PHILOSOPHICAL TRANSACTIONS B

royalsocietypublishing.org/journal/rstb

Research

**Cite this article:** House B, Silk JB, McAuliffe K. 2022 No strong evidence for universal gender differences in the development of cooperative behaviour across societies. *Phil. Trans. R. Soc. B* **378**: 20210439. https://doi.org/10.1098/rstb.2021.0439

Received: 28 February 2022
Accepted: 11 August 2022

One contribution of 17 to a theme issue 'Cooperation among women: evolutionary and cross-cultural perspectives'.

**Subject Areas:**
cognition, behaviour

**Keywords:**
gender differences, prosocial behaviour, fairness, child development, cross-cultural differences

**Author for correspondence:**
Bailey House
e-mail: bailey.house@york.ac.uk

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.6289742.

THE ROYAL SOCIETY PUBLISHING

# No strong evidence for universal gender differences in the development of cooperative behaviour across societies

Bailey House[1], Joan B. Silk[2] and Katherine McAuliffe[3]

[1]Department of Psychology, University of York, York YO10 5DD, UK
[2]School of Human Evolution and Social Change, Arizona State University, Phoenix, AZ, USA
[3]Department of Psychology and Neuroscience, Boston College, Chestnut Hill, MA 02467, USA

BH, 0000-0002-4023-9724

Human cooperation varies both across and within societies, and developmental studies can inform our understanding of the sources of both kinds of variation. One key candidate for explaining within-society variation in cooperative behaviour is gender, but we know little about whether gender differences in cooperation take root early in ontogeny or emerge similarly across diverse societies. Here, we explore two existing cross-cultural datasets of 4- to 15-year-old children's preferences for equality in experimental tasks measuring prosociality (14 societies) and fairness (seven societies), and we look for evidence of (i) widespread gender differences in the development of cooperation, and (ii) substantial societal variation in gender differences. This cross-cultural approach is crucial for revealing universal human gender differences in the development of cooperation, and it helps answer recent calls for greater cultural diversity in the study of human development. We find that gender has little impact on the development of prosociality and fairness within these datasets, and we do not find much evidence for substantial societal variation in gender differences. We discuss the implications of these findings for our knowledge about the nature and origin of gender differences in cooperation, and for future research attempting to study human development using diverse cultural samples.

This article is part of the theme issue 'Cooperation among women: evolutionary and cross-cultural perspectives'.

## 1. Introduction

Cooperation plays a fundamental role in human societies, but the form and scope of cooperation also vary considerably both across *and* within societies. Across societies, considerable variation in prosocial behaviour has been observed [1–3] and linked to variation in social norms that shape social preferences and regulate behaviour [4–6]. Within societies, gender is an important source of variation in behaviour, as division of labour between men and women is a universal feature of human societies, with each focusing on different types of subsistence tasks and sharing the products of their labour [7,8]. Developmental studies have informed our understanding of the sources of across-societal variation in prosocial behaviour, suggesting that the psychology underlying sensitivity to normative beliefs and values operates the same way across cultures, although the content of norms varies considerably [9]. By contrast, however, there has been relatively little consideration of whether there are gender differences in social preferences that underlie within-societal variation in prosocial behaviour (but see [10]) or whether there are systematic gender differences in the development of social preferences across societies. We consider the latter question in this paper.

### (a) Measuring prosocial social preferences

Experimental studies of prosocial behaviour provide insights about the psychological processes through which social preferences influence prosocial behaviour. In this work, researchers have borrowed methods developed by behavioural economists to assess the nature of preferences that underlie behaviour. Participants are presented with choices that have different outcomes on the material welfare of themselves and others, and the choices that they make reveal their preferences. By using different sets of choices, researchers can study distinct aspects of the psychological processes that govern cooperation in humans, such as generosity and fairness. For example, in one well-known experimental paradigm, the Dictator Game (DG), one individual (player 1) is given an endowment, usually a sum of money, that can be divided between themselves and another individual (player 2). Reputational concerns or expectations of reciprocity might influence the choices that individuals make, so offers are made anonymously and participants do not change roles. Any money allocated to player 2 reduces player 1's own pay-offs, so behaviour in this situation provides a measure of player 1's preference for outcomes that benefit others at a personal cost. However, there is no external factor compelling player 1 to give to player 2, so the DG can be interpreted as a measure of player 1's generosity. In the Ultimatum Game, a second stage is added to the standard DG: now, player 2 can accept or reject player 1's offer. If the offer is accepted, each takes home the amount that was offered. If the offer is rejected, both go home with nothing. The decision to reject any non-zero offer is costly and provides a measure of people's preferences for fair outcomes and aversion to inequity. Here, player 1 must take into account what player 2 will accept to prevent player 2 from rejecting their offer, but rejections by player 2 necessarily come second and so cannot change player 1's choices. The Ultimatum Game can therefore be interpreted as a measure of player 2's ideas about what is fair or unfair in this situation (in addition to a measure of player 1's belief about what player 2 thinks is fair).

These kinds of studies have been conducted in societies that span a wide range of subsistence strategies and levels of organization, from hunters and gatherers to western, educated, industrial, rich and democratic (WEIRD) societies. There is considerable diversity across societies in the behaviour observed in these games, suggesting considerable societal diversity in social preferences. For example, in some societies, people rarely reject any offers in the Ultimatum Game and in other societies people regularly reject offers that do not approach half of the original endowment [1–3]. Adults' choices in these kinds of economic games are influenced by what they themselves believe to be the correct behaviour and also by what *others* in their societies believe to be the correct behaviours [9]. The latter provides evidence of the importance of norms in shaping social preferences that underlie prosocial behaviour in games like the DG. Adults' choices are also correlated across games which measure different aspects of social preferences, such as the DG (which measures prosociality or preferences for fairness in dyadic interactions) and the Third-Party Punishment Game (which measures preferences for fairness in third-party interactions) [11]. These findings support the idea that prosocial behaviour in the DG is at least partly motivated by social preferences which are general across contexts.

It is important to note that a reasonable critique of games like the DG is that they often omit features of real-world social interactions which would make them more informative about real-world cooperative behaviour [12,13]. For example, they could (but often do not) ask participants to make decisions involving recipients who are from familiar ethnolinguistic groups or with whom they have real relationships [13,–16]. This means that one must be cautious about using behaviour in these games to draw inferences about real-world behaviour. However, this limitation of these games does not necessarily argue against evidence that individuals' choices in the games are motivated by general social preferences which vary across different social groups.

## (b) The development of prosocial social preferences

Developmental studies provide important insights about the origins of social preferences and the processes that shape them. One set of cross-cultural developmental studies was based on discrete versions of the DG [17]. In the Prosocial Game, child 1 is given a choice between two options. One option provides one reward to them and an identical reward for child 2 (1/1), the other option provides one reward to them and nothing to child 2 (1/0). In the Costly Sharing Game (CSG), child 1 is given a choice between one reward to them and an identical reward for child 2 (1/1), the other option provides two rewards to them and nothing to child 2 (2/0). Another set of developmental experiments was designed to specifically test children's aversion to unfair outcomes. In the Inequity Game (IG), child 1 is presented with a sequence of decisions about whether to accept or reject allocations of rewards for themselves and another child, child 2 [18,19]. The pay-off structure of these allocations is either equal (1/1), advantageous to child 1 (4/1) or disadvantageous to child 1 (1/4). If child 1 accepts the allocation, both children get the designated rewards, but if they reject the allocation neither child 1 nor child 2 receives anything. Although the IG is structurally like the DG because child 1 decides the pay-off outcome without input from child 2, it is functionally similar to player 2's role in the Ultimatum Game. This is because: (i) it involves a decision about whether to accept or reject a distribution chosen by someone else, and (ii) it reveals child 1's ideas about what is fair (i.e. whether the proposed division of rewards is better than an equal distribution of zero). The IG and CSG represent two different approaches to measuring social preferences in children, but they both require children to sacrifice personal gain to create more-equal outcomes: in the CSG, child 1 must incur costs to provide benefits to others and to avoid inequitable outcomes, and in the IG, child 1 must incur costs to avoid inequitable outcomes, but it is more costly for child 1 to avoid advantageous inequity (AI) than to avoid disadvantageous inequity (DI). We note that 'equitable' outcomes need not be 'equal' (and often are not), but in the design of these games equitable outcomes are always equal.

Several general findings emerge from this body of cross-cultural developmental work. First, developmental patterns are fairly uniform across cultures in some contexts and at some ages, but not others. For instance, behaviour in the Prosocial Game and responses to DI are relatively consistent across cultures. Children are increasingly likely to choose the 1/1 option over the 1/0 option in the Prosocial Game, and to reject the 1/4 option in the DI trials, as they mature from about 3 to 15 years of age. However, in situations in which children incur higher costs to create more-generous or more-fair outcomes, there is greater variation in behaviour across cultures. Second, cross-cultural variation emerges during middle childhood, the period of about 6–11 years of age. Children begin to reject offers in the AI trials when they are between about 7 and 10 years of age, but this occurs in only 3 of 7 societies in which the children were tested [18]. In the CSG, the probability of choosing the 2/0 option over the 1/1 option develops similarly across societies as children approached middle childhood and then begins to vary across societies as children converge on the behaviour of adults in their own societies in the same kinds of situations [9,17]. Third, additional experimental work which focuses on children's responses to novel normative instructions indicates that this is the period when children in very different societies become increasingly sensitive to normative instructions about correct behaviour [9,20]. This work suggests that the psychology underlying sensitivity to normative beliefs and values operates the same way across cultures, even though the content of norms varies considerably across those cultural groups.

