presented to children. Fourteen studies were cross-sectional, and two had both cross-sectional and longitudinal elements.

2.5 Methodological quality

Regarding the quality of the studies, in general the NOS scores were all of at least moderately good quality. Table 2 shows that the scores ranged from five to eight out of a maximum possible of nine.

[Table 2 here]

3 Results

3.1 Data Analysis

Heterogeneity was assessed using I^2 and χ^2 statistics. A meta-analysis showing tolerable heterogeneity (I^2 values < 30% and Chi Square *p* values >.10) will be analyzed using fixed effects models, and studies beyond these limits will be analyzed using random effects models.

3.2 Meta-analyses

The meta-analysis found that boys played with male-typed toys more than girls did (Figure 2a). The effect size of this difference was large (Cohen's d =1.03 ; $p < .0001$) and girls played

with female-typed toys more than boys did (Figure 2b). The effect size of this difference was large (Cohen's d = -0.91 ; p<.0001).

[Figures 2a & 2b here]

In interpreting Figures 2a and 2b, a minus sign in front of the *d* value indicates that girls engaged in playing with the toy more than boys did, and a plus sign in front of the *d* value indicates that boys engaged in playing with the toy more than girls did. Thus, in Figure 2a, showing the results of playing with male-typed toys, where d = [+]1.03, this means that boys played with male-typed toys more than girls did.

There was substantial heterogeneity (I^2 > 70%) so the random effects model was used. Figures 2a and 2b and Table 3 show that there were large sex differences in toy preference.

[Table 3 here]

3.3 Meta-Regression

A random-effects meta-regression was performed on the data using Stata's *metareg* command. A random effects model was chosen not only because of the heterogeneity seen in Table 4 but also because the random effects model generally reduces the chance of Type 1 error in meta-regression analysis (Harbord & Higgins, 2008). The two criterion variables were sex difference in play with 'male-typed' toys and 'female-typed' toys, and the six predictors were: (a) the age of the child in months; (b) the presence of parent (1 = absent; 2 = minimal interaction; 3 = moderate or full interaction); (c) setting of study (1= home; 2 = lab; 3 = nursery); (d) how gender-egalitarian the country was at the time the study took place; (e) years since publication; and (f) whether or not 'gender–neutral' as well as 'gender-typed' toys were included in the study.

[Table 4 here]

Table 4 shows that, for male-typed toys, the covariates predicted only 11.51% of the variation in the sex difference in time played with 'male-typed' toys, and the model as a whole did not perform significantly better than chance ($F(6,19) = 1.30$, $p = .301$). In this model none of the predictors were significant. The strongest predictor was Age, but was non-significant ($t = 1.77$, $p = .093$), indicating that the sex differences in play with 'male-typed' toys become non-significantly larger as children grow older.

For female-typed toys, although the covariates predicted 38.17% of the variation in the sex difference in time played with 'male-typed' toys, the model as a whole did not perform significantly better than chance ($F(6,19) = 2.55$, $p = .056$). In this model none of the predictors were significant. The strongest predictor was Age, but was non-significant ($t = -1.81$, $p = .086$), indicating that the sex difference in play with 'female-typed' toys become non-significantly smaller as children grow older.

3.4 Publication Bias

To assess publication bias, funnel plots were created (Figures 3a & 3b). Overall, most of the studies are within the 95% confidence intervals (CIs), despite the fact that study heterogeneity will tend to place the studies beyond these CIs. However, asymmetry is in evidence. The funnel plot of the findings for the sex differences in play with 'male-typed' toys (Figure 3a) is asymmetrical (seen in the tail in the lower right of the plot). The funnel plots for studies with 'female-typed' toys (Figure 3b) is similar to Figure 3a, except that the trend in the effect size was in the opposite direction (i.e. the tail is in the lower left) as would be expected from the Forest plots in Figures 2a and 2b. The three main outliers in Figure 3a (from right to left) are Pasterski et al., (2005; with father present), Todd et al. (2016; oldest group), and Pasterski et al., (2005; child alone), and the four main outliers in Figure 3b are Todd et al. (2016; oldest group; youngest group), Pasterski et al., (2005; child alone), Arthur (2014) and Doering, Zucker, Bradley, & MacIntyre (1989). Often asymmetry is caused by

publication bias due to 'small study effects' i.e. a bias towards small studies finding large effects and being published. Although many of the groups in this meta-analysis are, technically, small (N < 50) the outlying groups are not small relative to non-outlying groups. However, what the outlying groups all have in common is that they represent either relatively older age groups or relatively younger groups than the non-outlying groups. The NOS scores are not implicated because Table 2 shows that the outlying groups all had scores around the mid-range, either 6 or 7 from a NOS score range of 5 to 8 (Doering et al., 1989 = 6; Pasterski et al., 2005 = 7; Todd et al., 2016 = 7).

[Figures 3a & 3b here]

3.5 Analysis of percentage of time played with male-typed toys by boys and girls separately

The meta-regressions did not find any significant predictors of sex difference in toy play. However, in bivariate correlations between toy play and predictor variables some of the predictors showed different patterns of correlations when data from boys and girls were analysed separately. Such differences violate the assumptions for linear regression (homogeneity of regression slopes) and combining data from boys and girls, as in the above meta-regressions, is prone to Type 2 error and may result in null findings. For this reason, in the following section the regressions are performed separately for boys and girls. The outcome in these regressions cannot be 'sex differences in toy play' (because toy play cannot be compared by sex in an analysis where boys and girls are assessed separately), so the outcome variable in the following regressions is 'the percentage of time played with 'male-typed' toys relative to 'female-typed toys'. Thus, although in the previous section the toy preference was analysed separately for each toy type (combining data from boys and girls), in

the following section toy preference is analysed separately for boys and girls (combining male- and female-typed toys).

Table 5 shows the bivariate correlations between predictors and the percentage of time played with male-typed toys for boys and girls.

[Table 5 here]

Although Table 5 shows positive correlations for both boys and girls, further analysis reveals other differences. Figures 4a and 4b show, respectively, scatterplots of the relation between age and percentage of time played with male-typed toys for boys (Figure 4a), and age and percentage of time played with male-typed toys for girls (Figure 4b). Figure 4a shows a positive correlation, indicating that boys play relatively more with male-typed toys as they grow older (r_s = .43, n = 27, p < .0001, 2-tailed). In contrast, Figure 4b shows a negative quadratic correlation, indicating that girls' play with male-typed toys increases to a peak at about 4 years old, and then reduces thereafter. This quadratic trend was not significant using linear correlations *(*r_s = .03, n = 27, p = .357, 2-tailed), but was closer to significance using the SPSS curve fit analysis for quadratic functions (R^2 = .11, [r = .34], n = 27, p = .223, 2-tailed).

[Figures 4a & 4b here]

Figures 5a and 5b show, respectively, scatterplots of the relation between including gender-neutral toys in the stimuli and the percentage of time played with 'male-typed' toys for boys (Figure 5a) and girls (Figure 5b). Figure 5a shows that boys play non-significantly less with 'male-typed' toys when 'gender-neutral' toys are included (r_s = -.18, n = 27, p = .422, 2-tailed), and Figure 5b shows that girls play non-significantly more with 'male-typed' toys when 'gender-neutral' toys are included (r_s = -24, n = 27, p = .272, 2-tailed).

[Figures 5a & 5b here]

Figures 6a and 6b show, respectively, scatterplots of the relation between testing children at home and testing children in a laboratory for boys (Figure 6a) and girls (Figure 6b). Table 5 shows that boys play significantly less with male-typed toys at home (r_s = -.21, n = 27, p = .017, 2-tailed), while girls play with male-typed toys non-significantly more in a home setting compared to a laboratory (r_s =.05, n = 27, p = .253, 2-tailed).

[Figures 6a & 6b here]

Because of these differences, these variables were analysed separately for boys and girls in multiple linear regression, using SPSS 22.

Table 6 shows the results of the multiple linear regressions of the predictors, with percentage of time played with male-typed toys relative to female-typed toys as the dependent variable. Age was a significant predictor for boys, indicating that older boys played more with male-typed toys relative to female-typed toys than younger boys did. The length of time since publication of the study was a significant predictor for girls, showing that girls play less with male-typed toys and female-typed toys in more recent studies. The regression model for boys was a strong predictor of play (Ad R^2 = .57, F =5.69 p < .001). For girls only, the regression model was a moderate predictor of play (Ad R^2 = .43, F = 3.68, p = .012).

[Table 6 here]

3.6: Analysis of raw seconds of play with each toy type by boys and girls separately

In Section 3.2, toy preference was analysed separately for each toy type (combining boys' and girls' data), and in the Section 3.5 toy preference was analyzed separately for boys and girls (combining male- and female-typed toys). In this section, the actual duration of play with each toy type was assessed for boys and girls separately. Thus the number of seconds of

play with male- and female-typed toys, rather than the percentage of play with 'male-typed' toys relative to 'female-typed' toys, was assessed. Table 7 shows the Spearman bivariate correlations for both boys' and girls' play behaviour for male- and female-typed toys. It was found that for boys, age correlates significantly with the raw seconds of play with 'male-typed' toys (r = .54, n = 27, p < .05) while play with female-typed toys negatively correlates (r = -.51, n = 27, p < .05). Additionally, it was found that for girls, the older the study the longer they played with both male-typed (r = .65, n = 27, p < .001) and female-typed toys (r = .50, n = 27, p < .005). For boys, the older the study the longer they play with 'male-typed' toys (r = .71, n = 27, p < .0005) but there was no effect of time on their play with 'female-typed' toys (r = .05, n = 27, p = .816).

4. Discussion

This meta-analysis found sex differences in the preferences for male-typed and female-typed toys among boys and girls aged between one and eight years in the predicted directions; boys played with male-typed toys more than girls did and girls played more with female-typed toys than boys did. The effect sizes of these differences were large (Cohen's d = 1.03 and d = -0.91, respectively). This finding, from 16 observational studies measuring boys' and girls' independent choice of 'gender-typed' toys in free play situations, is consistent with the wider literature on sex-differences in play; both observational and questionnaire data indicate that typically developing boys and girls prefer different toys (Hines, 2004). Meta-regression detected no significant effect of the presence/absence of an adult, the study setting, the gender equality status of the country, year of publication, and presence of 'gender-neutral' toys indicating a consistent effect of child sex on toy preference across a range of environmental circumstances. However, further analysis revealed significant effects of age and indicated

that boys' and girls' preferences develop differently with age and may be subject to differential change over contexts and historical time.

4.1 The effect of age

The included studies of children's independent behaviour varied by social, geographical and historical context and by stimuli representative of toy type. Although the meta-regressions reported in Section 3.3 found no effect of age, this may have been due to nonlinearity in the data for girls (Figure 4b). In Section 3.5, the analysis of percentage of time played with 'male-typed' toys relative to 'female-typed' toys by boys and girls separately indicates that boys and girls show different developmental patterns with respect to toy preference. For boys there is a clear positive correlation between time playing with 'male-typed' toys and age, but the findings for girls do not show a mirror image of that profile. For girls there is a clear negative quadratic (inverted U shaped) correlation with age. Figure 4a suggests that younger boys (up to ~20 months) play relatively more with 'female-typed' toys, whereas Figure 4b suggests that girls' interest in 'male-typed' toys peaks ~4yrs and then tails off somewhat. Although Table 5 shows that this finding for girls was nonsignificant, the significance of the trend cannot be detected by linear regression because the trend is nonlinear. Our findings, therefore, are consistent with an effect of cognitive development and differential socialisation on the strength of boys' and girls' preferences for gender-typed toys.