## (c) Gender and prosociality

Local cultural norms can help to explain variation in the development of children's social preferences across societies, but this largely ignores possible sources of variation within societies, such as gender. A rich body of work demonstrates that cooperative decision-making is influenced by social categories such as ethnicity and group membership [21]. This is true in both adults and children. For instance, children share more with ingroup members than outgroup members in some experimental settings [22]. Gender is a key social category that is privileged in early ontogeny [23–25] and even at older ages continues to shape different aspects of social development including preferences [26] and behaviour [27]. Though adults' and children's prosocial behaviour does seem to differ based on the gender of a cooperative partner, there is less certainty about whether or how an individual's gender directly influences their prosocial behaviour. In particular, there is little consensus about whether gender differences in prosociality are universal features of human behaviour, or even what the direction of gender differences is (i.e. whether females are more prosocial than males or vice versa).

A number of theoretical proposals have offered explanations for why prosocial behaviour should vary based on gender, although there is no consensus about the direction of the effects [28]. Empirical work with adults does little to clarify this issue because the vast majority of data come from a small number of industrialized societies (see [10]) which has limited our ability to make compelling claims about universal features of human psychology [29]. Developmental data can potentially help shed light on the nature of gender differences in humans. If gender differences in prosocial preferences are a universal element of human behaviour, then we might expect them to develop in a uniform way across cultures, to emerge early in development, and be predictive of gender differences in adulthood. Alternatively, if gender differences are the product of cultural forces that operate within societies, we might expect them to emerge during middle childhood as children become sensitive to local social norms and begin to develop adult-like patterns of behaviour [30]. Meta-analyses of the influence of gender on prosocial behaviour in adults and children suggest that adult males are generally more prosocial than adult females [31], but some reviews have found this pattern to be inconsistent [32], while other reviews have obtained the reverse pattern for children [33]. For example, a recent study showed that among two populations of Americans and Canadians, girls shared stickers in a more egalitarian manner than did boys [34]. However, the effects of gender among children may vary for different types of prosocial behaviour, such as instrumental helping and sharing [33], and findings about gender differences may depend greatly on the context in which this behaviour is being measured [35,36]. These findings illustrate the potential importance of developmental research for fully characterizing gender differences in human cooperation, but they also illustrate the importance of standardizing tasks to facilitate comparisons between children and adults. Additionally, prior studies of gender differences in children have focused on WEIRD populations [29], a persistent problem in developmental psychology [37].

## (d) Present research

Here, we examine the impact of gender on the development of prosocial preferences across a diverse set of societies. We use measures of prosocial behaviour that are similar across societal samples, and that are comparable to measures used with adults. If gender differences reflect a phylogenetic legacy of evolutionary pressures that favour different behavioural strategies among males and females, we might expect gender differences in prosocial preferences to emerge early in life and be similar across cultures (prediction 1). Cultural evolutionary theories do not make specific claims about gender differences in social preferences. However, it is possible that gender differences in status, social roles and scope of social networks might be associated with gender differences in prosocial preferences. If this is the case, we might expect gender differences in prosocial behaviour to vary across societies and to begin to diverge during middle childhood (prediction 2).

To determine whether there are gender differences in the development of prosocial behaviour and fairness, we need data on children's prosocial behaviour and fairness behaviour. To this end, we draw on data which were collected using experimental methods applied consistently across different age groups and in diverse societies. These data come from cross-cultural studies using the binary-choice DG [9,17] that provide information about the behaviour of 663 children aged 4–14 years across 14 different societies. Data on the development of fairness behaviour are provided by a cross-cultural study using the IG [18] that provides data from about 430 children (429–437 depending on the condition) aged 4–15 years old across seven different societies. The IG protocol explores the development of two distinct forms of fairness: an aversion to outcomes which advantage oneself (AI) and an aversion to outcomes which disadvantage oneself (DI). In the studies below, we explore gender differences in the development of prosocial behaviour in the DG (study 1), fairness behaviour in the Advantageous Inequity Game (AIG) (study 2a) and fairness behaviour in the Disadvantageous Inequity Game (DIG) (study 2b). In each of these studies, we address two research questions which test predictions 1 and 2:

(i) prediction 1: are there gender differences in the development of children's behaviour which are widespread across societies? Our culturally diverse dataset allows

**Table 1.** Populations sampled for the DG for study 1.

| | population (location); description | n = SS (female) | n = obs. |
|---|---|---|---|
| **dataset DG1** | Aka (Congo Basin, Central African Republic); rural, foraging, hunting | 35 (13) | 35 |
| | Fijian (Yasawa Island, Fiji); rural, marine foraging, small-scale horticulture | 75 (33) | 75 |
| | Himba (Omuhonga Basin, Namibia); rural, pastoralism, small-scale horticulture | 82 (48) | 82 |
| | American (Los Angeles, United States); urban | 72 (35) | 72 |
| | Martu (Western Desert, Australia); rural, foraging, hunting | 22 (10) | 22 |
| | Shuar (Amazonia, Ecuador); rural, small-scale horticulture, hunting | 37 (13) | 37 |
| **dataset DG2** | German (Berlin, Germany); urban | 33 (18) | 33 |
| | Hadza (Great Rift Valley, Tanzania); rural, foraging, hunting | 26 (9) | 26 |
| | Argentinian (La Plata, Argentina); urban | 46 (21) | 46 |
| | American (Phoenix, United States); urban | 58 (28) | 58 |
| | Indian (Pune, India); urban | 50 (25) | 50 |
| | Shuar (Amazonia, Ecuador); rural, small-scale horticulture, hunting | 17 (8) | 17 |
| | Tanna (Tafea province, Vanuatu); rural, small-scale horticulture, hunting | 81 (43) | 81 |
| | Wichí (Misión Chaqueña, Argentina); rural, sedentarized hunter–gatherers | 29 (15) | 29 |
| | total | 663 (319) | 663 |

us to draw more plausible conclusions about universal human gender differences than is possible with samples from only one or a few societies, and permits a good opportunity to test prediction 1; and

(ii) prediction 2: is there cultural variation in how gender shapes the development of prosocial behaviour? Although this is a question of central importance, samples from some of the societies included in the analyses are small, and this limits the strength of conclusions that we can draw about gender differences in particular populations. We address this question and its test of prediction 2 with this limitation of our dataset in mind.

## 2. Study 1: Dictator Game

### (a) Methods

#### (i) Participants and experimental procedure

Data on children's choices in the binary-choice DG were obtained by combining datasets from two different cross-cultural research projects [9,17]. These two studies used similar versions of the DG, and the subjects in both studies spanned a similar age range. However, the tasks were presented to the children in somewhat different ways which we describe below.

#### (ii) Dataset: Dictator Game 1 (DG1)

This project examined the behaviour of children aged 3–14 years at six different sites ([17]; table 1; see original paper for full description of the sites and sample sizes). Children were presented with two training trials followed by four experimental trials (the order of which was counterbalanced across participants). In the experimental trial relevant to the present study (CSG, 'Social' condition in the original study), one child (child 1) was assigned to the role of actor and presented with two options: (i) they could keep two rewards and give none to a peer recipient (2/0, the self-maximizing option), or (ii) they could keep one reward and give one to the recipient (1/1, the prosocial option). The actor was face-to-face with the familiar peer recipient (child 2) and decided between the two options while being observed by both recipient and the experimenter. Participants were recruited at schools or opportunistically in compounds/villages, and children were randomly paired with partners with a few constraints. Efforts were made to pair children with peers of similar ages and to avoid pairing them with immediate family members, but no other sampling criteria (such as gender or familiarity) were used. Data from three additional trials using other versions of this experimental task are not included in the present study because the tasks differ considerably from the one used to produce dataset DG2.

#### (iii) Dataset: Dictator Game 2 (DG2)

This project collected data with samples of children aged 4–17 years at eight different sites ([9]; table 1; see original paper for full description of the sites and sample sizes). For the purposes of comparability across datasets, we have limited the analyses in this paper to participants between the ages of 4 and 14 years (excluding six participants older than 14 years). The procedure consisted of several training trials followed by the experimental trial relevant to the present study. In this trial, one child (child 1) was assigned to the role of actor and presented with the same two options as in dataset DG1: (i) they could keep two rewards and give none to a peer recipient (the 2/0, self-maximizing option), or (ii) they could keep one reward and give one to the recipient (1/1, the prosocial option). In this project, the actor was told that the recipient (child 2) was another child in the community, but the recipient was not physically present and anonymous. When participants decided between the two options, they were observed only by the experimenters. We want to highlight this significant methodological difference, in which there was no anonymity for participants in the experiment which produced dataset DG1, but considerable anonymity in the experiment which produced dataset DG2. In continuous versions of the DG, adults are more prosocial where they are less anonymous

[38]. Males and females are generally influenced by anonymity to a similar degree, but there are some gender differences in the distribution of their prosocial choices [39]. For this reason, we consider whether datasets DG1 and DG2 present different patterns of gender differences in children's prosocial behaviour.

Before participants made their choice, children viewed a short video in which an adult verbalized novel information about the two options. Before trials in the condition relevant to the present study (the 'both OK' condition in the original study), the model stated that 1/1 and 2/0 were both 'OK' and 'OK to choose'. This language is not strongly normative and does not preferentially bias children towards choosing either option and is very similar to language used to introduce the experimental task in both dataset DG1 and dataset DG2 (i.e. 'you can choose whichever one you want'). This video manipulation was intended as a baseline condition in the original study [9] and was contrasted with other between-subjects conditions which presented children with videos containing normative information intended to bias children's choices. Data from these other conditions are not included in the present study because these videos bias the children's choices in a way that makes the experimental task not comparable to that used in dataset DG1. Videos used a standardized script but were recorded at each field site using local translations of the script and local adults as models.