These findings may indicate the differential effects of social experiences on boys' and girls' cognitive development. Infants' application of rigid stereotypes may be consequential on the acquisition of a gendered identity and their accumulation of knowledge about typical gendered behaviour in others (Zosuls et al., 2009). This view is consistent with our finding of an increase in the magnitude of boys' and girls' free preference for toys typed to their own gender across this age period. After infancy, social influences (from parents, peers, educators and the media) are more accessible to children and are likely to differ, by gender, in both

focus and intensity, according to prevailing norms of gendered attitudes and behaviour. Children's stereotypical beliefs increase from the ages of 3 to 5 years (Halim, Ruble, Tamis-LeMonda & Shrout, 2013), peak between the ages of 5-7 years and become more flexible thereafter (Trautner et al., 2005), perhaps because children gain an appreciation of individual differences in preferences. The trend for age-related changes identified in this meta-analysis fits the view that the strength of stereotypes will "wax and wane across development" (Martin & Ruble, 2004, p. 68) but indicates that transition away from greater rigidity may begin earlier than suggested.

A less flexible application of stereotypes in younger children is indicated by the finding that three- and five-year-old children ridicule same-sex peers who play with 'cross-sex' toys (Langlois & Downs, 1980) and this tendency possibly persists longer for boys. The oldest children in the included studies were aged about 8 years, which may have been too early to detect a reduction in boys' stereotyped behaviour. Our findings indicate that reduction in gender-typed toy preference may apply only, or more, to girls. This change is consistent with the finding from a questionnaire study that, as age increased, both girls and boys increasingly requested 'male-typed' toys as gifts (Etaugh & Liss, 1992).

Of the four included studies reporting age comparisons, one found that both boys and girls played longer with female-typed toys as age increased (Schau et al., 1980) and another that boys' play time with a truck and doll were relatively consistent across the ages of 24 and 36 months whereas girls' play time with both of these toys increased slightly with age (Zosuls, et al., 2014). The others show that female-typed toys became less interesting to girls as well as boys with age (Servin et al., 1999; Todd et al., 2016). One possible explanation for the inconsistency might be differences in the selected stimuli; for example, Schau et al., (1980) included both a doll house and a mixer as 'female-typed' toys, both of which were popular with boys.

4.2 Presence/absence of an adult and setting of study

Contrasting studies or conditions where children played alone or in interaction with a parent have the potential to compare the immediate effect of social influence and individual play which might better reflect the child's intrinsic preferences. In the meta-regression, the immediate social context of testing (child playing alone or in the presence of or in interaction with a parent) did not predict the size of sex differences in play with male-typed or female-typed toys. Of the included studies which compare infants across conditions in which the same participant plays (a) alone and (b) in interaction with a parent, findings are inconsistent. One study found stable sex differences across 'play alone' and 'play with parent' conditions (Alexander & Saenz, 2012), whereas others found that infants played with gender-typed toys less when with their mother than when alone (Zosuls et al., 2009) and more with male-typed toys when with their father than alone (Servin et al., 1999). In studies involving older children, no differences were found between 'play alone' or 'play with parent' conditions for 3-5 year-olds (Servin et al., 1999) but typically-developing 3-10 year-old girls were found to play more with 'female-typed' toys when alone than with either parent (Pasterski et al., 2005). These findings have implications for the design of studies of sex differences in toy preference and may indicate nuances in the immediate social influences on boys and girls of different ages. Typically, fathers are under-represented in studies which include play with a parent and this may influence the results when sex of parent is not controlled for.

In the majority of studies, an adult observer (often the researcher) is present. Conditions described as 'play alone' do not necessarily preclude the presence of an inactive parent or adult researcher (especially with younger participants). The effect of such variables could be explored further. Schau, Kahn, Diepold, & Cherry (1980) found no effects of adult observer presence/absence or observer sex on typically-developing pre-school children's gender-typed toy selection. However, an effect of adult presence may vary by age/developmental stage;

pre-school-aged boys and girls rated as 'gender-aschematic' displayed more gender-typical play behaviour in the presence of an adult observer than when they were alone, whereas no such difference was found for children rated as 'gender-schematic' (Wilansky-Traynor & Lobel, 2008). Taken together, findings relating to social context indicate a potential interplay of factors, including the sex of the play partner/observer and the child's sex and developmental stage and the level and focus of interaction between adult and child. In modern Western societies, at least, there may be a move towards parental encouragement of gender-neutral play (see Zosuls et al., 2009). Studies which strictly control for immediate social influences on children at different ages may have the potential to inform our understanding of how expectations of social behaviour are modified by specific contexts, including self-presentation in the presence of other children.

Study locations (home, laboratory, nursery) may also impact on children's sex-typed behaviour. Single studies of sex differences in children's free play have not contrasted study settings but the meta-analysis provides an opportunity to do so. Interestingly, as Table 5 shows, boys played with 'male-typed' toys less when at home than in a laboratory setting (r_s = -.46; $p < .05$), however, whilst the trend is in the same direction, the finding for girls is non-significant (r_s = -.23; $p > .05$),. It is difficult to make conclusions from this finding but the relative formality and a potential 'testing' element of the context may draw children, especially boys, to toys which they are aware are typed to their own gender.

4.4 The effect of country (gender egalitarianism) in which the study was conducted

The opportunity to make comparisons of children's toy preferences across different countries allows a test of wider social influence on both children and their families. Although the influence of adults and of the media is likely to vary by cultural context, the size of sex

differences in children's preferences for male-typed and female-typed toys did not appear to be smaller in studies conducted in more egalitarian countries; large effects of sex were apparent across regions where the GII varied. It could be argued that non-significant findings were due to low power. Although the number of groups was small, the number of participants was about 26 times greater than the number of groups. Thus, although technically the regressions were of low statistical power, the β values may generalise well to the general population of children, and relatively large beta values (e.g. .594 for GII for girls) should be deemed as worth consideration, even if not statistically significant.

There is probably no perfect way to measure gender egalitarianism in any country at any given time and sub-sections of the country-wide population are likely to vary on the included measures. The GII has an advantage over some other indexes in having scores for many countries going back to 1995. This is not ideal for the range of study dates in this meta-analysis, but is better than the gender gap index, another popular tool, which goes back only to 2006. The GII is only produced every few years, so the year of study publication and year GII was conducted cannot be matched exactly. However, only the three earliest studies were published more than two years prior to the dates of GII reports and these do not vary much within countries over time (e.g. the GII in the UK in each of the four years measured: 2000 = 0.228; 2005 = 0.216; 2010 = 0.207; 2012 = 0.205; 2013 = 0.19). Despite the relative consistency across regions showing some support for biological determinants of preferences, a limitation of this meta-analysis, and of this research area in general, is that most of the research has been conducted in Western countries and findings cannot be generalised elsewhere.

4.4 The effect of the year the study was published

Table 5 shows that, in bivariate correlation, the year of publication of the study was a significant predictor for boys ($r_s = .47$; $p < .05$), and more so for girls ($r_s = .70$; $p < .0001$). Table 5 shows that this effect remains significant when the effect of other variables is accounted for. Girls play significantly less with female-typed toys and male-typed toys in more recent studies. Therefore, the findings of this study indicate a possible effect of historical time on toy preference. This finding may indicate moves towards greater gender equality in Western societies (Schwartz & Rubel-Lifschitz, 2009). One interpretation is that, over the years, girls have been exposed to increasing social pressure to play with neutral-typed toys as the volume of advertising to children and expansion in the number and type of children's media outlets in the US, and doubtless in other developed countries, has increased enormously over time (Bakir, Blodgett, & Rose, 2008). A further possible explanation would be that female-typed toys have become more interesting or attractive to girls over time, yet the included studies have typically and fairly consistently used dolls, cosmetics and kitchen equipment as female-typed toys.

A possibility to be considered relates to differences in the researchers' selections of male-type' toy stimuli in different time periods. Earlier studies included in the meta-analysis are more likely to offer toys associated with aggression, for example a dart gun and army shirt (Doering et al.,1989) and fighting figures (Servin et al., 1999) yet a gun was used as a 'male-typed' toy in one recent study (Li & Wong, 2016). Of course, other toys, not necessarily male-typed can afford aggressive play, as Schau et al. (1980) note in their observation that boys played with a toy food mixer as if it were a machine gun. It is difficult to be precise about the year in which a study was conducted as there is typically some time lag between this and the date of the resulting publication. The finding that girls spend a comparatively shorter time playing with gender-typed toys in recent studies is of interest, given the indication that some parents and educators may currently promote gender-neutral

toys for girls. The number and specific characteristics of toys vary by study are not always reported, yet no systematic changes over time are apparent from the information available. However, media coverage, cultural trends or current events may affect children's toy preferences. For example, at the time when Benjamin (1932) found that girls played with a toy aeroplane for similar times to boys female aviators were in the news; Amy Johnson's flew from England to Australia in 1930 and Amelia Earhardt flew solo across the Atlantic in 1932.

4.5 The effect of inclusion of 'gender-neutral' toys

Like age, the inclusion of gender-neutral as well as gender-typed toys as stimuli appeared to be associated with opposite (though statistically nonsignificant) relationships for boys and girls; boys played relatively less with male-typed toys when gender-neutral toys were included in the stimulus materials (Figure 5a), and girls played relatively more with male-typed toys when gender-neutral toys are included (Figure 5b). However, in both cases, although the directions of correlations remained the same in multiple regression, the apparent effect remained non-significant when the other variables were taken into account (Table 5). However, the opposite pattern in boys and girls is interesting, and may have implications for the design of future studies of sex differences in children's toy preferences, with the implication that some toys classified as 'gender-neutral' may appeal more to boys than girls.

In general, the studies included in the meta-analysis varied with respect to the number of toys chosen as representative of each toy-type category, the degree to which they are stereotyped to one gender and the duration of play sessions. Only three studies did not include gender-neutral toys and so no comparison of effect sizes between these and the other studies was conducted.

4.6 Limitations

One problem facing researchers is the choice of toys to include as gender-typed or gender-neutral and the number of toys offered. The most popular gender-typed toys are doll and vehicles, yet each of these come in many forms; for example, a doll can represent a baby, a male or female child or adult and be soft or rigid in composition and vary in colour. The suitability and attractiveness of toys is likely to differ with child age and some studies use the same stimuli across a relatively wide age range. In one study, the amount of time parents spent playing with specific toys in interactions with their children (aged 27-64 months) indicated that female-typed toys were the least entertaining (Idle et al.,1993) and this factor may have the potential to bias results.

In many individual studies the selection of stimuli is not discussed. However, one reported solution has been to use toys which have elicited gender differences in previous studies (e.g. Lamminmäki et al., 2012; Zosuls et al., 2014) or have been classified as masculine or feminine by parents and non-parents (e.g. Todd et al., 2016; van de Beek et al., 2009). Toys typically selected as neutral include books and puzzles but there is some overlap between categories, for example, Turner, Gervai & Hinde (1993) define a jigsaw puzzle as female-typed. Perhaps the most thorough consideration of the choice of stimuli comes from Idle et al. (1993); in this study, parents were asked to sort the stimuli toys into masculine, feminine and neutral categories. Interestingly, both mothers and fathers categorized many more toys as neutral than the researchers did. However, despite methodological variation between observational studies in the choice and number of stimuli offered, context of testing and age of child, the general consistency in finding sex differences in children's preferences for toys typed to their own gender indicates the strength of this phenomenon.

There was a substantial heterogeneity in findings from the included studies – roughly 70% for both male-typed and female-typed toys. Although any effects of this were mitigated by using the random effects model, we note that the fixed methods model produced slightly

smaller Cohen's *d* values for 'male-typed' toys (1.026, 95% CI 0.82 – 1.24) and female-typed toys (-0.91, 95% CI -1.12 - -0.40). On the whole, it would probably be reasonable to suggest that the effect size for the sex difference in preference for male-typed toys is large, and for female-typed toys is moderate-to-large.

Some studies of children's free preference for toys (for example O'Brien & Huston, 1985; Roopnarine, 1986) were excluded from the meta-analysis because reports did not include essential information, for example standard deviations were not reported and could not be obtained from the authors; these tended to be earlier studies, which may have affected interpretations based on the effect of year of publication.