### (iv) Coding and variables

In some cases, the same society was represented in both datasets (e.g. the Shuar and the USA (Los Angeles and Phoenix)), but these observations were obtained from different communities (sites) within these societies, and the data were collected 7–8 years apart. For this reason, when combining the data from both datasets, we assign each of the 14 samples a unique multi-level population code.

The data from the DG is binary (a participant chose either the '1/1' prosocial option [1] or the '2/0' self-maximizing option [0]). We predicted this binary dependent variable (DV) using variables capturing participants' gender (gender), age (age and age$^2$) and site.

*Gender*: this parameter is binary (female [1] and male [0]) and was coded either from consent forms provided by parents or by researchers at the time of the experimental sessions.

*Age* and *age*$^2$: we centred (i.e. *z*-score transformed) children's age by subtracting the full sample mean age (8.31 years) and dividing by the standard deviation (2.65). We include both this centred age parameter (age) and a squared version of the same parameter (age$^2$) to allow us to model non-monotonic age functions.

*Site:* the models also include multi-level parameters, including random intercepts and random slopes for all variables and interactions, for each site sample.

### (v) Statistical analyses

We constructed a Bayesian regression model that includes terms for gender, age and age$^2$. We also included interactions between gender and age/age$^2$ to evaluate whether the development of prosocial behaviour differs across genders. To assess and control for societal variation, we included multi-level parameters for the interactions for each site sample. We present the results of this model graphically because its complexity makes full numerical results difficult to interpret. Basic numerical model results are presented in table 2 (full numerical model results can be found in the electronic supplementary material).

We modelled participants' choices using regression with a binomial link function. The posterior distribution of the model was estimated using Markov chain Monte Carlo, in which model predictions were generated by processing many samples from the posterior distribution of the model. Data were analysed in the R Environment for Statistical Computing v. 3.5.1 [40], with models specified using the function 'map2stan' (R package 'rethinking' v. 1.59), a convenience tool for fitting different regression models [41]. Multi-level models were run using a variant of Hamiltonian Monte Carlo (an algorithm good with high-dimension models) implemented in RStan v. 2.18.2 [42]. Models were specified using weakly informative priors, which reduce overfitting and help the Markov chain to converge to the posterior distribution more effectively than flat priors. These were the same priors as were used in previously published analyses based on datasets DG1 and DG2 [9,17]. The posterior distribution we present here is based on 12 000 samples from three chains (after 3000 adaptation steps), for a total of 27 000 samples. These samples were sufficient to establish convergence to the target posterior distribution. We assessed convergence through: (i) visual inspection of the chains, (ii) the R-hat Gelman and Rubin statistic (approximately 1.00 for all parameters, R-hat values greater than 1.01 can indicate that the chain did not converge), and (iii) the effective number of samples for all parameters were reasonable. Statistics for convergence of the Markov chains can be found in the full numerical model results provided in the electronic supplementary material.

## (b) Results

### (i) Are there gender differences in the development of prosocial preferences?

Across the entire cross-cultural sample, males chose the prosocial option 34% of the time on average and females chose the prosocial option 27% of the time. Figure 1*a* shows the model estimates for the probability that males and females will choose the prosocial option in the DG. These estimates are largely similar across genders, but the model estimates for males are somewhat higher than the estimates for females between about age 7 and 10 years of age. However, the 95% percentile intervals for the estimates for males and females in this age range overlap considerably, suggesting that the gender difference is not very reliable. In figure 1*b* we plot the coefficient for gender as a function of child age (dotted black line). The 95% percentile interval of the coefficient for gender does not fully exclude zero at any age, suggesting that gender does not consistently affect the likelihood of choosing the prosocial option.

The inclusion of multi-level parameters for each of the 14 different site samples gives us confidence that the estimate of the coefficient for gender is not likely to be strongly biased by individual sites. To confirm this, we have also plotted coefficients for the multi-level parameters for gender for each site sample (figure 1*b*). Estimates for the coefficient of gender in each site are largely similar to the overall effect of gender, particularly through the range of 7–10 years where we have the largest samples of participants. The figure also shows that there is not any systematic difference in the parameters for

**Table 2.** Model structures and numerical model results for studies 1, 2a and 2b. (Model structures are presented using the formulae used for similar regression models in the R package 'lme4'. Differences in the structures of the models across studies are in italics. Full model results (including results for multi-level parameters) are provided in the electronic supplementary material.)

| | study 1 | | | | study 2a | | | | study 2b | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| parameter name | coef. | s.d. | 95% CI% | | coef. | s.d. | 95% CI% | | coef. | s.d. | 95% CI% | |
| intercept | −0.87 | 0.19 | −1.25 | −0.51 | −2.02 | 0.66 | −3.12 | −0.49 | −0.07 | 0.37 | −0.84 | 0.67 |
| gender (1 = female) | −0.45 | 0.26 | −0.96 | 0.06 | 0.02 | 0.40 | −0.78 | 0.81 | 0.00 | 0.36 | −0.72 | 0.71 |
| age (centred) | −0.04 | 0.22 | −0.47 | 0.39 | 0.01 | 0.35 | −0.70 | 0.71 | 0.71 | 0.37 | −0.08 | 1.40 |
| $age^2$ (centred) | 0.21 | 0.16 | −0.11 | 0.51 | 0.26 | 0.28 | −0.35 | 0.78 | −0.37 | 0.23 | −0.85 | 0.06 |
| trial number | | | | | −0.07 | 0.11 | −0.29 | 0.13 | 0.03 | 0.08 | −0.15 | 0.18 |
| gender × age (interaction) | −0.06 | 0.23 | −0.49 | 0.41 | 0.36 | 0.43 | −0.49 | 1.23 | −0.23 | 0.36 | −0.91 | 0.51 |
| gender × $age^2$ (interaction) | 0.12 | 0.21 | −0.30 | 0.54 | −0.14 | 0.38 | −0.94 | 0.56 | 0.13 | 0.28 | −0.38 | 0.73 |
| regression model structures | | | | | | | | | | | | |
| study 1 | [outcome variable] *'subject chose 1/1 in DG'*∼ [fixed effects] 1 + gender + age + $age^2$ + gender*(age + $age^2$) + [random effects] (1 + gender + age + $age^2$ + gender*(age + $age^2$) \| site ID) | | | | | | | | | | | |
| study 2a | [outcome variable] *'subject rejected 4/1 in Advantageous Inequity Game'*∼ [fixed effects] 1 + gender + age + $age^2$ + *trial number* + gender*(age + $age^2$) + [random effects] (1 + gender + age + $age^2$ + *trial number* + gender*(age + $age^2$) \| site ID) | | | | | | | | | | | |
| study 2b | [outcome variable] *'subject rejected 1/4 in Disadvantageous Inequity Game'*∼ [fixed effects] 1 + gender + age + $age^2$ + *trial number* + gender* (age + $age^2$) + [random effects] (1 + gender + age + $age^2$ + *trial number* + gender*(age + $age^2$) \| site ID) | | | | | | | | | | | |

gender across the sites in dataset DG1 (pink lines) and dataset DG2 (green lines). This indicates that despite the methodological differences in the anonymity of participant children across the two studies, the influence of gender on prosocial behaviour was similar.

#### (ii) Do gender effects vary across sites?

We can explore societal variation by plotting the coefficients for the multi-level parameters for gender for each site and comparing the estimated coefficients for gender across all sites (figure 2). We note at the outset that sample sizes for some of the sites are very small, and the estimates for these sites should be viewed with caution. The only apparent difference in the magnitude of the estimated coefficient for gender is between Los Angeles and the Martu when children are 3–4 years of age. However, this difference should be interpreted with great caution because our samples for this age are small at all sites, and the sample for the Martu is particularly limited. Thus, our data provide no compelling evidence for societal variation in the development of children's choices in the DG.

### (c) Study 1: discussion

There is very limited evidence for gender differences in the development of prosocial behaviour in the binary-choice version of the DG used here. Males are estimated to be somewhat more likely to select prosocial outcomes than are females, but overall this difference is relatively small (about 7–8%) and not very reliable even at 7–10 years of age where we have the most data. There is also very little evidence for societal variation in gender differences in the DG.

Overall, these analyses suggest that there could be very modest gender differences in prosocial behaviour in 7- to 10-year olds, but the effect is quite uncertain. A gender difference in the pattern of responses in the DG could indicate that there are gender differences in either generosity, or preferences for fair outcomes. To help distinguish between these possible explanations, we can examine children's responses in a task that directly measures responses to unfairness. To this end, in Study 2, we investigate gender differences in the development of children's aversion to AI.

## 3. Study 2: The Inequity Game

The IG assesses preferences for equitable outcomes. In this game, two children are paired in a face-to-face game. One child, child 1, is assigned to the role of actor, and this participant makes all the decisions in the game, and the other child, child 2, is assigned to the role of recipient. Experimenters present pairs with different allocations of treats (candies, cookies, etc.), half of which are equal and half of which are unequal. Two forms of inequity are tested. In the AI condition, the actor receives more rewards than the recipient (four treats for actor, one for recipient; 4/1). In the DI condition, the actor receives fewer rewards than the recipient (one for actor, four for recipient; 1/4). The actor can either accept or reject allocations using green or red handles, respectively, that control whether the treats are (i) delivered to both children, or (ii) to a middle bowl in which case neither child receives them.

There are similarities and differences between the DG and the IG. In both games, the equitable allocation is the least

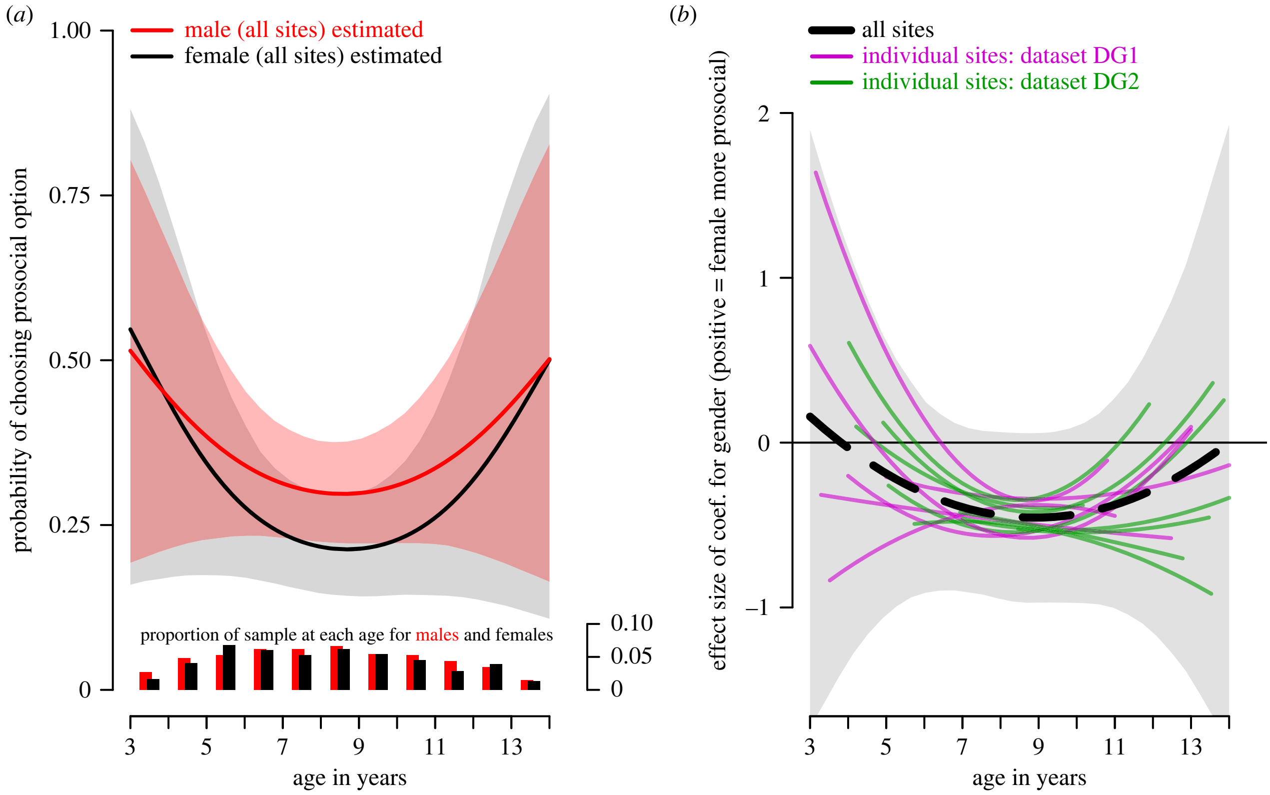


**Figure 1.** Results for the analysis of gender differences in the DG. (*a*) Lines represent regression estimates for the probability that females (black) and males (red) will choose 1/1, for the entire cross-cultural sample. Shaded regions denote 95% percentile intervals for those estimates (males = red, females = grey). Vertical bars along the *x*-axis represent the proportion of the total sample at each age in the distribution, for both males (red bars) and females (black bars). (*b*) Lines represent regression estimates for the coefficient of the gender parameter, for the entire cross-cultural sample (black dashed), for the individual sites in dataset DG1 (pink lines), and for the individual sites in dataset DG2 (green lines). Estimates for each site are presented for the age range of children in each site sample. Shaded region denotes 95% percentile intervals for the estimates for the entire cross-cultural sample (black dashed line). (Online version in colour.)

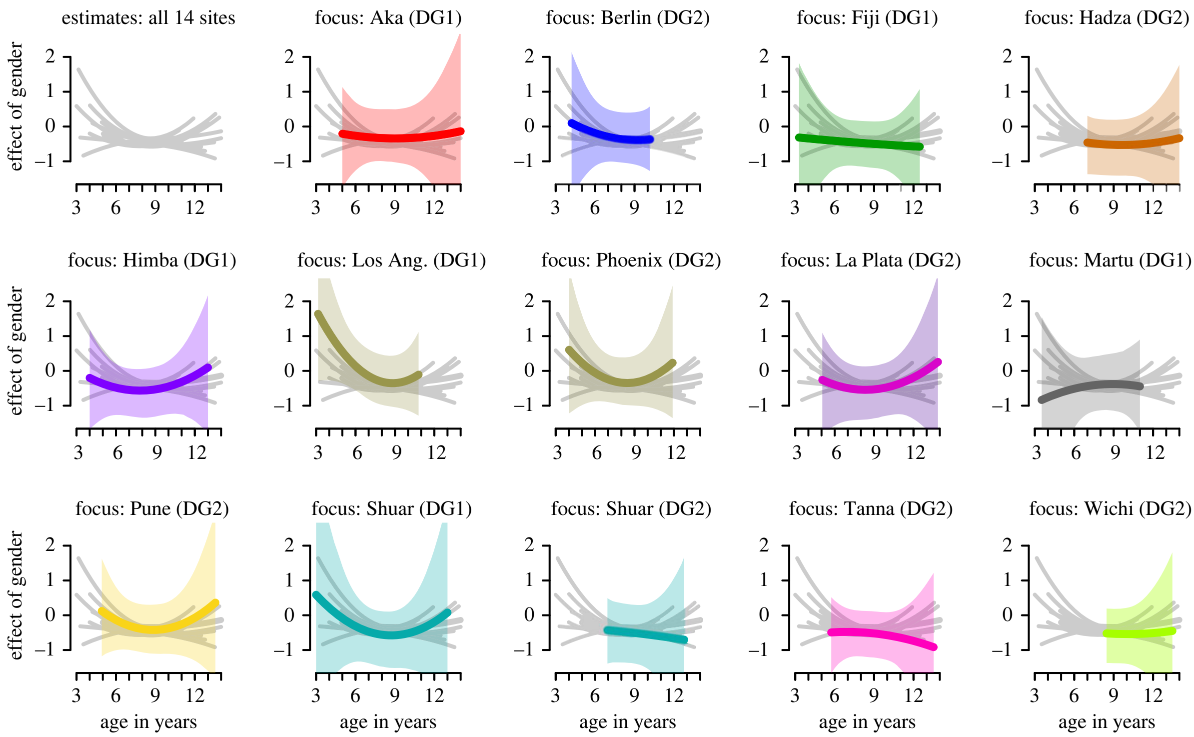


**Figure 2.** Results for the analysis of societal variation in gender differences in the DG. Top left panel: grey lines represent regression estimates for the coefficient of the gender parameter, for each of the 14 sites. All other panels: coloured lines and shaded regions represent regression estimates (and 95% percentile intervals) for the coefficient of the gender parameter for one of the sites, compared to the grey lines representing regression estimates for the coefficients for gender for each of the other sites. This allows comparisons to determine whether the estimates for each site differ from the estimates for the other sites. Estimates and percentile intervals for each site are presented for the age range of the children in each site sample. (Online version in colour.)

rational choice from a self-maximization perspective because the actor must make a personal sacrifice to generate equitable outcomes. In the DG, preferences for equitable and generous outcomes are confounded because the 1/1 option is both equitable and beneficial to the recipient. In both versions of the IG, the decision to reject an offer is not a generous choice, because children who reject an inequitable allocation deprive their partner (and themselves) of rewards. However, rejections always reduce inequality between the actor and recipient. Rejections in the IG thus provide a measure of preferences for equitable outcomes (fairness) that is not confounded with generosity.

Considering children's responses in the IG allow us to explore whether the small and uncertain gender difference observed in the DG is more likely to be driven by a gender difference in fairness, or a gender difference in generosity or rationality. Specifically, if we observe a similar (or stronger) gender difference in the AIG as we observed in the DG, then we might propose that this is driven by a gender difference in fairness motivations.

## 4. Study 2a: Advantageous Inequity Game

### (a) Methods

#### (i) Participants and experimental procedure

Data on children's choices in the binary-choice IG were taken from [18] which explored inequity aversion in children ages 4–15 across seven societies. In a between-participant design, children were presented with either AI (study 2a) or DI (study 2b). In all cases, participants sat face-to-face with a gender-matched peer. Participants were recruited at schools, through community contacts, or opportunistically in public areas like parks. Children were paired with same-gender peers who were as closely matched in age as possible. In some sites, it was possible to pair two children who did not know one another (e.g. USA), while in others this was virtually impossible owing to small community size (e.g. Mexico; see Blake *et al*. [18]) for more details regarding sampling). As the proportion of unfamiliar peers varies from 0.80 to 0.01 across sites (Blake *et al*. [18]; electronic supplementary material, table S2), thorough analysis of the effects of partner familiarity on children's behaviour is not possible with this data.