This meta-analysis comprised papers relating to children's free preferences for gender-typed toys and some did not meet the criteria for inclusion specified in Section 2. Many of the excluded studies showed interesting results, largely consistent with the findings of the included papers. For example, cross-cultural studies are particularly informative on the effect of culture; a conducted in Hungary and the UK (Turner, Gervai & Hinde, 1993) was excluded because children were in interaction with other children. The impact of hormonal variation on the toy preferences of typically developing children allows exploration of the biological explanation; a German study measured the association between digit ratio (2D:4D),as a measure of early testosterone exposure, and sex-typed play, but used a measure completed by parents rather than the observational method (Hönekopp & Thierfelder, 2009).

It could be argued that a bias to publish only positive findings of sex difference in toy choice is unlikely as a null finding in any age-group or context would be of interest because of the social and political implications of stereotyping. Some non-typical preferences are reported in studies which did not meet the inclusion criteria of this meta-analysis, for example, at 12 months, boys played with a vacuum cleaner more than girls and no significant difference were found in the time boys and girls played with a truck or a shovel (Snow,

Maccoby & Jacklin, 1983). In the included studies, as stated above, some atypical results were found and, of course, there was considerable variation in individual children's preferences.

There was no support for the hypothesis that the presence of adults will affect the magnitude of preference for gender-typed toys. However, fathers, and possibly male observers, are underrepresented in studies comparing boys' and girls' toy preferences in specific social contexts and this may limit the interpretation with respect to this factor.

One important point to note is that all of the studies included in this meta-analysis and, indeed, most studies of children's toy preferences are conducted in Western countries. It is important therefore to extend toy preference research to other geographical locations.

4.7 Methodological Concerns

It should be noted that the GII is a blunt measure of societal attitudes in a particular country; it is likely that environmental influences may vary more by regional and cultural sub-group norms within countries than between some of the counties countries where the included studies were conducted.

In general, the typicality of the participant samples can be questioned; participants' families may typically represent comparatively high socio-economic groups and therefore associated with non-traditional sex-typed attitudes (Zosuls et al., 2009). Nevertheless, this is not reflected in the findings of large sex differences in toy preferences in the meta-analysis. Studies involving relatives of children recruited in clinical contexts are likely to be drawn from the broader population (Berenbaum & Hines, 1992; Berenbaum & Snyder, 1995; Doering et al., 1989; Pasterski et al., 2005), as are studies conducted in multicultural

nurseries (Todd et al., 2016). One possible caution is that parents may encourage gender-stereotyped play more outside a research setting.

4.8 Directions for Future Research

Key recommendations for future research include the following:

(a) We recommend attention to age of child when making comparisons between typically developing children and those with atypical levels of hormones or other specific characteristics.

(b) We argue for the value of longer-term longitudinal studies in addressing the relationship between boys' and girls' toy preferences and the development of gendered identity. However, such studies are compromised by repeated exposure to same/similar toys, suitability of toys for the different age groups and the relative attractiveness of toys (Lobel & Menashri, 1993).

(c) We suggest that randomised controlled trials (RCTs) might be useful in testing for: geographical variation; typical and atypical children; and differences in behaviour across contexts of testing, for example study locations and social contexts. Where younger children are tested playing 'alone', ethical and practical constraints prevent researchers from isolating a young child completely and it may be more useful to test between play alone and play in social contexts in a safe and familiar home environment.

(d) We suggest attention to selection of test stimuli, particularly to the inclusion/non-inclusion and characteristics of toys classified as gender-neutral. Meta-regression reveals that the inclusion of gender-neutraltoys may have differential impact (albeit statistically nonsignificantly) on boys' and girls' toy preferences, with boys playing less with male-typed toys when gender-neutral toys are included and girls playing more with male-typed toys when gender-neutral toys are included. Therefore future studies could include a condition where gender-neutral toys are present and a condition where they are absent.

In general, suggestion for current researchers in this area include greater attention to the choice of stimuli for testing, as was exemplified in Idle et al (1993), a study with high methodological quality a replication of which might shed light on the changing perceptions and behaviours of parents and children across a range of ages. An extension of studies on other primates might provide further evidence for or against the biological theories.

5 Conclusion

In view of the societal interest in, and clinical implications of sex-typed behaviour, it is important to determine its origins: the present review analysed studies of boys' and girls' (aged between one and eight years) free preference for 'gender-typed' toys, asking questions related to the degree to which sex, age and various social factors impacted on their behaviour. In general, the finding of robust sex differences in boys' and girls' toy preferences across a range of ages, different time periods, countries and settings indicates an innate influence on this behaviour, an effect which appears to be subject to modification by developmental and social factors arising at different ages.

In observational studies of the type reviewed here, it is impossible to determine the degree to which findings are a result of biological predisposition or environment. However, this study attempted to assess the variables most likely to be indicative of either nature or nurture, and must conclude that there is some (circumstantial) evidence for both sides of the argument. When assessing the effect of Publication Year, we suggest that there is more evidence of the effect of environment on girls' than on boys' toy preferences. In assessing the effects of Age, it could be argued that social effects on boys are stronger or persist longer than those on girls, or alternatively, that there is a stronger biological predisposition for boys' attraction to particular object features. However, these suggestions are necessarily speculative and we recommend that experimental evidence is required in order to shed light on the

specific questions which are identified in the meta-analysis, including the nuances of social influence.

Vygotsky (1978) regarded toys as cultural artefacts or ‘tools of the mind’ which are designed with specific skills, either cognitive or social, in mind. One implication of the findings of the meta-analysis is that it may be important to ensure that toys which are attractive to each sex are not restrictive in the skills that they afford.

References

Alexander, G. M. (2014). Postnatal testosterone concentrations and male social development. *Frontiers in Endocrinology, 5,* 15. doi:10.3389/fendo.2014.00015

Alexander, G. M. & Hines, M. (2002). Sex differences in response to children´s toys in nonhuman primates (Cercopithecus aethiops sabaeus). *Evolution and Human Behavior*, *23*, 467-479.

Alexander, G. M., & Saenz, J. (2012). Early androgens, activity levels and toy choices of children in the second year of life. *Hormones and Behavior*, *62*(4), 500–504.

Arthur, A. E. (2014). Does social categorization affect toddlers' play preferences? an experimental test. Dissertation Abstracts International: Section B: The Sciences and Engineering, 74(11-B(E)),

Auyeung, B., Baron-Cohen, S., Ashwin, E., Knickmeyer, R., Taylor, K., Hackett, G., & Hines, M. (2009). Fetal testosterone predicts sexually differentiated childhood behavior in girls and in boys. *Psychological Science, 20*(2), 144–148.

Bakir, A., Blodgett, J. G., & Rose, G. M. (2008). Children's responses to gender-role stereotyped advertisements. *Journal of Advertising Research, 48*(2), 255–266.

Banerjee, R., & Lintern, V. (2000). Boys will be boys: The effect of social evaluation concerns on gender-typing. *Social Development*, *9*(3), 397-408.

Benenson, J. F., Tennyson, R., & Wrangham, R. W. (2011). Male more than female infants imitate propulsive motion. *Cognition, 121,* 262-267.

Benjamin, H. (1932). Age and sex differences in the toy-preferences of young children. *The Pedagogical Seminary and Journal of Genetic Psychology*, *41*(2), 417–429.

Berenbaum, S. A., & Hines, M. (1992). Early androgens are related to childhood sex-typed toy preferences. *Psychological Science*, *3*(3), 203–206.

Berenbaum, S. A., & Snyder, E. (1995). Early hormonal influences on childhood sex-typed activity and playmate preferences: Implications for the development of sexual orientation. *Developmental Psychology*, *31*(1), 31.

Blakemore, J. E., LaRue, A. A., & Olejnik, A. B. (1979). Sex-appropriate toy preference and the ability to conceptualize toys as sex-role related. *Developmental Psychology, 15*(3), 339.

Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist, 32*(7), 513.

Campbell, D. W., & Eaton, W. O. (1999). Sex differences in the activity level of infants. *Infant and Child Development, 8*(1), 1–17.

Cohen-Bendahan, C. C., van de Beek, C., & Berenbaum, S. A. (2005). Prenatal sex hormone effects on child and adult sex-typed behavior: methods and findings. *Neuroscience & Biobehavioral Reviews, 29*(2), 353–384.

Collaer, M. L., & Hines, M. (1995). Human behavioral sex differences: a role for gonadal hormones during early development? *Psychological Bulletin*, *118*(1), 55.

Diekman, A. B., & Murnen, S. K. (2004). Learning to be little women and little men: The inequitable gender equality of nonsexist children's literature. *Sex Roles, 50*(5-6), 373–385.

Doering, R. W., Zucker, K. J., Bradley, S. J., & MacIntyre, R. B. (1989). Effects of neutral toys on sex-typed play in children with gender identity disorder. *Journal of Abnormal Child Psychology*, *17*(5), 563–574.

Else-Quest, N. M., Hyde, J. S., Goldsmith, H. H., & Van Hulle, C. A. (2006). Gender differences in temperament: a meta-analysis. *Psychological Bulletin*, *132*(1), 33.

Etaugh, C., & Liss, M. B. (1992). Home, school, and playroom: Training grounds for adult gender roles. *Sex Roles, 26*(3-4), 129–147.

Fabes, R. A., Martin, C. L., & Hanish, L. D. (2003). Young Children's Play Qualities in Same-, Other-, and Mixed-Sex Peer Groups. *Child Development, 74*(3), 921–932.

Fagot, B. I. (1974). Sex differences in toddlers' behavior and parental reaction. *Developmental Psychology*, *10*(4), 554.

Fagot, B. I., & Hagan, R. (1991). Observations of parent reactions to sex-stereotyped behaviors: Age and sex effects. *Child Development*, 62(3), 617-628.

Fagot, B. I., & Leinbach, M. D. (1989). The young child's gender schema: Environmental input, internal organization. *Child Development*, 60(3), 663-672.

Ferrant, G. (2010). The Gender Inequalities Index (GII) as a new way to measure Gender Inequalities in Developing countries. Retrieved from https://halshs.archives-ouvertes.fr/halshs-00462463

Halim, M. L., Ruble, D., Tamis-LeMonda, C., & Shrout, P.E. (2013). Rigidity in gender-typed behaviors in early childhood: A longitudinal study of ethnic minority children. *Child Development*, 84, 1269–1284. doi:10.1111/cdev.12057Harbord, R. M., & Higgins, J. P. (2008). Meta-regression in Stata. *Meta*, *8*(4), 493–519.

Hassett, J.M., Siebert, E.R. & Wallen, K. (2008). Sex Differences in rhesus monkey toy preferences parallel those of children. *Hormones and Behavior, 54,* 359-364

Henderson, B. A., & Berenbaum, S. A. (1997). Sex-typed play in opposite-sex twins. *Developmental Psychobiology*, *31*(2), 115–123.

Hines, M. (2004). *Brain gender*. Oxford University Press.

Hines, M., Brook, C., & Conway, G. S. (2004). Androgen and psychosexual development: Core gender identity, sexual orientation, and recalled childhood gender role behavior in women and men with congenital adrenal hyperplasia (CAH). *Journal of Sex Research*, *41*(1), 75–81.

Jadva, V., Hines, M., & Golombok, S. (2010). Infants' preferences for toys, colors, and shapes: Sex differences and similarities. *Archives of Sexual Behavior*, 39(6), 1261-1273.

Karraker, K. H., Vogel, D. A., & Lake, M. A. (1995). Parents' gender-stereotyped perceptions of newborns: The eye of the beholder revisited. *Sex Roles*, *33*(9-10), 687–701.

Lamminmäki, A., Hines, M., Kuiri-Hänninen, T., Kilpeläinen, L., Dunkel, L., & Sankilampi, U. (2012). Testosterone measured in infancy predicts subsequent sex-typed behavior in boys and in girls. *Hormones and Behavior*, *61*(4), 611–616.

Langlois, J. H., & Downs, A. C. (1980). Mothers, fathers, and peers as socialization agents of sex-typed play behaviors in young children. *Child Development*, 1237–1247.

Lauer, J. E., Udelson, H. B., Jeon, S. O., & Lourenco, S. F. (2015). An early sex difference in the relation between mental rotation and object preference. *Frontiers in Psychology*, 6, 558.

Lavelli, M., & Fogel, A. (2002). Developmental changes in mother-infant face-to-face communication: birth to 3 months. *Developmental Psychology*, *38*(2), 288.

Leaper, C., Breed, L., Hoffman, L., & Perlman, C. A. (2002). Variations in the Gender-Stereotyped Content of Children's Television Cartoons Across Genres. *Journal of Applied Social Psychology*, *32*(8), 1653–1662.

Leeb, R. T., & Rejskind, F. G. (2004). Here's looking at you, kid! A longitudinal study of perceived gender differences in mutual gaze behavior in young infants. *Sex Roles*, *50*(1-2), 1–14.

Rebecca Y Hei Li, & Wong, W. I. (2016). Gender-typed play and social abilities in boys and girls: Are they related? *Sex Roles, 74*(9-10), 399. doi:10.1007/s11199-016-0580-7

Lobel, T. E., & Menashri, J. (1993). Relations of conceptions of gender-role transgressions and gender constancy to gender-typed toy preferences. *Developmental Psychology*, *29*(1), 150.

Lytton, H., & Romney, D. M. (1991). Parents' differential socialization of boys and girls: A meta-analysis. *Psychological Bulletin, 109*(2), 267.

Martin, C. L., & Ruble, D. (2004). Children's search for gender cues cognitive perspectives on gender development. *Current Directions in Psychological Science*, *13*(2), 67–70.

Martin, C. L., & Ruble, D. N. (2010). Patterns of gender development. *Annual Review of Psychology, 61*, 353.

McClure, E. B. (2000). A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. *Psychological Bulletin*, *126*(3), 424.

Meyer-Bahlburg, H. F. L., Dolezal, C., Baker, S. W., Carlson, A. D., Obeid, J. S., & New, M. I. (2004). Prenatal androgenization affects gender-related behavior but not gender identity in 5–12-year-old girls with congenital adrenal hyperplasia. *Archives of Sexual Behavior, 33*(2), 97-104. doi:10.1023/B:ASEB.0000014324.25718.51

Miller, C. L. (1983). Developmental changes in male/female voice classification by infants. *Infant Behavior and Development*, 6(2), 313-330.

Moore, D. S., & Johnson, S. P. (2008). Mental rotation in human infants a sex difference. *Psychological Science*, *19*(11), 1063–1066.

Nagy, E., Kompagne, H., Orvos, H., & Pal, A. (2007). Gender-related differences in neonatal imitation. *Infant and Child Development*, *16*(3), 267–276.

Nordenstrom, A., Servin, A., Bohlin, G., Larsson, A., & Wedell, A. (2002). Sex-typed toy play behavior correlates with the degree of prenatal androgen exposure assessed by CYP21 genotype in girls with congenital adrenal hyperplasia. *The Journal of Clinical Endocrinology & Metabolism*, *87*(11), 5119–5124.

O'Brien, M., & Huston, A. C. (1985). Development of sex-typed play behavior in toddlers. *Developmental Psychology*, *21*(5), 866.

O'Brien, M., Huston, A. C., & Risley, T. R. (1983). Sex-typed play of toddlers in a day care center. *Journal of Applied Developmental Psychology*, 4(1), 1-9.

Pasterski, V. L., Geffner, M. E., Brain, C., Hindmarsh, P., Brook, C., & Hines, M. (2005). Prenatal hormones and postnatal socialization by parents as determinants of male-typical toy play in girls with congenital adrenal hyperplasia. *Child Development*, *76*(1), 264–278.

Quinn, P. C., & Liben, L. S. (2008). A sex difference in mental rotation in young infants. *Psychological Science*, *19*(11), 1067–1070.

Quinn, P. C., & Liben, L. S. (2014). A sex difference in mental rotation in infants: convergent evidence. *Infancy*, *19*(1), 103–116.

Quinn, P. C., Yahr, J., Kuhn, A., Slater, A. M., & Pascalis, O. (2002). Representation of the gender of human faces by infants: A preference for female. *PERCEPTION-LONDON-*, *31*(9), 1109–1122.

Roopnarine, J. L. (1986). Mothers' and fathers' behaviors toward the toy play of their infant sons and daughters. *Sex Roles*, *14*(1-2), 59–68.

Rubin, J. Z., Provenzano, F. J., & Luria, Z. (1974). The eye of the beholder: Parents' views on sex of newborns. *American Journal of Orthopsychiatry*, *44*(4), 512.

Rust, J., Golombok, S., Hines, M., Johnston, K., Golding, J., Team, A. S., et al. (2000). The role of brothers and sisters in the gender development of preschool children. *Journal of Experimental Child Psychology*, *77*(4), 292–303.

Schau, C. G., Kahn, L., Diepold, J. H., & Cherry, F. (1980). The relationships of parental expectations and preschool children's verbal sex typing to their sex-typed toy play behavior. *Child Development*, 266–270.

Schwartz, S. H., & Rubel-Lifschitz, T. (2009). Cross-national variation in the size of sex differences in values: effects of gender equality. *Journal of Personality and Social Psychology*, *97*(1), 171.

Schweinle, A., & Wilcox, T. (2004). Sex differences in infants' ability to represent complex event sequences. *Infancy*, *6*(3), 333–359.

Serbin, L. A., Connor, J. M., Burchardt, C. J., & Citron, C. C. (1979). Effects of peer presence on sex-typing of children's play behavior. *Journal of Experimental Child Psychology*, *27*(2), 303–309.

Serbin, L. A., Connor, J. M., & Citron, C. C. (1981). Sex-differentiated free play behavior: Effects of teacher modeling, location, and gender. *Developmental Psychology*, 17(5), 640-646

Serbin, L. A., Poulin-Dubois, D., Colburne, K. A., Sen, M. G., & Eichstedt, J. A. (2001). Gender stereotyping in infancy: Visual preferences for and knowledge of gender-stereotyped toys in the second year. *International Journal of Behavioral Development*, 25(1), 7-15.

Servin, A., Bohlin, G., & Berlin, L. (1999). Sex differences in 1-, 3-, and 5-year-olds' toy-choice in a structured play-session. *Scandinavian Journal of Psychology*, *40*(1), 43–48.

Servin, A., Nordenström, A., Larsson, A., & Bohlin, G. (2003). Prenatal androgens and gender-typed behavior: a study of girls with mild and severe forms of congenital adrenal hyperplasia. *Developmental Psychology*, *39*(3), 440.

Shell, R., & Eisenberg, N. (1990). The role of peers' gender in children's naturally occurring interest in toys. *International Journal of Behavioral Development*, *13*(3), 373–388.

Signorella, M. L., Bigler, R. S., & Liben, L. S. (1993). Developmental Differences in Children′ s Gender Schemata about Others: A Meta-analytic Review. *Developmental Review*, *13*(2), 147–183.

Slaby, R. G., & Frey, K. S. (1975). Development of gender constancy and selective attention to same-sex models. *Child Development*, 46(4), 849-856.

Smith, P. K., & Daglish, L. (1977). Sex differences in parent and infant behavior in the home. *Child Development*, 1250–1254.

Snow, M. E., Jacklin, C. N., & Maccoby, E. E. (1983). Sex-of-child differences in father-child interaction at one year of age. *Child Development*, 227–232.

Spelke, E. S. (2005). Sex differences in intrinsic aptitude for mathematics and science?: a critical review. *American Psychologist*, *60*(9), 950.

Table 4: Gender Inequality Index | Human Development Reports. (n.d.). Retrieved June 10, 2015, from http://hdr.undp.org/en/content/table-4-gender-inequality-index

Todd, B. K., Barry, J. A., & Thomason, S. A. O. (2016). Preferences for 'Gender-typed' toys in boys and girls aged 9 to 32 months: Young children's 'Gender-typed' toy preferences. *Infant and Child Development,* doi:10.1002/icd.1986

Touwen, B. C. L. (1976). *Neurological development in infancy* (Vol. 58). Heinemann Educational Books.

Trautner, H. M., Ruble, D. N., Cyphers, L., Kirsten, B., Behrendt, R., & Hartmann, P. (2005). Rigidity and flexibility of gender stereotypes in childhood: Developmental or differential? *Infant and Child Development*, *14*(4), 365–381.

Turner, P. J., & Gervai, J. (1995). A multidimensional study of gender typing in preschool children and their parents: Personality, attitudes, preferences, behavior, and cultural differences. *Developmental Psychology*, *31*(5), 759.

Turner, P. J., Gervai, J., & Hinde, R. A. (1993). Gender-typing in young children: Preferences, behaviour, and cultural differences. *British Journal of Developmental Psychology*, *11*, 323-342.

Van de Beek, C., van Goozen, S. H., Buitelaar, J. K., & Cohen-Kettenis, P. T. (2009). Prenatal sex hormones (maternal and amniotic fluid) and gender-related play behavior in 13-month-old infants. *Archives of Sexual Behavior*, *38*(1), 6–15.

Vygotsky, L. S. (1978). *Mind and Society: the Development of Higher Mental Processes,(edited by Cole, M., et. al)*. Harvard University Press, Cambridge, Massachussets.

Whiting, B. B., Edwards, C. P., & Edwards, C. P. (1992). *Children of different worlds: The formation of social behavior*. Harvard University Press. Retrieved from https://books.google.co.uk/books?hl=en&lr=&id=b96jDjN3M_MC&oi=fnd&pg=PR7&dq=Children+of+different+worlds:+The+formation+of+social+behavior.+&ots=PN1r5D-W4j&sig=oIqGb1bn-6JckgUY9ddllGBQmec

Wilansky-Traynor, P., & Lobel, T. E. (2008). Differential effects of an adult observer's presence on sex-typed play behavior: A comparison between gender-schematic and gender-aschematic preschool children. *Archives of Sexual Behavior*, *37*(4), 548–557.

Wilcox, T. (2003). Event-mapping tasks: Investigating the effects of prior information and event complexity on performance. *Infant Behavior and Development*, *26*(4), 568–587.

Wolf, T. M. (1973). Effects of live modeled sex-inappropriate play behavior in a naturalistic setting. *Developmental Psychology*, *9*(1), 120.

Yarrow, M. R., Campbell, J. D., & Burton, R. V. (1970). Recollections of childhood a study of the retrospective method. *Monographs of the Society for Research in Child Development*, iii–83.

Zosuls, K. M., Ruble, D. N., Tamis-LeMonda, C. S., Shrout, P. E., Bornstein, M. H., & Greulich, F. K. (2009). The acquisition of gender labels in infancy: Implications for gender-typed play. *Developmental Psychology*, *45*(3), 688.

Zosuls, K. M., Ruble, D. N., & Tamis-LeMonda, C. S. (2014). Self-Socialization of gender in african american, dominican immigrant, and mexican immigrant toddlers. *Child Development*, *85*(6), 2202-2217. doi:10.1111/cdev.12261

Table 1a. Information extracted from the included papers with reference to 'male-typed' toys.

Study	Age (months)	Boys N	Boys Mean play	Boys SD	Girls N	Girls Mean play	Girls SD	Neutral toys	Year	Publication years	Parent present	Setting	Co
Alexander & Saenz (2012)	19.5	47	438.72	166.08	47	248.64	161.28	1	2012	3	3	2	.
Arthur (2014)	19	22	148	105	18	54	53	0	2014	2	2	2	.
Berenbaum & Hines (1992)	65.79	18	325.25	270.21	18	135.13	201.72	1	1992	23	2	2	.
Berenbaum & Snyder (1995)	93.1	19	484.21	141.28	19	202.48	168.14	1	1995	20	2	1	.
Doering et al (1989)	87	15	260.1	67.7	15	98.4	101.3	1	1989	26	1	2	.
Idle et al. (1993)	46	10	393.2	136.2	10	271.2	139.7	1	1993	22	3	3	.
Lamminmäki & Hines (2012)	14	21	46.2	59.67	21	33.54	47.16	1	2012	3	2	2	.
Li & Wong (2016)	77	38	214.7	132.26	38	43.38	85.79	1	2016	0	1	1	H Kong
Pasterski et al. (2005)	81	12	244.8	118.8	12	7.2	14.4	1	2005	10	3	2	.
Pasterski et al. (2005)	81	25	144	144	25	43.2	104.4	1	2005	10	3	2	.
Pasterski et al. (2005)	81	24	252	100.8	24	46.8	93.6	1	2005	10	1	2	.