In the task, one participant, the actor, made decisions by manipulating a wooden apparatus which allowed them to either accept or reject allocations of resources by pulling one of two handles. The procedure for this task was largely the same across all sites in that within each session participants were presented with multiple trials, half of which were unequal trials (in which accepting the allocations would give participants more rewards than a peer) and half of which were equal trials (in which accepting the allocations would give participants the same number of rewards as a peer). In this study, we only analyse the unequal trials. Aspects of the procedure that varied across sites are discussed in detail in Blake *et al*. [18] and included: (i) differences in food rewards used across sites; (ii) differences in the blocking of trials, with children in some sites being presented with counterbalanced blocks of 12 equal or unequal trials, while children in other sites were presented with 16 randomized equal or unequal trials; (iii) across some sites, the resources presented were variable within experimental session (half were more-preferred items, half less-preferred) while in other sites resources were stable within-session; and (iv) the proportion of unfamiliar versus familiar pairs of children varied depending on location.

#### (ii) Coding and variables

The outcome data for the AIG are binary (i.e. choosing to reject [1] or accept [0] an option). Note that here and in study 2b, we examine only responses to unequal options (see [18]). We then predicted this binary DV using variables capturing participants' gender (gender) and age (age and age$^2$), similar to study 1 (age was centred using a mean of 8.47 [years] and a s.d. of 2.85). We also included a centred covariate parameter capturing the order of presentation of a given trial (trial number; mean of 7.86, s.d. of 4.41), as this was included in the models for the original study. We also include multi-level parameters which included: (i) random intercepts and random slopes (gender, age, age$^2$, trial number) for each site sample (i.e. population of origin), and (ii) random intercepts for individual participants.

#### (iii) Statistical approach

We use an identical modelling approach as that described for study 1. The only difference is that the regression models also include random intercepts for participant identity (ID) and a covariate for trial number (along with random slopes for this for each site).

### (b) Results

#### (i) Are there gender differences in preference for fair outcomes?

We plot the model's predictions for the probability that males and females reject an inequitable distribution in the AIG (figure 1*a*). These estimates are very similar for males and females, and there is no evidence of a general gender difference across all of the sites at any age.

The coefficient for the overall effect of gender across sites (figure 2) does not reliably deviate from zero at any age, indicating that there is no clear effect of gender at any age. The estimates for most of the sites appear largely similar to the overall pattern across sites (figure 2, right panel, pink lines).

#### (ii) Do gender effects vary across sites?

As in study 1, we can explore societal variation by plotting the coefficients for the multi-level parameters for gender for each site with 95% percentile intervals, and comparing the estimated coefficients for gender at each site to all of the other sites (figure 3). We note that this dataset provides more data for each site than was available for study 1 (table 3), particularly at the youngest ages (figure 4*a*, vertical bars next to *x*-axis), so estimates for individual sites can be viewed with greater confidence. The clearest evidence for societal variation is that the estimate for gender in Canada is reliably more negative than the estimate for Uganda, though only for children younger than age 5–6 years. Given that the majority of the data comes from younger children, this finding is potentially meaningful.

### (c) Study 2a discussion

This study's primary finding is that there is little evidence for cross-cultural gender differences in AI aversion, suggesting that males and females are relatively similar in their tendency

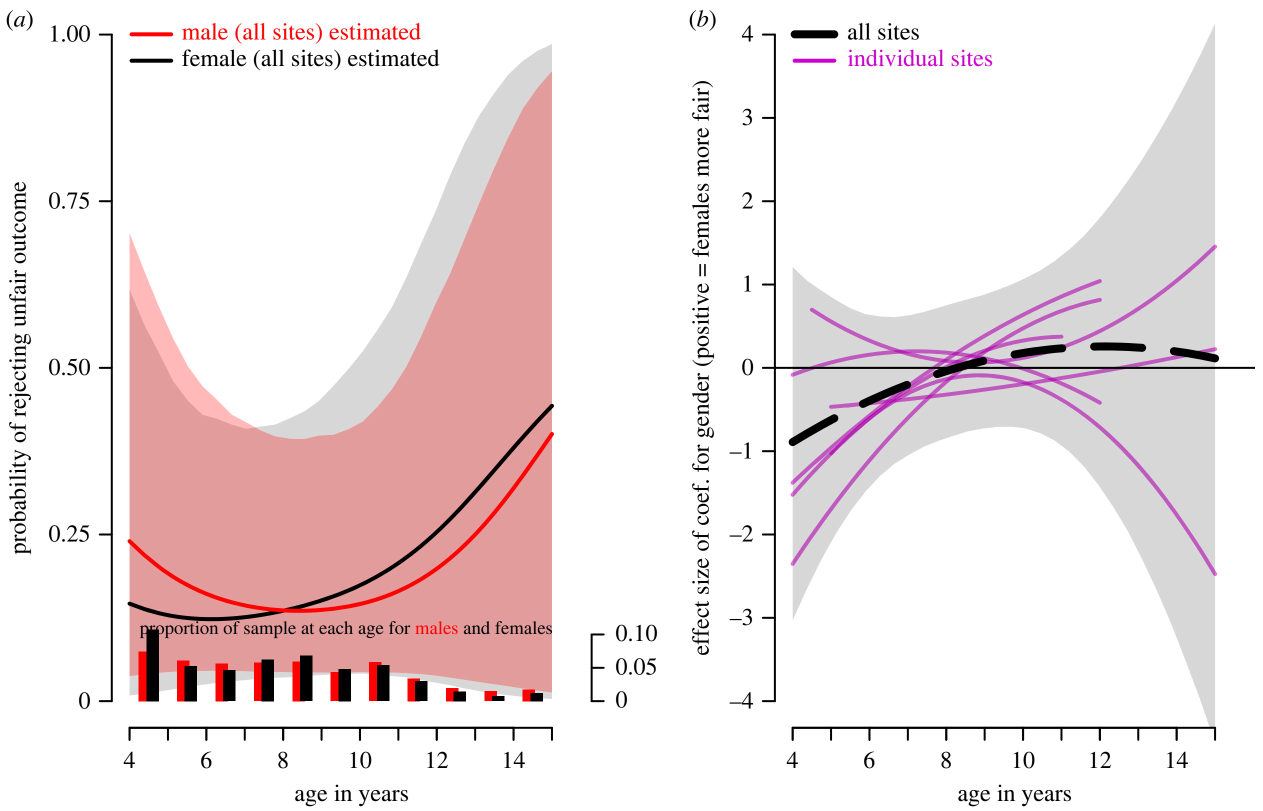


**Figure 3.** Results for the analysis of societal variation in gender differences in the AIG. (*a*) Lines represent regression estimates for the probability that females (black) and males (red) will reject the unfair distribution of rewards, for the entire cross-cultural sample. Shaded regions denote 95% percentile intervals for those estimates (males=red, females=grey). Vertical bars along the *x*-axis represent the proportion of the total sample at each age in the distribution, for both males (red bars) and females (black bars). (*b*) Lines represent regression estimates for the coefficient of the gender parameter, for the entire cross-cultural sample (black dashed line), and for the individual sites (pink lines). Estimates for each site are presented for the age range of the children in each site sample. Shaded region denotes 95% percentile intervals for the estimates for the entire cross-cultural sample (black dashed line). (Online version in colour.)

**Table 3.** Populations sampled for the AIG (study 2a) and Disadvantageous Inequity Game (study 2b).

| | population (location); economy | *n* = SS (female) | *n* = obs. |
|---|---|---|---|
| study 2a: **AIG** | Antigonish (Canada); professional, trade/service, agriculture | 47 (19) | 377 |
| | Andhra Pradesh (India); agriculture, labour | 57 (30) | 342 |
| | Xculoc (Puuc region, Mexico); agriculture, labour | 37 (15) | 297 |
| | San Pedro de Saño (Peru); agriculture, labour | 75 (36) | 444 |
| | Dakar (Senegal); trade/service, labour, fishing | 63 (31) | 504 |
| | Fort Portal (Uganda); agriculture, labour | 59 (29) | 472 |
| | Boston (United States); professional, trade/service, labour | 99 (60) | 792 |
| | total | 437 (220) | 3228 |
| study 2b: **DIG** | Antigonish (Canada); professional, trade/service, agriculture | 49 (28) | 393 |
| | Andhra Pradesh (India); agriculture, labour | 47 (30) | 282 |
| | Xculoc (Puuc region, Mexico); agriculture, labour | 31 (14) | 248 |
| | San Pedro de Saño (Peru); agriculture, labour | 74 (37) | 439 |
| | Dakar (Senegal); trade/service, labour, fishing | 65 (33) | 517 |
| | Fort Portal (Uganda); agriculture, labour | 55 (28) | 438 |
| | Boston (United States); professional, trade/service, labour | 108 (51) | 863 |
| | total | 429 (221) | 3180 |

to reject unfair outcomes that advantage themselves across these seven sites. There may be some societal variation in gender differences in AI aversion, but this effect is limited to children in two societies that are younger than age 5–6 years.