Schau et al. (1980)	48	26	357.04	209.02	26	131.48	107.9	1	1980	35	2	2	.
Servin et al. (1999)	12	19	182.29	103.93	19	114.48	75.02	1	1999	16	3	2	.
Servin et al. (1999)	36	13	279.88	67.09	13	139.75	77.45	1	1999	16	3	2	.
Servin et al. (1999)	60	14	244.07	123.3	14	150.42	104.55	1	1999	16	3	2	.
Todd et al. (2016)	13	20	31.65	29.15	20	5.75	11.55	0	2014	0	2	3	.
Todd et al. (2016)	21.5	18	46.95	29.7	18	12.3	17.65	0	2014	0	2	3	.
Todd et al. (2016)	28	16	75	28.6	16	13.6	17.6	0	2014	0	2	3	.
van de Beek (2009)	13	63	169.98	118.23	63	112.47	94.54	1	2009	6	2	2	.
Wilansky-Traynor & Lobel (2008)	65.83	26	158.33	80.79	26	80.45	66.38	0	2008	7	2	3	.
Wilansky-Traynor & Lobel (2008)	65.83	31	130.89	89.36	31	82.01	75.83	0	2008	7	1	3	.
Zosuls et al. (2009)	18	36	72	90	36	30	36	1	2009	6	2	1	.
Zosuls et al. (2009)	22	36	78	114	36	30	42	1	2009	6	2	1	.
Zosuls et al. (2009)	18	36	54	84	36	42	54	1	2009	6	3	1	.
Zosuls et al. (2009)	22	36	66	66	36	36	84	1	2009	6	3	1	.

Zosuls, Ruble & Tamis-LeMonda (2014)	24	93	64.8	56.7	99	32.4	35.1	1	2014	2	2	2	.
Zosuls, Ruble & Tamis-LeMonda (2014)	36	63	67.5	48.6	77	43.2	43.2	1	2014	2	2	2	0

Notes: N = sample size; Mean play = mean seconds played with toy; SD = standard deviation; Neutral toys (1= 'neutral toys included'; 2 = 'neutral toys not included'); Publication years = years since publication of the study; Parent present (1= absent; 2 = minimal interaction; 3= moderate or full interaction); Setting = location of study (1= home; 2 = lab; 3 = nursery); Country = Gender Inequality Index, a measure of how gender-egalitarian the country was at the time the study took place.

Table 1b. Information extracted from the included papers with reference to 'female-typed' toys.

		Boys			Girls				
Study	**Age (months)**	**N**	**Mean play**	**SD**	**N**	**Mean play**	**SD**	**Neutral toys**	**Yea**
Alexander & Saenz (2012)	19.5	47	377.28	170.88	37	538.56	214.08	1	201
Arthur (2014)	19	22	126	67	18	244	54	0	201
Berenbaum & Hines (1992)	65.79	18	33.5	69.55	15	247.93	246.5	1	199
Berenbaum & Snyder (1995)	93.1	19	26.42	91.95	13	265.42	240.71	1	199
Doering et al. (1989)	87	15	7.2	21.5	15	145.2	109.6	1	198
Idle et al. (1993)	46	10	136.6	92.4	10	140.5	126.8	1	199
Lamminmäki & Hines (2012)	14	21	80.65	68.63	26	80.09	79.6	1	201
Li & Wong (2016)	77	38	26.96	62.5	38	183.28	117.5	1	201
Pasterski et al (2005)	81	12	21.6	36	12	115.2	169.2	1	200
Pasterski et al (2005)	81	25	43.2	93.6	27	147.6	169.2	1	200
Pasterski et al (2005)	81	24	21.6	68.4	26	219.6	136.8	1	200
Schau et al. (1980)	48	26	92.08	102.25	26	166.92	159.62	1	198
Servin et al. (1999)	12	19	220.82	103.27	19	275.08	98.2	1	199
Servin et al. (1999)	36	13	73.19	56.9	18	215.83	156.92	1	199
Servin et al. (1999)	60	14	23.75	44.08	21	139.1	110.38	1	199
Todd et al. (2016)	21.5	18	7.35	13.45	11	42.05	38.35	0	201
Todd et al. (2016)	28	16	6.1	7.4	16	57.35	28.95	0	201
Todd et al. (2016)	13	20	14	17.65	20	58.9	31.85	0	201
van de Beek (2009)	13	63	151.97	117.28	63	189.82	112.96	1	200

Wilansky-Traynor & Lobel (2008)	65.83	26	63.43	72.29	30	148.3	77.38	0	200
Wilansky-Traynor & Lobel (2008)	65.83	31	61.52	74.03	29	125.03	79.32	0	200
Zosuls et al. (2009)	18	36	78	120	46	90	108	1	200
Zosuls et al. (2009)	22	36	54	60	46	150	162	1	200
Zosuls et al. (2009)	18	36	96	102	46	108	116	1	200
Zosuls et al. (2009)	22	36	78	84	46	120	136	1	200
Zosuls, Ruble & Tamis-LeMonda (2014)	24	93	29.7	40.5	99	62.1	64.8	1	201
Zosuls, Ruble & Tamis-LeMonda (2014)	36	63	32.4	43.2	77	75.6	72.9	1	201

Notes: N = sample size; Mean play = mean seconds played with toy; SD = standard deviation; Neutral toys (1= 'neutral toys included'; 2 = 'neutral toys not included'); Publication years = years since publication of the study; Parent present (1= absent; 2 = minimal interaction; 3= moderate or full interaction); Setting = location of study (1= home; 2 = lab; 3 = nursery); Country = Gender Inequality Index, a measure of how gender-egalitarian the country was at the time the study took place.

Table 2.

Evaluation of the methodological quality of the 16 studies comparing toy preference in boys and girls in studies included in the meta-analysis.

Quality is assessed using Newcastle-Ottowa Quality Assessment Scale (NOS) criteria adapted for this study (see notes below the table).

Study	Case definition adequate	Representativeness of cases	Selection of Controls	Parental opinion	Comparability of both groups	Ascertainment of behavior	Same ascertainment method for both groups	Non-Response rate	**NOS Score**
Schau et al. (1980)	*	*	*	*	*	X	*	X	6
Doering et al (1989)	*	*	*	X	*	*	*	X	6
Berenbaum & Hines (1992)	X	*	*	*	*	*	*	X	6
Idle et al (1993)	*	*	*	*	*	*	*	*	8
Berenbaum & Snyder (1995)	*	X	*	*	*	*	*	*	7
Servin et al (1999)	*	*	*	X	*	*	*	*	7
Pasterski et al (2005)	*	*	*	*	*	*	*	X	7
Wilansky-Traynor & Lobel (2008)	*	*	*	X	*	*	*	X	6
Zosuls et al. (2009)	*	*	*	X	*	*	*	*	7

van de Beek (2009)	*	*	*	X	**	*	*	*	8
Lamminmäki & Hines (2012)	*	*	*	*	*	*	*	X	7
Alexander & Saenz (2012)	*	*	X	X	*	*	*	X	5
Todd et al. Li & Wong (2016)	*	*	X	*	*	*	*	*	7
Zosuls, Ruble & Tamis-LeMonda (2014)									
Zosuls, Ruble & Tamis-LeMonda (2014)									
Arthur (2014)									

Notes: '*' indicates that the study fulfilled this criterion; 'X' indicates that the study did not fulfil this criterion.
Case definition adequate: clear justification for 'gendered nature' of toy, e.g. based on research
Representativeness of cases: recruitment of consecutive participants
Selection of Controls: whether boys and girls were comparable in terms of social background
Parental opinion: whether parents' views on gender were measured
Comparability of both groups: give a star for either of the following: the toys were comparable (in size, shape etc) and the boys and girls were comparable in age
Ascertainment of behavior: play behavior was clearly defined
Same ascertainment method for both groups: the measurement of the outcome (time spent playing with toy) was clearly defined
Non-Response rate: whether either non-uptake or dropout rates reported

Table 3.

Findings of the meta-analysis of 16 studies, with 27 groups of children in total. Subgroup analyses are based on the type of toy ('male-typed' or 'female-typed') that children played with.

Toy type	**Cohen's *d*** **(95% CI)**	**Z** ***(p)***	χ^2 ***(p)***	**I^2**
'Male-typed' Toys	1.026 (0.816 – 1.236)	9.56 ($p < 0.0001$)	96.56 ($p < 0.0001$)	3.1%
'Female-typed' Toys	-0.909 (-1.1231 – -0.395)	-8.32 ($p < 0.0001$)	104.22 ($p < 0.0001$)	5.1%

Notes: The random effects model was used due to the substantial homogeneity. Using the fixed methods model, the *d* values were smaller for 'male-typed' toys (0.861, 95% CI 0.76 – 0-97) and 'female-typed' toys (-0.759, 95% CI -0.861 - -0.657).

Table 4.
Findings of the meta-regression of 16 studies, with 27 groups of children in total. Subgroup analyses are based on the type of toy ('male-typed' or 'female-typed') that children played with.

Variable	**Male-typed Toys**			**Female-typed Toys**		
	β	SE β	*t*	β	SE β	*T*
Present	.042	. 221	0.19	.085	.184	.46
Setting	.143	.302	0.47	.096	.281	.34
Country (GII)	.068	1.900	0.04	.594	1.698	.35
Publication years	.005	.018	0.31	.000	.016	.03
Age	.010	.006	1.77	-.009	.005	-1.81
Gender neutral toys	-.250	.471	-0.53	-1.722	.844	.477

Notes: * $p < .05$, ** $p < .01$, *** $p < .001$, **** $p < .0001$ (two-tailed).
Presence = presence of parent (1= absent; 2 = minimal interaction; 3= moderate or full interaction); Setting = location of study (1= home; 2 = lab; 3 = nursery); Country = Gender Inequality Index, a measure of how gender-egalitarian the country was at the time the study took place; Publication years = years since publication of the study: Age = age of group in months; Gender neutral toys = whether or not 'gender–neutral' as well as 'gender-typed' toys were included in the study as stimuli.

Table 5.

Spearman's (bivariate) correlations between eligible predictors and the percentage of time played with 'male-typed' toys for boys (N = 27) and girls (N = 27).

	Present	**(GII)**	**Publication years**	**Age**	**Home vs Laboratory[a]**	**Nursery vs Laboratory[a]**	**Gender-neutral toys**
Boys	-.168	-.034	.464*	682****	-.456*	.184	-.126
Girls	.139	-.020	.700****	.176	-.228	-.065	.172

* $p < .05$, ** $p < .01$, *** $p < .001$, **** $p < .0001$ (two-tailed).

* $p < .05$, ** $p < .01$, *** $p < .001$, **** $p < .0001$ (two-tailed).

Notes: Presence = presence of parent (1 = absent; 2 = minimal interaction; 3 = moderate or full interaction); Setting = location of study (1 = home; 2 = lab; 3 = nursery); Country = Gender Inequality Index, a measure of how gender-egalitarian the country was at the time the study took place; Publication years = years since publication of the study: Age = age of group in months; Gender neutral toys = whether or not 'gender–neutral' as well as 'gender-typed' toys were included in the study as stimuli.

A: in order to not violate the assumptions of the regression the categorical data was split

Table 6.

Multiple linear regression for boys (left side) and girls (right side), with percentage of time played with 'male-typed' toys relative to 'female-typed' toys as the dependent variable.

Variable	**Boys**			**Girls**		
	B	SE B	β	β	SE B	β
Present	2.196	5.330	.062	2.102	3.815	.098
Setting[a]:						
– Home vs Lab	-13.730	9.849	-.271	-.514	5.810	.502
– Nursery vs Lab	6.774	11.661	.127	16.256	8.038	.502
Country (GII)	6.303	42.943	.024	1.164	1.639	.487
Publication year	.594	.479	.243	1.141*	.333	.769
Age	.464	.140	.553	-.061	.097	-.119
Gender neutral toys	-9.196	14.454	-.172	7.164	10.324	.221

* $p < .05$, ** $p < .01$ (two-tailed).

Notes: Presence = presence of parent (1 = absent; 2 = minimal interaction; 3 = moderate or full interaction); Setting = location of study (1 = home; 2 = lab; 3 = nursery); Country = Gender Inequality Index, a measure of how gender-egalitarian the country was at the time the study took place; Publication years = years since publication of the study: Age = age of group in months; Gender neutral toys = whether or not 'gender–neutral' as well as 'gender-typed' toys were included in the study as stimuli.

A: in order to not violate the assumptions of the regression the categorical data was split

Table 7a.

Spearman bivariate correlations for predictors and boys play behaviour with ‘male-typed’ toys and ‘female-typed’ toys

	Present	**(GII)**	**Publication years**	**Age**	**Home vs Laboratory**	**Nursery vs Laboratory**	**Gender-neutral toys**
‘Male-typed’	.092	.165	.712**	.540*	-.336	-.195	.377
‘Female typed’	.412*	.088	.142	-.510*	.043	-.282	.252

* $p < .05$, ** $p < .01$, *** $p < .001$, **** $p < .0001$ (two-tailed).

Notes: Presence = presence of parent (1 = absent; 2 = minimal interaction; 3 = moderate or full interaction); Setting = location of study (1 = home; 2 = lab; 3 = nursery); Country = Gender Inequality Index, a measure of how gender-egalitarian the country was at the time the study took place; Publication years = years since publication of the study: Age = age of group in months; Gender neutral toys = whether or not ‘gender–neutral’ as well as ‘gender-typed’ toys were included in the study as stimuli.

Table 7b.

Spearman bivariate correlations for predictors and girls play behaviour with ‘male-typed’ toys and ‘female-typed’ toys

	Present	(GII)	Publication years	Age	Home vs Laboratory	Nursery vs Laboratory	Gender-neutral toys
‘Male-typed’	.169	-.033	.652**	.262	-.266	-.185	.320
‘Female typed’	.056	-.028	.498*	.195	-.239	-.358	.320

* $p < .05$, ** $p < .01$, *** $p < .001$, **** $p < .0001$ (two-tailed).

Notes: Presence = presence of parent (1 = absent; 2 = minimal interaction; 3 = moderate or full interaction); Setting = location of study (1 = home; 2 = lab; 3 = nursery); Country = Gender Inequality Index, a measure of how gender-egalitarian the country was at the time the study took place; Publication years = years since publication of the study: Age = age of group in months; Gender neutral toys = whether or not ‘gender–neutral’ as well as ‘gender-typed’ toys were included in the study as stimuli.

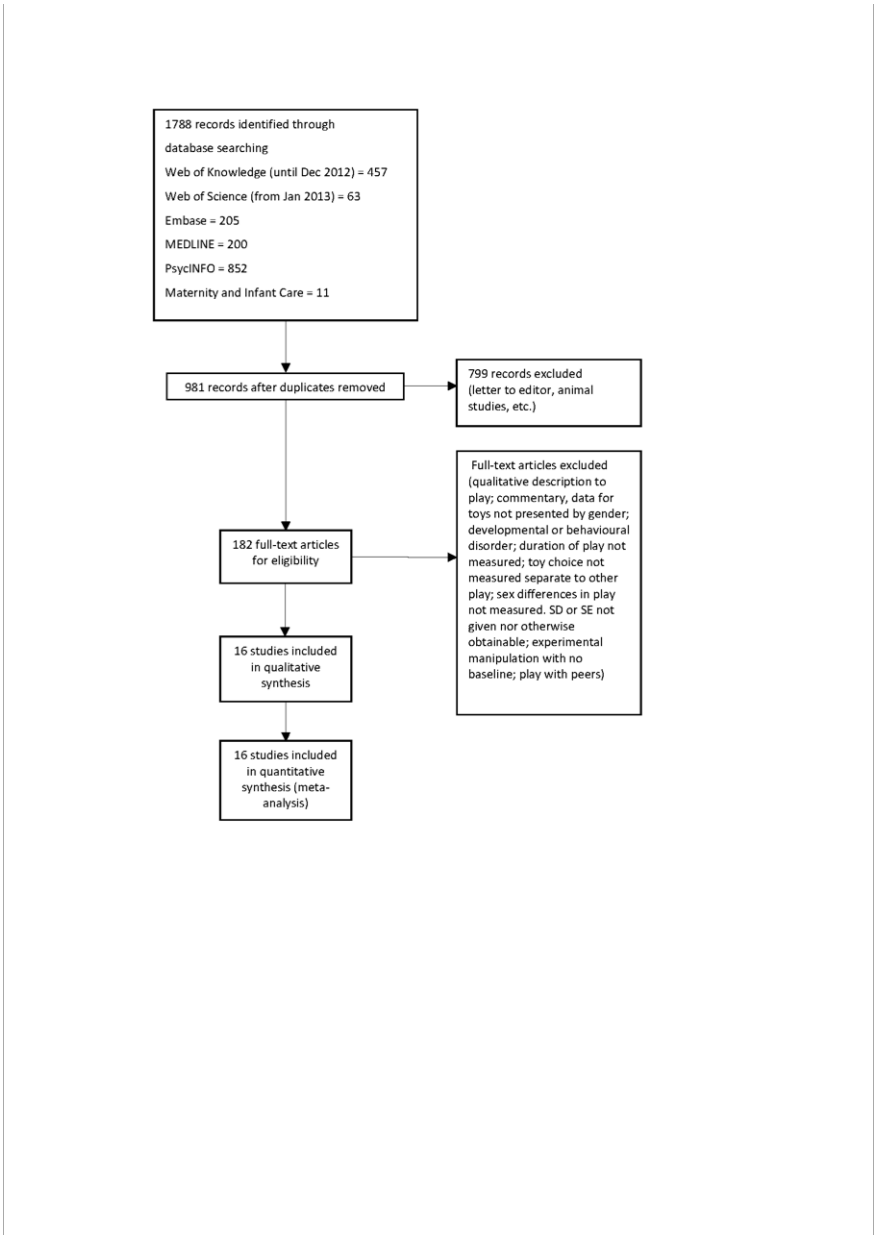
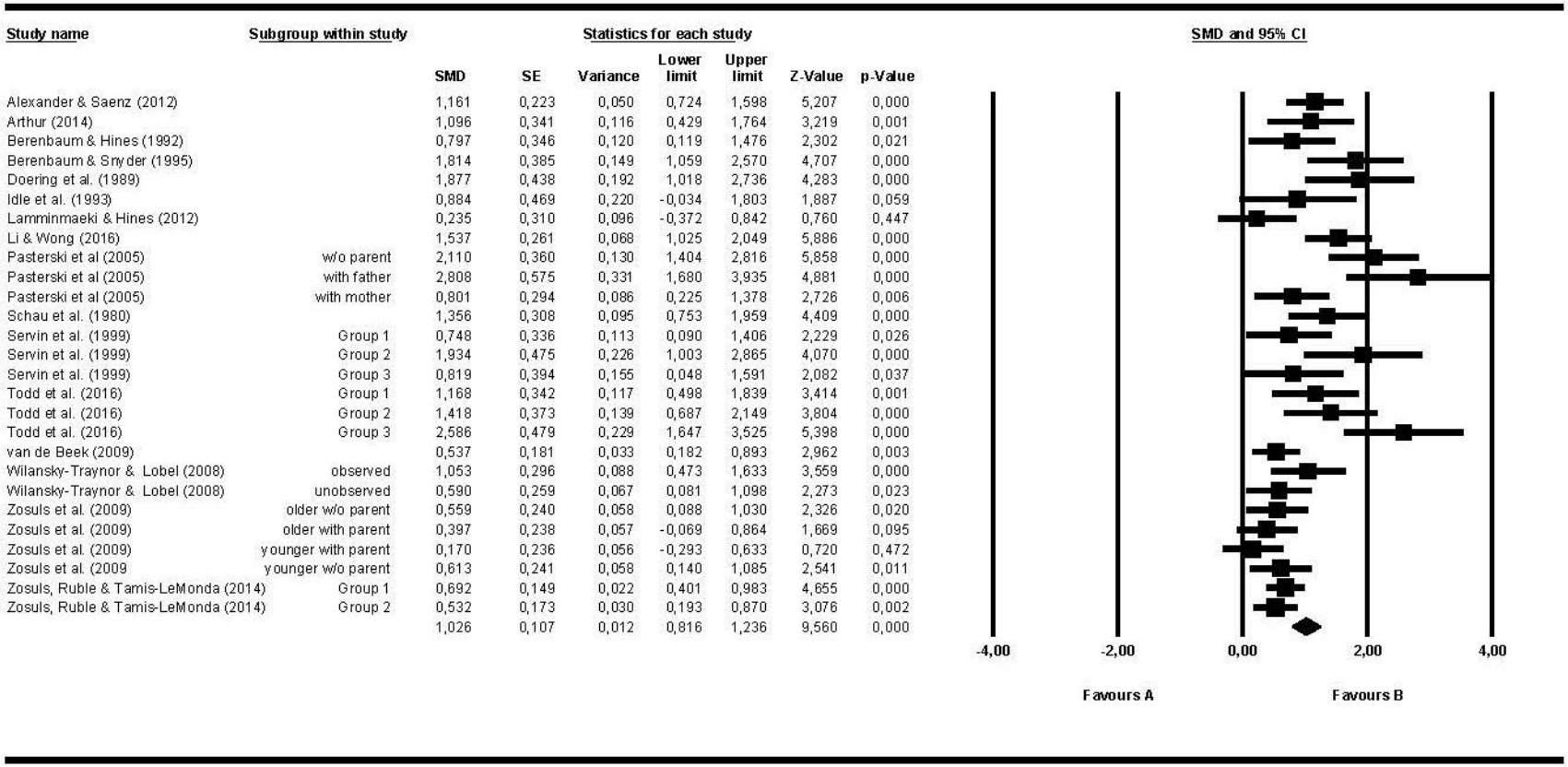

1788 records identified through
database searching
Web of Knowledge (until Dec 2012) = 457
Web of Science (from Jan 2013) = 63
Embase = 205
MEDLINE = 200
PsycINFO = 852
Maternity and Infant Care = 11
981 records after duplicates removed
799 records excluded (letter to editor, animal studies, etc.)
182 full-text articles for eligibility
Full-text articles excluded (qualitative description to play; commentary, data for toys not presented by gender; developmental or behavioural disorder; duration of play not measured; toy choice not measured separate to other play; sex differences in play not measured. SD or SE not given nor otherwise obtainable; experimental manipulation with no baseline; play with peers)
16 studies included in qualitative synthesis
16 studies included in quantitative synthesis (meta-analysis)

Figure 1

Figure 2a. Forest plot of meta-analysis of sex difference in children's play with 'male-typed' toys