Overall, these results do not support the idea that males and females show pronounced differences in fairness motivations; specifically, in their tendency to give up self-advantageous pay-offs so as to increase fairness between themselves and a peer. However, even if there are no gender differences in the

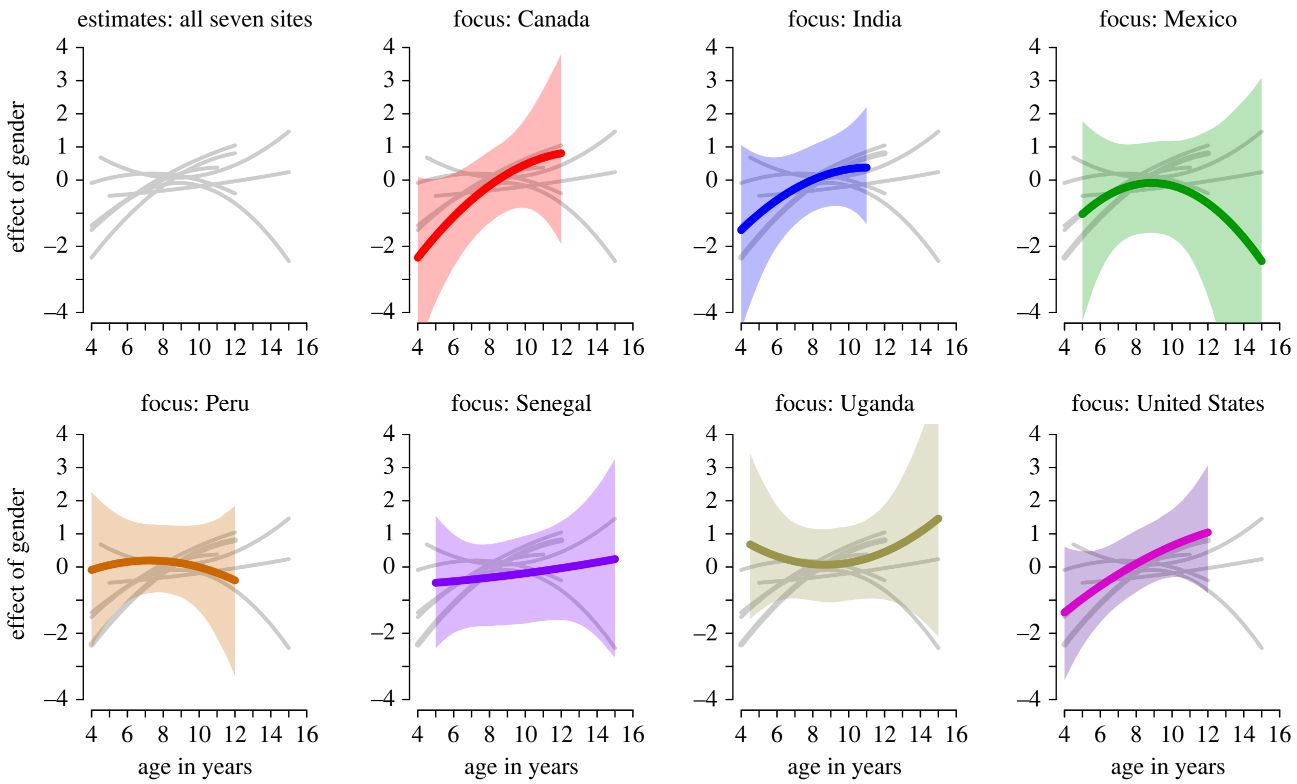


**Figure 4.** Results for the analysis of societal variation in gender differences in the AIG. Top left panel: grey lines represent regression estimates for the coefficient of the gender parameter, for each of the seven sites. All other panels: coloured lines and shaded regions represent regression estimates (and 95% percentile intervals) for the coefficient of the gender parameter for one of the sites, compared to the grey lines representing regression estimates for the coefficients for gender for each of the other sites. This allows comparisons to determine whether the estimates for each site differ from the estimates for the other sites. Estimates and percentile intervals for each site are presented for the age range of the children in each site sample. (Online version in colour.)

development of AI aversion, gender differences in other forms of fairness may still be possible. In study 2b, we explore gender differences in the development of a related form of fairness known as 'disadvantageous inequity aversion', in which an individual is averse to unfair outcomes which put themselves at a personal disadvantage relative to others.

DI aversion can be measured in a similar way to AI aversion, using what we will call the DIG [19]. In this game, children again accept or reject unfair allocations of rewards between themselves and a peer, but the reward distribution is reversed: one reward for the child themselves and four rewards for the peer. If children reject this distribution then no one gets anything, but if they accept the distribution then they get one and the peer gets four, putting themselves at a disadvantage relative to the peer.

As with the AIG, the only way to reduce unfairness in the DIG is to reject the unfair distribution, but rejection does not require giving up a personal advantage. For this reason, the DIG is less similar to the DG (relative to the AIG). However, reducing unfairness in the DIG still requires paying a personal cost (i.e. giving up rewards), which is also true in both the DG and AIG.

## 5. Study 2b: Disadvantageous Inequity Game

### (a) Methods

#### (i) Participants and experimental procedure

As with study 2a, data on children's choices in the binary-choice DIG were taken from Blake *et al.* [18] which explored inequity aversion in children ages 4–15 years across seven societies. Different children participated in studies 2b and 2a, meaning that children were assigned to be presented with either advantageous or disadvantageous allocations, but children were never presented with both forms of inequality. However, participants in both studies were from the same populations and data were collected at the same time. The experimental procedure was identical to that described in study 2a, the only difference between the DIG and the AIG is the pay-off distribution that children were presented with: one for themselves and four for the recipient (the inverse of the pay-offs in the AIG).

#### (ii) Coding, variables and statistical approach

For study 2b, age was centred using a mean of 8.45 (years) and a s.d. of 2.76, and trial number with a mean of 7.88 and s.d. of 4.43. Our approach was identical to studies 1 and 2a, except that in the model analyses, the Markov chains were less efficient and required more samples to converge to the posterior distribution. The posterior distribution we present is based on 20 000 samples from three chains (after 6000 adaptation steps), for a total of 36 000 samples. These samples were sufficient to establish convergence to the target posterior distribution, though the chains were inefficient.

### (b) Results

#### (i) Are there gender differences in preference for fair outcomes?

As with the DG and AIG, we plot the model's predictions for the probability that males and females will reject an unfair distribution in the DIG (figure 5*a*). These estimates are similar across males and females, and there is again no evidence of a

royalsocietypublishing.org/journal/rstb *Phil. Trans. R. Soc. B* **378**: 20210439

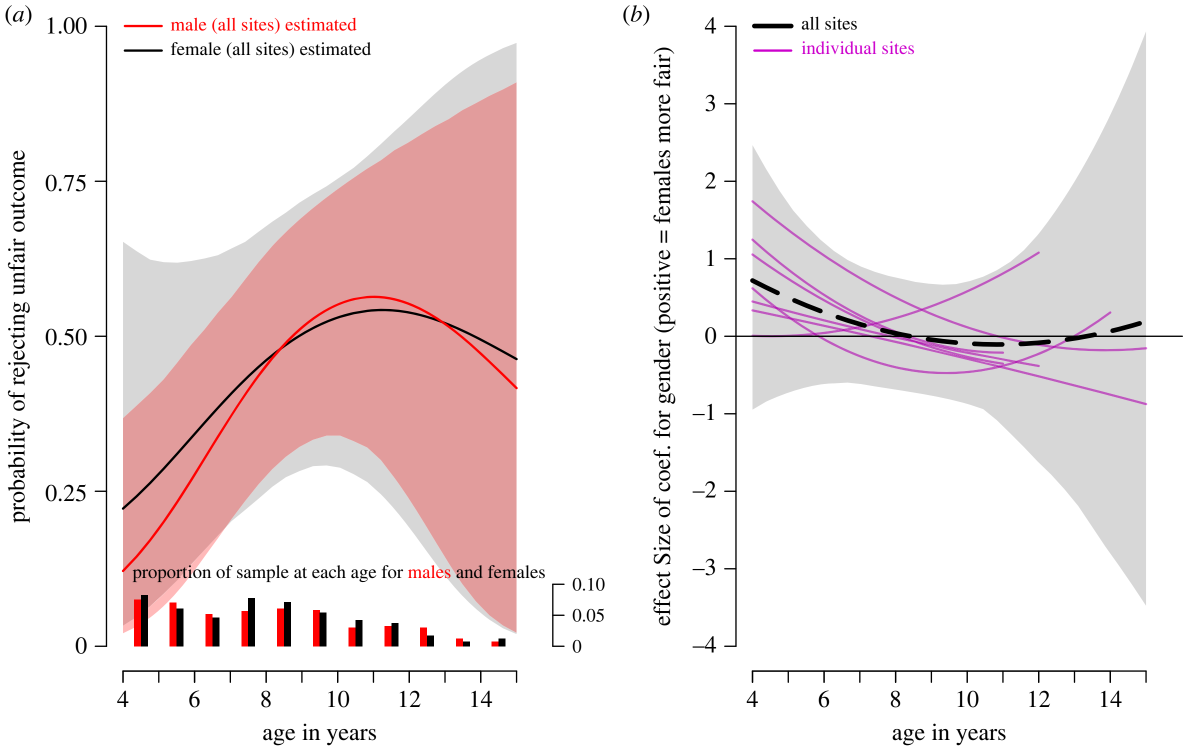


**Figure 5.** Results for the analysis of gender differences in the DIG. (*a*) Lines represent regression estimates for the probability that females (black) and males (red) will reject the unfair distribution of rewards, for the entire cross-cultural sample. Shaded regions denote 95% percentile intervals for those estimates (males = red, females = grey). Vertical bars along the *x*-axis represent the proportion of the total sample at each age in the distribution, for both males (red bars) and females (black bars). (*b*) Lines represent regression estimates for the coefficient of the gender parameter, for the entire cross-cultural sample (black dashed line) and for the individual sites (pink lines). Estimates for each site are presented for the age range of the children in each site sample. Shaded region denotes 95% percentile intervals for the estimates for the entire cross-cultural sample (black dashed line). (Online version in colour.)

general gender difference at any age, when looking across all of the sites. The gender coefficient does not reliably deviate from zero at any age (figure 5*b*). By plotting the coefficients for each of the multi-level parameters for gender for each site (figure 5*b*, pink lines), we can see that the estimates for most of the sites appear largely similar to the overall pattern across sites.

### (ii) Do gender effects vary across sites?

We can again explore societal variation by plotting the coefficients for the multi-level parameters for gender for each site along with 95% percentile intervals and comparing the estimated coefficients for gender at each site to all of the other sites (figure 6). As with study 2a, this dataset provides substantial data for each site, particularly for participants of the youngest ages, leading to greater confidence about comparisons across individual sites. There is little indication of variation across sites in the estimates for gender for children before the age of 10 years. For older children, there is some indication of variation, with coefficients becoming increasingly more negative in the USA and Senegal, relative to Canada. However, sample sizes are relatively small at these ages, so this difference ought to be viewed with considerable caution.

### (iii) Study 2b discussion

This study's primary finding is that there is little evidence for overall cross-cultural gender differences in DI aversion. Males and females are relatively similar in their tendency to reject unfair outcomes that disadvantage themselves across these seven sites. There is also no substantial evidence for societal variation in DI aversion. Overall, these results largely replicate the findings of study 2a and show no indication of widespread gender differences in the development of fairness, across our target age range of about 4–15 years of age.

## 6. General discussion

The results of our analyses are consistent with the idea that gender has little impact on the development of prosocial behaviour and fairness during early childhood and early adolescence. These results strongly suggest that there is little evidence of substantial, widespread gender differences in experimental measures of prosociality and fairness within a large and diverse sample. In addition, there is relatively little evidence for variation in the effects of gender across societies. This research provides the strongest evidence available that gender does not strongly influence the development of prosociality and fairness in human children, but these results are based on samples from individual societies that were quite variable in size, limiting our ability to draw conclusions about cultural variation in gender differences. Below, we consider some implications of these findings in more detail.

First, we simply highlight that to date, no study has explored how gender influences on the development of prosociality and fairness in such a large sample of children from so many societies. In recent years, it has become widely acknowledged that there is cultural bias towards Western participant samples in psychology in general [29] and developmental psychology specifically [37], which makes it

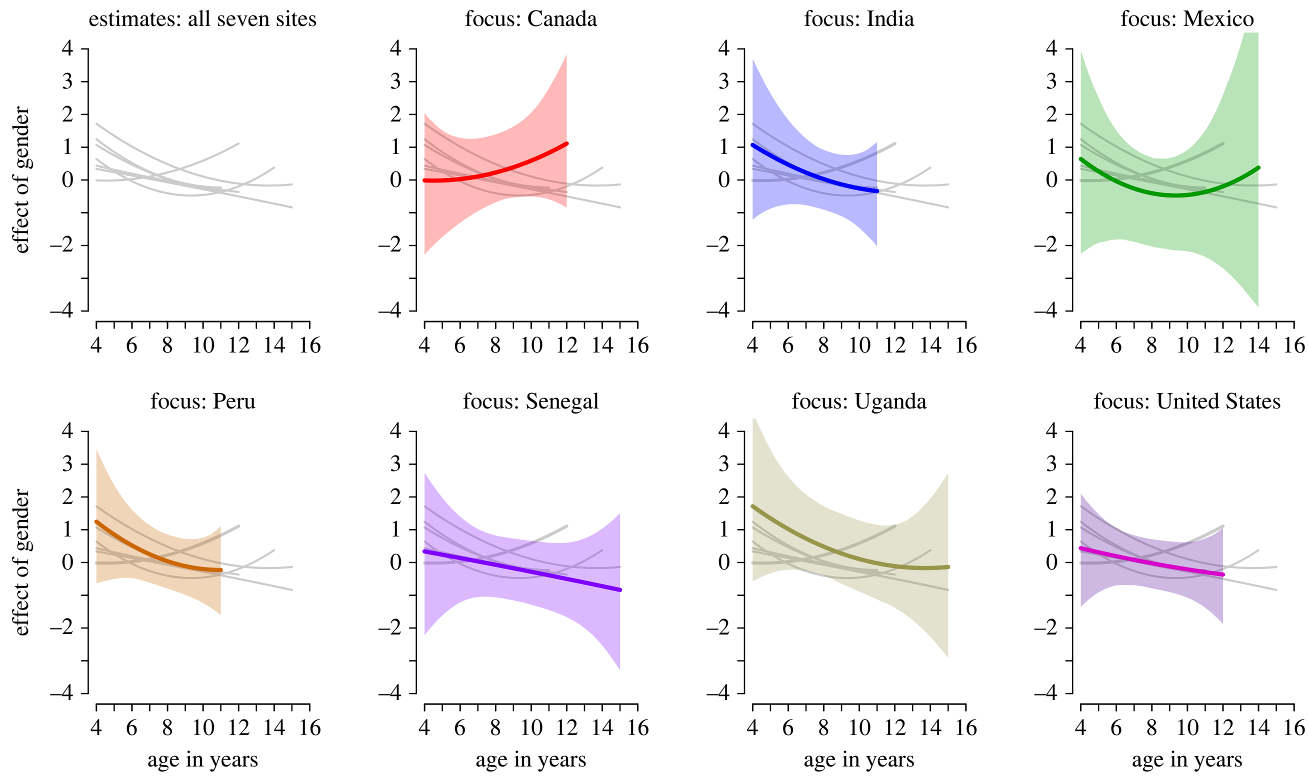


**Figure 6.** Results for the analysis of societal variation in gender differences in the DIG. Top left panel: grey lines represent regression estimates for the coefficient of the gender parameter, for each of the seven sites. All other panels: coloured lines and shaded regions represent regression estimates (and 95% percentile intervals) for the coefficient of the gender parameter for one of the sites, compared to the grey lines representing regression estimates for the coefficients for gender for each of the other sites. This allows comparisons to determine whether the estimates for each site differ from the estimates for the other sites. Estimates and percentile intervals for each site are presented for the age range of the children in each site sample. (Online version in colour.)

difficult to draw conclusions about findings in terms of human universals and cultural diversity. This has led to increasing calls for more diverse samples, and here we use datasets based on hundreds of children and adolescents spread across 7–14 populations that span a range of geographical regions, economic activities and subsistence strategies. The nature of this sample strengthens our primary conclusion that there is little evidence for universal human gender differences in the development of prosociality and fairness, but there are important caveats to consider.

Most importantly, we must be careful about extending these conclusions beyond contexts that are similar to the DG and the IG. These tasks are effective for measuring unilateral resource allocation decisions (i.e. sharing, generosity) in a systematic way. The advantage of using uniform methods across sites is that it maximizes the validity of our cross-cultural comparisons, but the disadvantage is that it limits the possibility of contextualizing methods or results for specific populations. Thus, while these methods offer some degree of ecological validity in that they include real stakes and peer interactions, they are of course not directly reflective of the decisions that children make in many real-life situations. It is therefore important to be careful about extending conclusions from these studies beyond the specific task context in which the data were recorded. It also means that the more diverse the sample is, the more difficult it will become for researchers to balance the advantages of uniform methods across sites (i.e. providing greater task comparability) against the disadvantages of using tasks which are necessarily decontextualized. Drawing from recent critiques of economic games which lack ecological validity [13], one approach would be to present participants in different societies with games that incorporate locally sourced information found in real-world social interactions. For example, presenting participants with recipients identified as belonging to local religious or cultural groups [15,16], or identifying specific recipients with whom the participants personally know [14]. This could make the behaviour elicited in these games more relevant for understanding behaviour in other contexts, while keeping the basic structure of the game comparable across sites. Our analysis (i.e. comparing data from datasets DG1 and DG2) did not suggest that making recipients identifiable to participants changed the degree to which gender predicted their behaviour, but future studies should provide children with information about recipients while *also* keeping their own identities anonymous [13].

Although we did find some hints that males may be somewhat more prosocial than females in the DG task, this finding is not clearly reliable and requires replication, and overall our results provide no strong evidence for universal gender differences in the development of prosocial preferences or egalitarianism. Theoretically, we could have observed no *overall* effect of gender across societies, but also large asymmetric effects of gender within each society—yet similarities in the effects of gender across societies suggest that this did not occur in these datasets. However, the present research does not necessarily contradict prior studies reporting gender differences in prosocial behaviour in both adults [31,32,36] and children [33,34]. Rather, our research leads to the conclusion that if these gender differences are real they are less likely to reflect universal human traits, and more likely to reflect society-specific patterns of socialization. Though in this study, we found little evidence for such societal variation in gender differences, this could simply mean that these differences are relatively small. Many of the individual societal samples from study 1 were small, and estimates for gender differences in individual societies here are interpreted with caution. What hints of societal variation in gender differences we did find were largely limited to the youngest children in the IG, where our samples were largest, but still these effects are tenuous. From these data, we can conclude that it is unlikely that children's behaviour in these games is substantially influenced by universal gender differences, or by large society-specific gender differences. Overall, this is consistent with the idea that any gender differences that exist are likely to be small and are unlikely to be directly caused by the individuals' gender. This suggests that other causal factors which are associated with gender should be explored, and that studying societal variation in the development of gender differences of cooperation will require very large samples of children.

To summarize, this project studied the development of gender differences in human prosocial behaviour across a large, culturally diverse sample of children from different societies. We find little evidence for widespread human gender differences in the development of children's choices in two specific experimental measures of prosociality and fairness: the DG and the IG. We also find little evidence for substantial societal variation in gender differences, particularly among children older than age 5–6 years, though samples from some societies are small. Although we are cautious about extending these findings too far beyond the specific task contexts that we studied, the data from our analyses do not support the idea that there are strong universal gender differences in prosocial behaviour and fairness behaviour in childhood and early adolescence, or that different societies display large (but asymmetric) gender differences in the development of these behaviours.