Study name	Subgroup within study	SMD	SE	Variance	Lower limit	Upper limit	Z-Value	p-Value
Alexander & Saenz (2012)		1,161	0,223	0,050	0,724	1,598	5,207	0,000
Arthur (2014)		1,096	0,341	0,116	0,429	1,764	3,219	0,001
Berenbaum & Hines (1992)		0,797	0,346	0,120	0,119	1,476	2,302	0,021
Berenbaum & Snyder (1995)		1,814	0,385	0,149	1,059	2,570	4,707	0,000
Doering et al. (1989)		1,877	0,438	0,192	1,018	2,736	4,283	0,000
Idle et al. (1993)		0,884	0,469	0,220	-0,034	1,803	1,887	0,059
Lamminmaeki & Hines (2012)		0,235	0,310	0,096	-0,372	0,842	0,760	0,447
Li & Wong (2016)		1,537	0,261	0,068	1,025	2,049	5,886	0,000
Pasterski et al (2005)	w/o parent	2,110	0,360	0,130	1,404	2,816	5,858	0,000
Pasterski et al (2005)	with father	2,808	0,575	0,331	1,680	3,935	4,881	0,000
Pasterski et al (2005)	with mother	0,801	0,294	0,086	0,225	1,378	2,726	0,006
Schau et al. (1980)		1,356	0,308	0,095	0,753	1,959	4,409	0,000
Servin et al. (1999)	Group 1	0,748	0,336	0,113	0,090	1,406	2,229	0,026
Servin et al. (1999)	Group 2	1,934	0,475	0,226	1,003	2,865	4,070	0,000
Servin et al. (1999)	Group 3	0,819	0,394	0,155	0,048	1,591	2,082	0,037
Todd et al. (2016)	Group 1	1,168	0,342	0,117	0,498	1,839	3,414	0,001
Todd et al. (2016)	Group 2	1,418	0,373	0,139	0,687	2,149	3,804	0,000
Todd et al. (2016)	Group 3	2,586	0,479	0,229	1,647	3,525	5,398	0,000
van de Beek (2009)		0,537	0,181	0,033	0,182	0,893	2,962	0,003
Wilansky-Traynor & Lobel (2008)	observed	1,053	0,296	0,088	0,473	1,633	3,559	0,000
Wilansky-Traynor & Lobel (2008)	unobserved	0,590	0,259	0,067	0,081	1,098	2,273	0,023
Zosuls et al. (2009)	older w/o parent	0,559	0,240	0,058	0,088	1,030	2,326	0,020
Zosuls et al. (2009)	older with parent	0,397	0,238	0,057	-0,069	0,864	1,669	0,095
Zosuls et al. (2009)	younger with parent	0,170	0,236	0,056	-0,293	0,633	0,720	0,472
Zosuls et al. (2009	younger w/o parent	0,613	0,241	0,058	0,140	1,085	2,541	0,011
Zosuls, Ruble & Tamis-LeMonda (2014)	Group 1	0,692	0,149	0,022	0,401	0,983	4,655	0,000
Zosuls, Ruble & Tamis-LeMonda (2014)	Group 2	0,532	0,173	0,030	0,193	0,870	3,076	0,002
		1,026	0,107	0,012	0,816	1,236	9,560	0,000

Figure 2b. Forest plot of meta-analysis of sex difference in children's play with 'female-typed' toys

Study name	Subgroup within study	SMD	SE	Variance	Lower limit	Upper limit	Z-Value	p-Value
Alexander & Saenz (2012)		-0,844	0,229	0,053	-1,293	-0,395	-3,683	0,000
Arthur (2014)		-1,918	0,383	0,147	-2,669	-1,166	-5,003	0,000
Berenbaum & Hines (1992)		-1,236	0,381	0,145	-1,983	-0,489	-3,242	0,001
Berenbaum & Snyder (1995)		-1,422	0,401	0,161	-2,209	-0,635	-3,542	0,000
Doering et al. (1989)		-1,747	0,429	0,184	-2,589	-0,906	-4,071	0,000
Idle et al (1993)		-0,035	0,447	0,200	-0,912	0,841	-0,079	0,937
Lamminmaeki & Hines (2012)		0,007	0,293	0,086	-0,568	0,583	0,025	0,980
Li & Wong (2016)		-1,661	0,266	0,071	-2,183	-1,140	-6,243	0,000
Pasterski et al. (2005)	w/o parent	-1,808	0,336	0,113	-2,467	-1,150	-5,384	0,000
Pasterski et al. (2005)	with father	-0,765	0,423	0,179	-1,594	0,064	-1,809	0,070
Pasterski et al. (2005)	with mother	-0,756	0,287	0,083	-1,319	-0,193	-2,630	0,009
Schau et al. (1980)		-0,558	0,283	0,080	-1,112	-0,004	-1,975	0,048
Servin et al. (1999)	Group 1	-0,538	0,330	0,109	-1,186	0,109	-1,630	0,103
Servin et al. (1999)	Group 2	-1,136	0,392	0,153	-1,903	-0,368	-2,901	0,004
Servin et al. (1999)	Group 3	-1,278	0,377	0,142	-2,017	-0,538	-3,386	0,001
Todd et al. (2016)	Group 1	-1,744	0,371	0,138	-2,472	-1,016	-4,694	0,000
Todd et al. (2016)	Group 2	-1,352	0,422	0,178	-2,179	-0,525	-3,205	0,001
Todd et al. (2016)	Group 3	-2,426	0,466	0,217	-3,338	-1,513	-5,208	0,000
van de Beek (2009)		-0,329	0,179	0,032	-0,680	0,023	-1,833	0,067
Wilansky-Traynor & Lobel (2008)	observed	-1,131	0,288	0,083	-1,696	-0,565	-3,919	0,000
Wilansky-Traynor & Lobel (2008)	unobserved	-0,829	0,269	0,072	-1,356	-0,301	-3,079	0,002
Zosuls et al. (2009)	older w/o parent	-0,106	0,223	0,050	-0,542	0,331	-0,475	0,635
Zosuls et al. (2009)	older with parent	-0,109	0,223	0,050	-0,545	0,327	-0,489	0,625
Zosuls et al. (2009)	younger w/o parent	-0,751	0,230	0,053	-1,202	-0,300	-3,264	0,001
Zosuls et al. (2009)	younger with parent	-0,362	0,224	0,050	-0,801	0,078	-1,612	0,107
Zosuls, Ruble & Tamis-LeMonda (2014)	Group 1	-0,596	0,148	0,022	-0,885	-0,306	-4,035	0,000
Zosuls, Ruble & Tamis-LeMonda (2014)	Group 2	-0,704	0,175	0,031	-1,047	-0,361	-4,023	0,000
		-0,909	0,109	0,012	-1,123	-0,695	-8,323	0,000

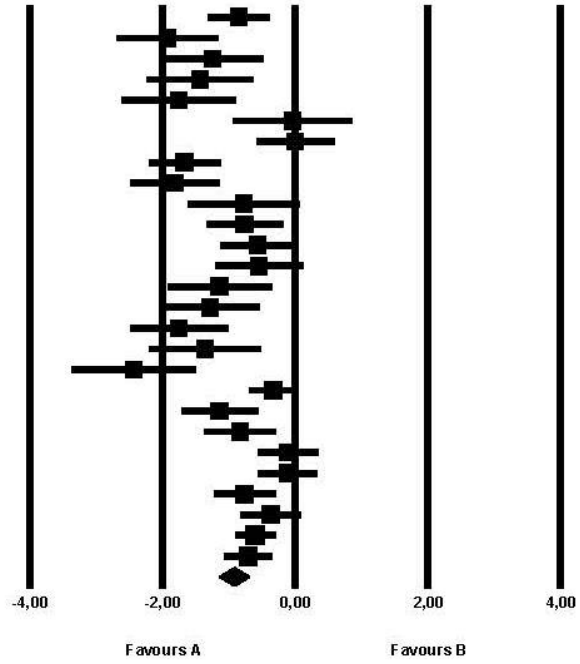


Figure 3a. Funnel plot for the sex difference in play with ‘male-typed’ toys

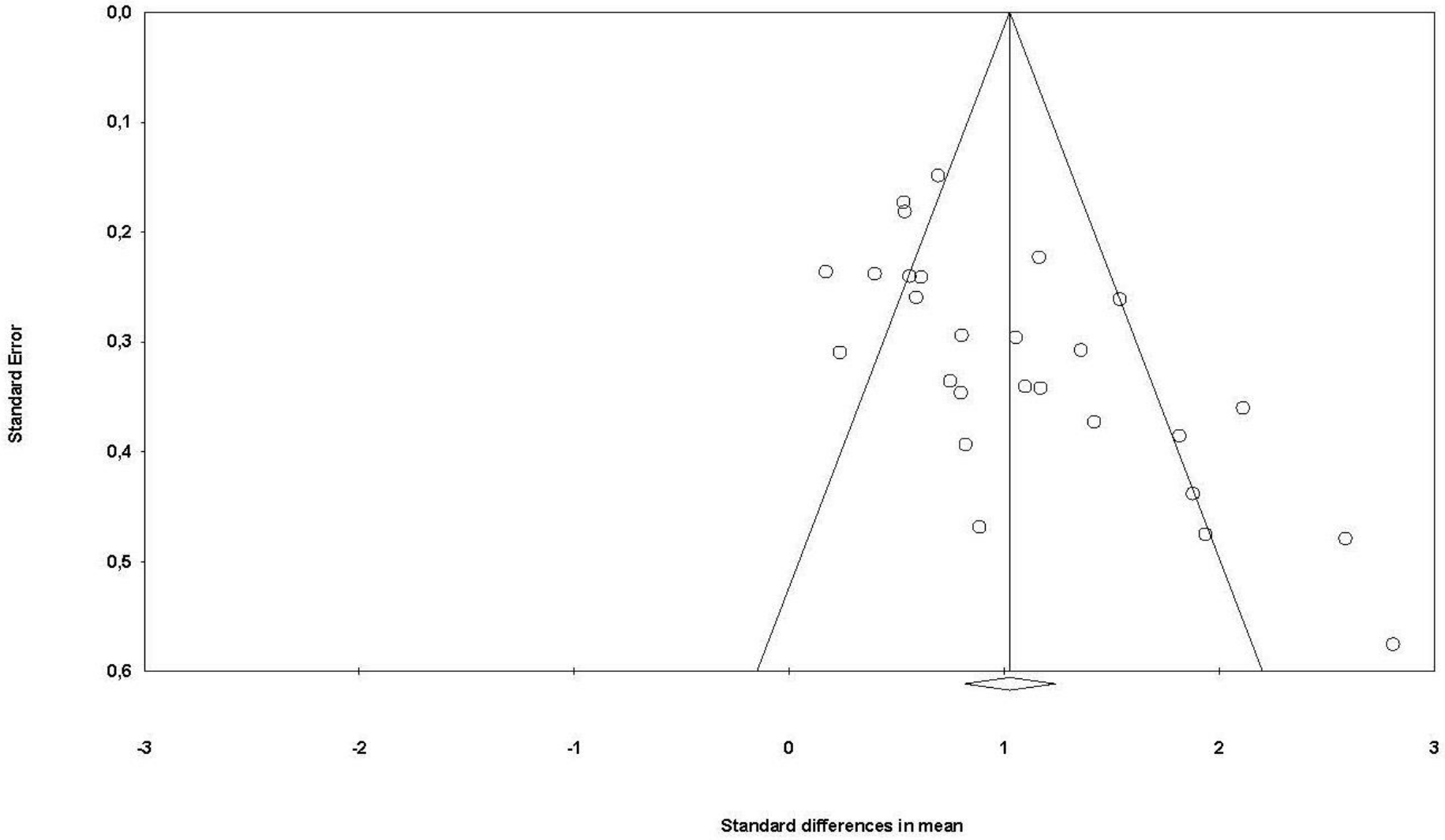


Figure 3b. Funnel plot for the sex difference in play with ‘female-typed’ toys

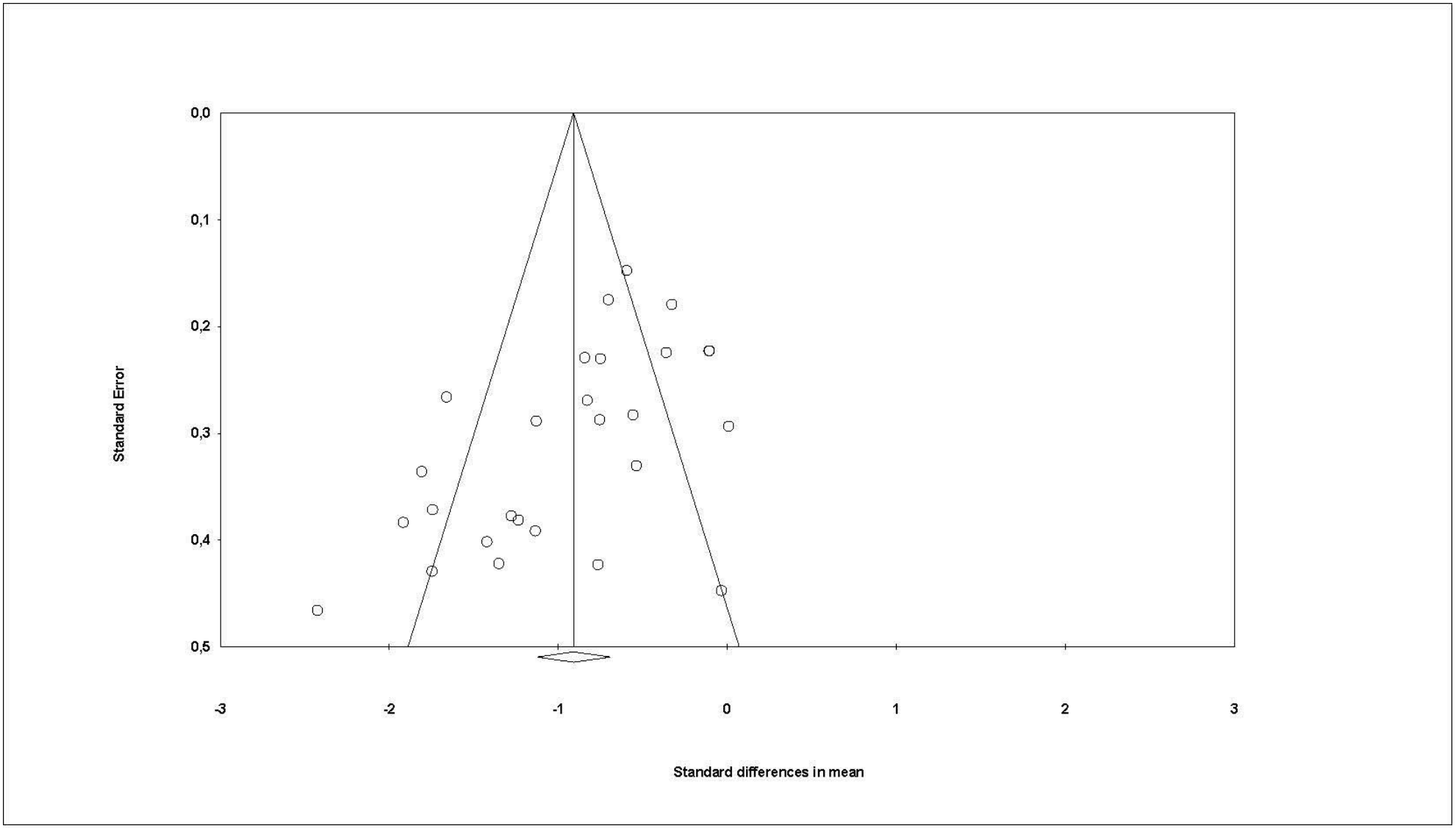
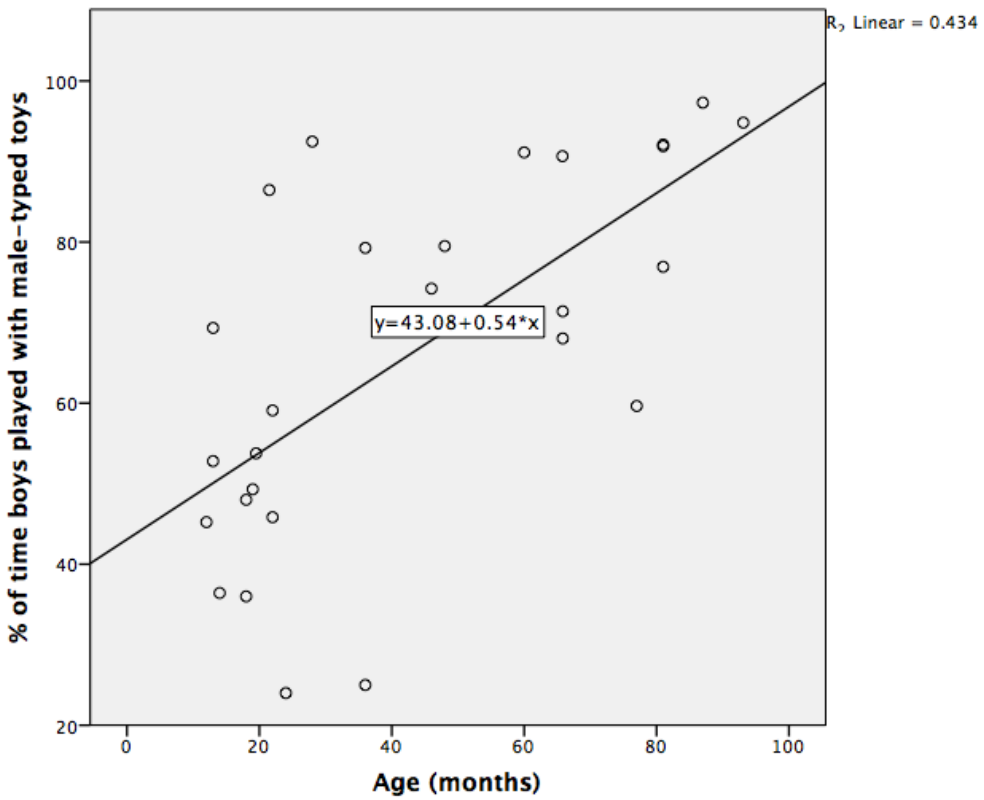


Figure 4a (left, boys) and **Figure 4b** (right, girls). Percentage of time played with male-typed toys relative to female-typed toys, by age.



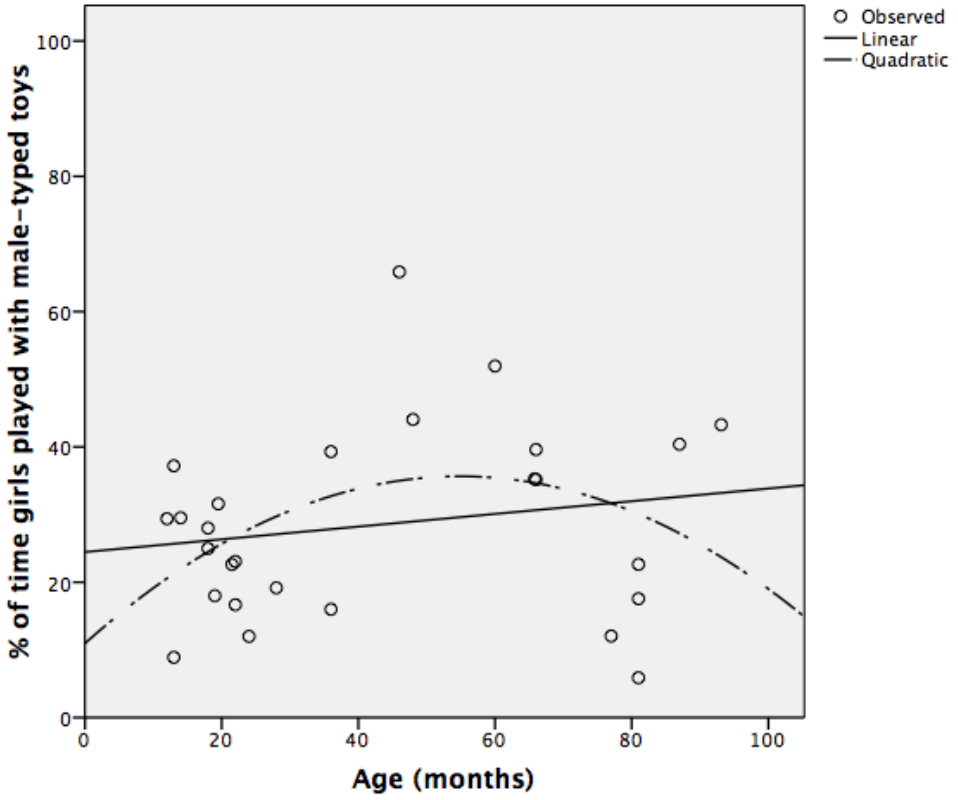
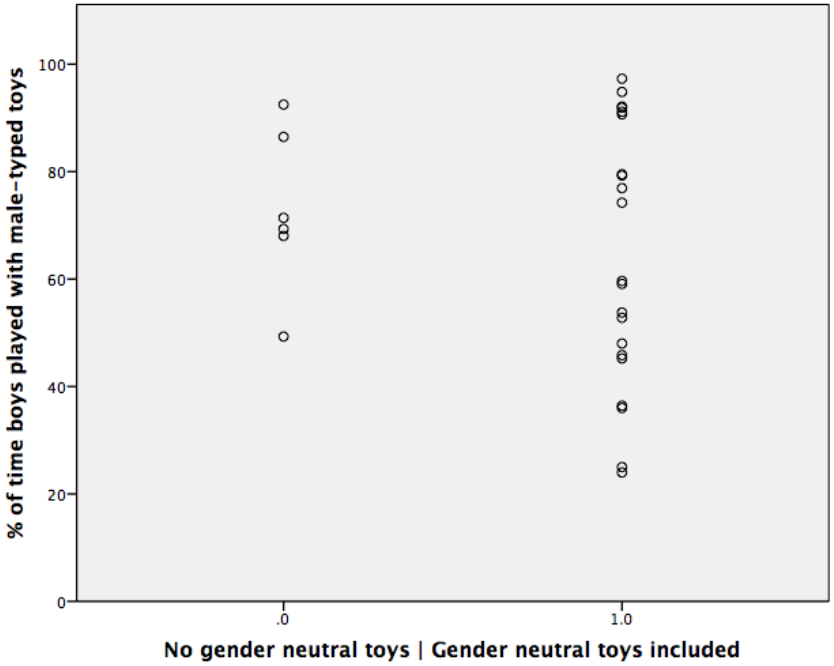


Figure 5a (left, boys) and **Figure 5b** (right, girls). Percentage of time played with 'male-typed' toys relative to 'female-typed' toys, when 'gender-neutral' toys were included in the stimuli.



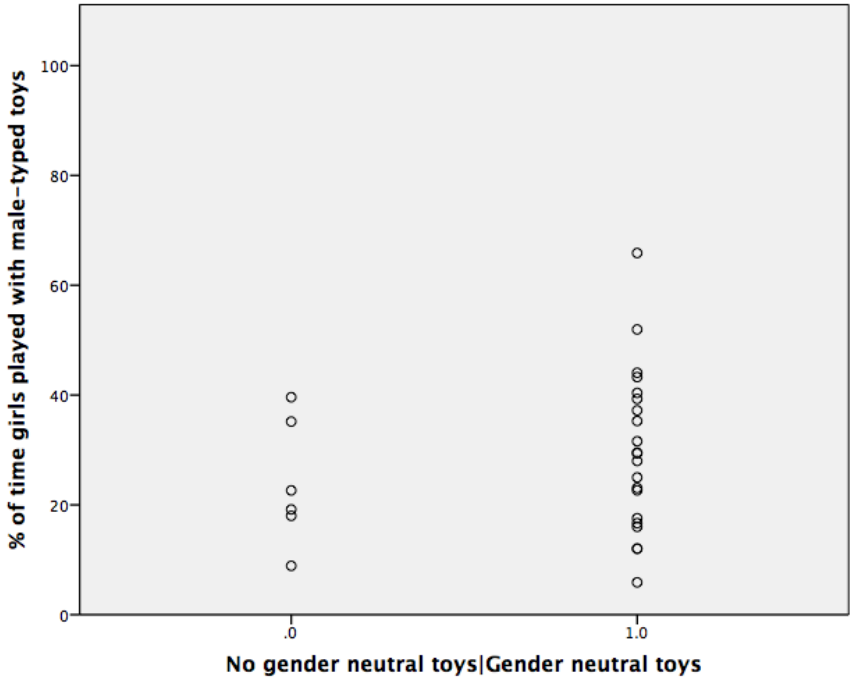


Figure 6a (left, boys) and **Figure 6b** (right, girls). Percentage of time played with 'male-typed' toys relative to 'female-typed' toys, in a home compared to a laboratory setting