Data accessibility. Data, model code and full model results for this manuscript can be accessed at: https://osf.io/g796s/, and also in the electronic supplementary material [43].

Authors' contributions. B.R.H.: conceptualization, data curation, formal analysis, methodology, validation, visualization, writing—original draft, writing—review and editing; J.S.: conceptualization, methodology, writing—original draft, writing—review and editing; K.M.: conceptualization, methodology, writing—original draft, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. We received no funding for this study.

Acknowledgements. The following people were instrumental to the collection of the data on which the current studies are based: P. Kanngiesser, H.C. Barrett, T. Broesch, S. Cebioglu, A.N. Crittenden, A. Erut, S. Lew-Levy, C. Sebastian-Enesco, A.M. Smith, S. Yilmaz, P. Blake, J. Corbit, T.C. Callaghan, O. Barry, A. Bowie, L. Kleutsch, K.L. Kramer, E. Ross, H. Vongsachang and R. Wrangham & F. Warneken.

## References

1. Henrich J *et al.* 2010 Markets, religion, community size, and the evolution of fairness and punishment. *Science* **327**, 1480–1484. (doi:10.1126/science.1182238)
2. Henrich J, Boyd R, Bowles S, Camerer C, Fehr E, Gintis H, McElreath R. 2001 In search of homo economicus: behavioral experiments in 15 small-scale societies. *Am. Econ. Rev.* **91**, 73–78.
3. Henrich J *et al.* 2005 In cross-cultural perspective: behavioral experiments in 15 small-scale societies. *Behav. Brain Sci.* **28**, 795–815. (doi:10.1017/S0140525X05000142)
4. Ensminger J, Henrich J. 2014 *Experimenting with social norms: fairness and punishment in cross-cultural perspective*. New York, NY: Russell Sage Foundation.
5. Fehr E, Fischbacher U, Gächter S. 2002 Strong reciprocity, human cooperation, and the enforcement of social norms. *Hum. Nat.* **13**, 1–25. (doi:10.1007/s12110-002-1012-7)
6. Fehr E, Fischbacher U. 2004 Social norms and human cooperation. *Trends Cogn. Sci.* **8**, 185–190. (doi:10.1016/j.tics.2004.02.007)
7. Brown JK. 1970 A note on the division of labor by sex. *Am. Anthropol.* **72**, 1073–1078. (doi:10.1525/aa.1970.72.5.02a00070)
8. Murdock GP, Provost C. 1973 Factors in the division of labor by sex: a cross-cultural analysis. *Ethnology* **12**, 203–225. (doi:10.2307/3773347)
9. House BR *et al.* 2020 Universal norm psychology leads to societal diversity in prosocial behaviour and development. *Nat. Hum. Behav.* **4**, 36–44. (doi:10.1038/s41562-019-0734-z)
10. Spadaro G, Jin S, Balliet D. 2022 Gender differences in cooperation across 20 societies: a meta-analysis. *Phil. Trans. R. Soc. B* **378**, 20210438. (doi:10.1098/rstb.2021.0438)
11. Henrich J. 2006 Cooperation, punishment, and the evolution of human institutions. *Science* **312**, 60–61. (doi:10.1126/science.1126398)
12. Gurven M, Winking J. 2008 Collective action in action: prosocial behavior in and out of the laboratory. *Am. Anthropol.* **110**, 179–190.
13. Pisor AC, Gervais MM, Purzycki BG, Ross CT. 2020 Preferences and constraints: the value of economic games for studying human behaviour. *R. Soc. Open Sci.* **7**, 192090. (doi:10.1098/rsos.192090)
14. Gervais MM. 2017 RICH economic games for networked relationships and communities: development and preliminary validation in Yasawa, Fiji. *Field Methods* **29**, 113–129. (doi:10.1177/1525822X16643709)
15. Pisor AC, Gurven M. 2018 When to diversify, and with whom? Choosing partners among out-group strangers in lowland Bolivia. *Evol. Hum. Behav.* **39**, 30–39. (doi:10.1016/j.evolhumbehav.2017.09.003)
16. Purzycki BG, Apicella C, Atkinson QD, Cohen E, McNamara RA, Willard AK, Xygalatas D, Norenzayan A, Henrich J. 2016 Moralistic gods, supernatural punishment and the expansion of human sociality. *Nature* **530**, 327–330. (doi:10.1038/nature16980)
17. House BR, Silk JB, Henrich J, Barrett HC, Scelza BA, Boyette AH, Hewlett BS, McElreath R, Laurence S. 2013 Ontogeny of prosocial behavior across diverse societies. *Proc. Natl Acad. Sci. USA* **110**, 14 586–14 591. (doi:10.1073/pnas.1221217110)
18. Blake PR *et al.* 2015 The ontogeny of fairness in seven societies. *Nature* **528**, 258–261. (doi:10.1038/nature15703)
19. Blake PR, McAuliffe K. 2011 'I had so much it didn't seem fair': eight-year-olds reject two forms of inequity. *Cognition* **120**, 215–224. (doi:10.1016/j.cognition.2011.04.006)
20. House BR, Tomasello M. 2018 Modeling social norms increasingly influences costly sharing in middle childhood. *J. Exp. Child Psychol.* **171**, 84–98. (doi:10.1016/j.jecp.2017.12.014)
21. Rusch H. 2014 The evolutionary interplay of intergroup conflict and altruism in humans: a review of parochial altruism theory and prospects for its extension. *Proc. R. Soc. B* **281**, 20141539. (doi:10.1098/rspb.2014.1539)
22. Fehr E, Bernhard H, Rockenbach B. 2008 Egalitarianism in young children. *Nature* **454**, 1079–1083. (doi:10.1038/nature07155)
23. Leinbach MD, Fagot BI. 1993 Categorical habituation to male and female faces: gender schematic processing in infancy. *Infant Behav. Dev.* **16**, 317–332. (doi:10.1016/0163-6383(93)80038-A)

24. Quinn PC, Yahr J, Kuhn A, Slater AM, Pascalis O. 2002 Representation of the gender of human faces by infants: a preference for female. *Perception* **31**, 1109–1121.
25. Quinn PC, Uttley L, Lee K, Gibson A, Smith M, Slater AM, Pascalis O. 2008 Infant preference for female faces occurs for same- but not other-race faces. *J. Neuropsychol.* **2**, 15–26. (doi:10.1348/174866407X231029)
26. Dunham Y, Baron AS, Carey S. 2011 Consequences of 'Minimal' group affiliations in children. *Child Dev.* **82**, 793–811. (doi:10.1111/j.1467-8624.2011.01577.x)
27. Renno MP, Shutts K. 2015 Children's social category-based giving and its correlates: expectations and preferences. *Dev. Psychol.* **51**, 533–543. (doi:10.1037/a0038819)
28. Kramer K. 2022 Female cooperation. Evolutionary, cross-cultural and ethnographic evidence. *Phil. Trans. R. Soc. B* **378**, 20210425. (doi:10.1098/rstb.2021.0425)
29. Henrich J, Heine SJ, Norenzayan A. 2010 Beyond WEIRD: towards a broad-based behavioral science. *Behav. Brain Sci.* **33**, 111–135. (doi:10.1017/S0140525X10000725)
30. House BR. 2018 How do social norms influence prosocial development? *Curr. Opin. Psychol.* **20**, 87–91. (doi:10.1016/j.copsyc.2017.08.011)
31. Eagly AH, Crowley M. 1986 Gender and helping behavior: a meta-analytic review of the social psychological literature. *Psychol. Bull.* **100**, 283.
32. Croson R, Gneezy U. 2009 Gender differences in preferences. *J. Econ. Literature* **47**, 448–474.
33. Fabes RA, Eisenberg N. 1998 Meta-analyses of age and sex differences in children's and adolescents' prosocial behavior. In *Handbook of child psychology, 3* (ed. W Damon), pp. 1–29. New York, NY: Wiley.
34. Benenson JF, Durosky A, Nguyen J, Crawford A, Gauthier E, Dubé É. 2019 Gender differences in egalitarian behavior and attitudes in early childhood. *Dev. Sci.* **22**, e12750. (doi:10.1111/desc.12750)
35. Camerer C. 2003 *Behavioral game theory: experiments in strategic interaction*. Princeton, NJ: Princeton University Press.
36. Eagly AH. 2009 The his and hers of prosocial behavior: an examination of the social psychology of gender. *Am. Psychol.* **64**, 644–658.
37. Nielsen M, Haun D, Kärtner J, Legare CH. 2017 The persistent sampling bias in developmental psychology: a call to action. *J. Exp. Child Psychol.* **162**, 31–38. (doi:10.1016/j.jecp.2017.04.017)
38. Franzen A, Pointner S. 2012 Anonymity in the dictator game revisited. *J. Econ. Behav. Org.* **81**, 74–81. (doi:10.1016/j.jebo.2011.09.005)
39. Krysowski E, Tremewan J. 2021 Why does anonymity make us misbehave: different norms or less compliance? *Econ. Inquiry* **59**, 776–789. (doi:10.1111/ecin.12955)
40. R Development Core Team. 2018 *R: a language and environment for statistical computing (3.5.1)*. Vienna, Austria: R Foundation for Statistical Computing. See www.R-project.org.
41. McElreath R. 2015 *Statistical rethinking*. London, UK: Chapman and Hall.
42. Stan Development Team. 2018 RStan: the R interface to Stan. (2.18.2). See http://mc-stan.org.
43. House B, Silk JB, McAuliffe K. 2022 No strong evidence for universal gender differences in the development of cooperative behaviour across societies. Figshare. (doi:10.6084/m9.figshare.c.6289742)